Multimodal, Embodied and Location-Aware Interaction

A dissertation

submitted for the degree of

Doctor of Philosophy

by

Steven Strachan, BSc., MSc.

Supervisor: Roderick Murray-Smith

Hamilton Institute

National University of Ireland, Maynooth

Ollscoil na hÉireann, Má Nuad

January 2007

Contents

Abstract

This work demonstrates the development of mobile, location-aware, eyes-free applications which utilise multiple sensors to provide a continuous, rich and embodied interaction. We bring together ideas from the fields of gesture recognition, continuous multimodal interaction, probability theory and audio interfaces to design and develop location-aware applications and embodied interaction in both a small-scale, egocentric body-based case and a large-scale, exocentric 'world-based' case.

BodySpace is a gesture-based application, which utilises multiple sensors and pattern recognition enabling the human body to be used as the interface for an application. As an example, we describe the development of a gesture controlled music player, which functions by placing the device at different parts of the body. We describe a new approach to the segmentation and recognition of gestures for this kind of application and show how simulated physical model-based interaction techniques and the use of real world constraints can shape the gestural interaction.

GpsTunes is a mobile, multimodal navigation system equipped with inertial control that enables users to actively explore and navigate through an area in an augmented physical space, incorporating and displaying uncertainty resulting from inaccurate sensing and unknown user intention. The system propagates uncertainty appropriately via Monte Carlo sampling and output is displayed both visually and in audio, with audio rendered via granular synthesis. We demonstrate the use of uncertain prediction in the real world and show that appropriate display of the full distribution of potential future user positions with respect to sites-of-interest can improve the quality of interaction over a simplistic interpretation of the sensed data. We show that this system enables eyes-free navigation around set trajectories or paths unfamiliar to the user for varying trajectory width and context. We demonstrate the possibility to create a simulated model of user behaviour, which may be used to gain an insight into the user behaviour observed in our field trials. The extension of this application to provide a general mechanism for highly interactive context aware applications via density exploration is also presented. AirMessages is an example application enabling users to take an embodied approach to scanning a local area to find messages left in their virtual environment.

Declaration

I hereby certify that this material, which I now submit for assessment on the program study leading to the award of Doctor of Philosophy in Computer Science is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

signed: ID No.: 63152312

Date :

Contributing Publications

The papers contributing to chapter 3 are:

- S. Strachan, R. Murray-Smith, S. O'Modhrain, "BodySpace: inferring body pose for natural control of a music player", Extended abstracts of ACM SIG CHI Conference, San Jose, 2007.

- S. Strachan, R. Murray-Smith, "BodySpace: Multi-Modal Interaction with Mobile Devices", 2nd Joint Workshop in Multimodal Interaction and Related Machine Learning Algorithms, Royal College of Physicians, Edinburgh, UK, July 2005.

- S. Strachan, R. Murray-Smith, "Muscle Tremor as an Input Mechanism", UIST 2004, Santa Fe, 2004.

- S. Strachan, R. Murray-Smith, I. Oakley, J. Ängeslevä,"Dynamic Primitives for Gestural Interaction", Mobile Human-Computer Interaction MobileHCI 2004: 6th International Symposium, Glasgow, UK, September 13-16, 2004. Proceedings. Stephen Brewster, Mark Dunlop (Eds), LNCS 3160, Springer-Verlag, p325-330, 2004.

Papers contributing to chapter 4 are listed below. In the first two papers the sections on Monte Carlo propagation for browsing virtual environments were contributed principally by John Williamson. The construction of the application and the user study were both contributed by the author of this thesis:

- S. Strachan, J. Williamson, R. Murray-Smith, "Show me the way to Monte Carlo: density-based trajectory navigation", Proceedings of ACM SIG CHI Conference, San Jose, 2007.

- J. Williamson, S. Strachan, R. Murray-Smith, "Its a long way to Monte Carlo: Probabilistic display in GPS navigation", Proceedings of Mobile HCI 2006, Helsinki, 2006, p89-96.

- S. Strachan, P. Eslambolchilar, R. Murray-Smith, S. Hughes, S. O'Modhrain, "gpsTunes - controlling navigation via audio feedback", Mobile HCI 2005, p275-278.

The paper contributing most to chapter 5 was:

- S. Strachan, R. Murray-Smith, "airMessages: embodied, location-aware interaction", CHI 2007 Workshop on Mobile Spatial Interaction, San Jose, April 28th, 2007.

Acknowledgments

First and foremost I'd like to thank Dr Roderick Murray-Smith for his outstanding supervision during the course of this work. Without his almost never ending stream of ideas, talks and general guidance, this work could not have been completed. I'd also like to thank Dr John Williamson for his contribution to this work. The complementary nature of our two streams of work has undoubtedly aided the completion of this thesis.

I'd like to acknowledge Stephen Hughes who built the hardware, which was vital for all of my research and Andrew Ramsay who wrote the software, which made the use of this hardware so much easier than it could have been.

A big thank you should also go to Dr Ian Oakley and Dr Sile O'Modhrain for their supervision in the early part of my research at Media Lab Europe, when there was a lot to be learned very quickly, and to Jussi Ängeslevä for the kind use of his design concept. Also at this time and throughout the Ph.D. Dr Andrew Crossan and Parisa Eslambolchilar were a very great help in everything from what whisky to drink in the pub to the intricacies of eMbedded Visual C++.

The Higher Education Authority and Science Foundation Ireland funded me throughout this project and for that I am grateful. A big thank you goes also to my co-workers at the Hamilton Institute and to the people I have interacted with within NUI Maynooth who both professionally and socially have made the past three years very enjoyable.

Chapter 1

Introduction

1.1 Designing Interaction

The computer is no longer just for performing calculations, writing and printing documents or storing information, it acts as a tool for access to a vast source of information, it acts as a link to our loved ones, to social interactions and networks, it effectively acts as a portal to the world. The emphasis here then should be on the word *tool*. The rise of mobile computing has acted as a catalyst for the emergence of a new kind of location-aware interaction. The design of a richer, more embodied interaction with our devices in this location-aware context is highly desirable and to achieve this, we must consider the field of interaction design. Interaction design is defined by Preece *et al.* (2002) as the design of spaces for human communication and interaction. In particular it is about experiences that enhance the way in which people communicate and interact (Preece *et al.* 2002). Our mobile device will act as tool for this interaction and in this work we attempt to design and demonstrate highly interactive *tools* for use with location-aware mobile computing. In order to achieve this we must adopt a new approach to interaction design, one more appropriate for this new, highly mobile and embodied kind of location-aware interaction. First we should learn from more traditional forms of interaction.

1.2 Instrumenting Interaction

Human beings have always used tools to interact with their surroundings. The most fundamental tool we poses as humans is our hands but the hand alone is not enough to achieve many tasks. People needed to *extend* their powers in order to achieve more and this led to the development of the first hand-held tools. Hammers, axes and chisels, all very primitive, are the most basic of human inventions and may be thought of as some of the earliest examples of interaction design. A skilled craftsman is well practiced and perceives his tool effectively as an extension of himself or his powers. He becomes part of a tightly coupled feedback loop involving himself and his tool. He brings together his skills and intentions via this feedback loop and learns both the subtleties of his tool and the consequences of his actions to the point where the tool itself becomes almost transparent. This is the most basic and fundamental notion of a tool and these notions of extension, transparency and ubiquity are something we should strive to achieve when developing new tools or designing interaction for our purposes. Through

Figure 1.1: Basic tools.

the industrial age we saw the development of advanced mechanical tools designed to extend the power of the human hand to a far greater range of applications. This serves to broaden the definition of a tool, since strictly speaking these tools were no longer handheld. McCullough (1998) created the following definition:

> *A tool is a moving entity whose use is initiated and actively guided by a human being, for whom it acts as an extension, toward a specific purpose*

we can think of the entity as being either physical or conceptual, the motion may be manual or machine powered, the guidance may be manual or by indirect control. A tool, by definition, is for serving intent, whereas a machine can operate on its own. But this does not mean that a machine, which possess a certain amount of autonomy, can not be thought of as a real tool. A pilot can still perceive the plane to be his tool, even if much of the low level functionality of the plane is automated because he is still a very important part of the whole control loop. It is for this reason that the notion of a computer as a tool was originally met with some skepticism, but by thinking of the computer and user as a tightly coupled loop of control, we can begin to visualise the computer as a tool, which the user manipulates depending on their intentions. This notion of using control theory in interaction design is a important part of this thesis. By thinking of the 'loop of control' as the basis for our interaction design and building our interaction from here, we offer a new method of interaction design to the community. This is important for the emergence of handheld computing since the computer (as in figure 1.2), for the first time, becomes a truly handheld *tool* more suited to a continuous control style of interaction as apposed to the more discrete traditional forms of interaction.

1.3 Computer-Human Interaction

The early skepticism regarding the use of a computer as a tool, despite the undeniable 'machine-like' qualities of early computers, was most likely due to the fact that the earliest interfaces were designed by engineers for engineers and although they were highly functional for their particular needs, they had not necessarily been explicitly *designed* with general usability concerns in mind. The emergence of personal computers used by the general public brought with it a necessity for the design of interfaces and interaction accessible to all but was ultimately hindered by remnants from these early interfaces. Beale and Edwards (1999) argue that current input devices, such as a keyboard, are more oriented towards the requirements of the computer than the user and state that human-computer interaction should be more akin to everyday human-human interaction. With the

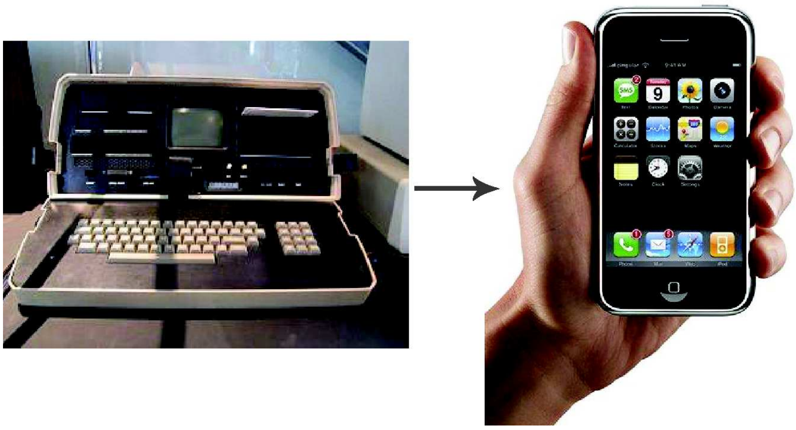


Figure 1.2: Old and new.

emergence of the field of human-computer interaction and the inclusion of psychologists, sociologists and computer scientists in the field of interaction design, computers have ultimately emerged in recent years to be one of our most powerful tools for creativity (McCullough 1998).

1.3.1 Input

Traditionally computers have produced large amounts of output for relatively little input. The use of discrete key pressing activity on a keyboard provides a low bit-rate when compared to modern day continuously controlled applications, which take data simultaneously from multiple sensors. This kind of system has become more prominent recently as these systems become more mobile and consequentially more location-aware, more context-aware and are equipped with more and more sensing. We are set to see the amount of input increase dramatically, perhaps even surpassing the level of output for the first time. With our own equipment, described later, we have the potential to take data simultaneously from up to thirteen different sensors while only displaying information to users via the visual, audio or haptic channels. It is essential then that we take advantage of this increased input bandwidth and develop tools which take real advantage of this switch.

There are a considerable number of input devices on the market for desktop-based computers. Apart from the obvious keyboard and mouse, we have joysticks, touch pads, trackballs, touch-screens, tablets and gloves to name but a few. In contrast, mobile devices have relatively little in the way of input capability. The only readily accepted and available forms of input for mobile computing are key-pads, buttons, touch-screens and for some devices, externally attachable qwerty keyboards, but even these are considered inadequate for effective input, especially when used whilst mobile.

The startlingly paced emergence of mobile computing has exposed old interaction metaphors as being inadequate in this domain. In the literature, two main problems regarding mobile usability are cited. The first problem cited is limited screen size (Fishkin *et al.* 2000). And the second is text entry, which is still extremely limited when using a mobile device, although we have seen some positive research into developing new methods of text entry for mobile computers (Zhai and Kristensson 2003, Ward *et al.* 2000, Williamson and Murray-Smith 2005*a*). The use of mice is obviously not practical for any mobile device and the compromised software keyboard causes significant difficulty for most people. The stylus may obscure the small screen, obscuring information the user may like to interact with and involves the use of both hands. Also, the repeated tapping movements required can become tedious to use for a long period of time and they demand a great deal of visual attention. They are reasonable for the entry of small amounts of data in almost any environment, but as soon as it is required to enter a large amount of data this method quickly becomes inadequate (Fallman 2002).

Handwriting recognition has also been commonly employed on mobile devices. Various handwriting recognition systems, such as *Graffiti*, have been relatively successful but the fundamental weakness of handwriting recognition is its speed, typically around 15 wpm (Card *et al.* 1983). This is poor compared to the average 30 wpm with a keyboard on a desktop computer. It can be argued then that mobile computing has almost *demanded* a rethink of the fundamentals of interaction design in the search for a more natural and less obtrusive interaction technique.

1.3.2 Novel Interaction

New kinds of interaction with computers such as voice control, gestural interaction and more recently, brain-computer interaction (Blankertz *et al.* 2006, Blankertz *et al.* 2007) have begun to emerge but are still in their infancy and it remains to be seen whether these kinds of new technology can surpass the traditional keyboard and mouse.

The use of gesture is often mentioned as one possibility for a more natural and expressive interaction technique. Gestures are a natural form of communication for humans. Purcell (1985) notes:

> "*the fluidity and expressiveness of human gesture is a fundamentally important component of interpersonal communication*"

It is argued that this interpersonal communication between humans should equally be applied to the communication that takes place between humans and computers and Wexelblat (1998) suggests that:

> "*gestural interaction could herald a new kind of environment, one in which people could interact with computers in a natural, conversational manner*"

It seems natural then that we would want to incorporate at least some form of gestural interaction into our mobile tools. It is one of the aims of this thesis to initiate the development of software tools specifically designed for location-aware applications, since by definition mobile computers are *mobile*. We wish to introduce the concept of using the mobile computer as an instrument or tool, which incorporates the both fundamentals of gesture and allows interaction with the environment in a rich and embodied manner.

1.3.3 Embodied Interaction

The Graphical User Interface was an important invention for desktop computing. It removed the need for a knowledge of complex command-line languages and presented the inner workings of the computer to the user in a graphical form. It was a metaphoric representation of everyday objects with which a novice user could interact with and have at least an inclination of what would happen if they, for example, moved the file symbol onto the trash can symbol. Fishkin *et al.* (2000) notes:

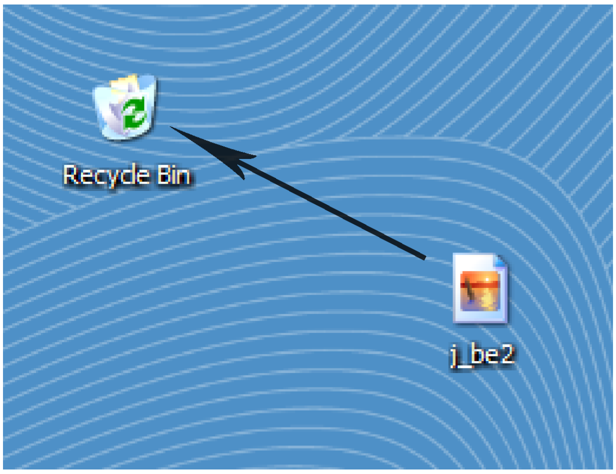


Figure 1.3: The traditional metaphor used for deleting a file on the desktop

> *It is interesting to note that while a whole sensory-motor world is created within the confines of the computer display, the physical computer itself has become an anonymous, invisible "box". All the attention is on a disembodied display.*

With the emergence of mobile computing it has become necessary to move forward from this disembodied GUI. One of the main differences between mobile computing and desktop computing is the ability to take a mobile device out into the real world. They are held in the hand and touched and the addition of sensing opens up the opportunity to shake, flick, gesture or tilt the devices, we can point them at things or point them at ourselves, we can use our natural movements and involve our body and our limbs in ways that evolution has allowed us to do for a very long time. McCullough (2005) notes:

> *Place begins with embodiment. Body is place, and it shapes your perceptions. Embodiment is not just a state of being but an emergent quality of interactions.*

We must then recreate these everyday metaphors, utilised so effectively by the GUI, in the context of mobile computing and embodied interaction. Should it now be possible, instead of using your mouse to drag the file symbol to the trash can symbol, to take the file from the place you stored it on your body and physically place it in the virtual trash can in front of you? Rukzio *et al.* (2006) describe an experimental comparison of some currently used embodied mobile interaction techniques, i.e. touching, pointing and

Figure 1.4: A file is removed from the body and placed in the virtual trash can.

scanning, finding that location is by far the most important factor for the selection of the correct metaphor within a given context. This leads us then to think about how we might construct a new kind of location-aware interface and what kind of embodied metaphors for interaction should we choose? In chapter 3 we design our interaction around the body to produce an egocentric interface with embodied gestures to different parts of the body. In chapter 4 we expand the interaction out into the space of the real world to produce an exocentric interface which requires users to interact with and explore the space in an embodied manner. These two very different kinds of embodied interaction demonstrate the potential for the use of embodiment and embodied interaction in our designs, which is made possible by the emergence of mobile computing. Again though we must define new ways of thinking about the design of this interaction and lay down some new principles for interaction design in this embodied and spatial context.

1.4 Spatial Interfaces

The notions of "space" and spatial organisation to facilitate and structure our interaction is something intuitive. The world in which we live is structured in a way which moulds and guides our actions and interactions. We are always arranging our most commonly used tools around ourselves. Our mobile phones are always within arm's reach. In the office we have our computers, our books and our filing cabinets all close at hand and arranged in a spatial way. It is likely that you could reach for something on your desk and grab it even with your eyes closed because spatial mapping and the representation of our own personal space in our minds is very powerful. So it is important that we exploit this power in the design of our interfaces and this is demonstrated in chapter 3.

Harrison and Dourish (1996) argue that a focus on spatial models is misplaced. Drawing on understandings from architecture and urban design they highlight a critical distinction between "space" and "place". While designers use spatial models to support interaction, they show how it is actually a notion of "place", which frames interactive behaviour. They explain that space is about the structure of the three dimensional world in which objects and events occur, and in which they have relative position and direction. Physically, a place is a space with which we poses understandings of behavioral appropriateness and cultural expectations.

> *We are located in "space", but we act in "place". Furthermore, "places" are spaces that are valued. The distinction is rather like that between a "house" and a "home"; a house might keep out the wind and the rain, but a home is where we live.*

The notion of egocentric and exocentric interfaces becomes important here. Marentakis and Brewster (2005) define, in their work with a spatial audio interface, an egocentric display to be one where the sound position is fixed to the user and remains unchanged no matter what the direction of the user is, whereas in an exocentric presentation, sound position is fixed to the world and is updated in real time based on the direction of the user relative to the sounds. In an egocentric setting we have the ability to build highly personal spatial interfaces, such as that in chapter 3 where users have the ability to associate different files or functionality with different parts of their body, which is obviously a very personal, intimate and egocentric thing. Exocentric spatial interfaces are much more open and expansive. The work conducted in chapter 4 shows that it is possible to treat the world as our interface and here is where it becomes important to consider the notion of 'place'. This kind of interface opens up the possibility to interact with objects placed in our interface or even other people in other places, creating new opportunities for the exploration of the social mechanisms for this kind of interaction.

More and more often researchers involved with mobile computing and interaction design are considering the relationship between the human and their mobile device. Socially, the mobile computer has become very important and people have come to rely on their devices as an important part of their lives. Ask a typical user how they would feel if they lost their device and they would likely tell you that they would feel lost, disconnected, invisible or even naked. This is due to the subconscious connection that people make between their mobile phones and their social lives, and with their connections to friends and family. Removing the device causes a subconscious connection to a loved one to be broken or a potential invite to a social gathering to be lost. These emotional responses to the current state of mobile technology are a hint that our devices are becoming an important part of a user's life.

Although there has been a considerable amount of work conducted in the field of audio interfaces, there is still a desire to develop more tools for use in this eyes-free, location-aware context. We are set to see a move away from more traditional, visual, screen-based user interfaces. Audio interfaces are considered more appropriate for mobile computers being used 'on the move' since most visual attention should be allocated to more safety critical tasks, such as avoiding passing cars, and there has been work conducted on 'eyes-free' interaction (Cohen and Ludwig 1991, Savidis *et al.* 1996, Brewster *et al.* 2003). In a mobile computing context the environment in which the user of a mobile device finds himself is constantly changing but existing interfaces are largely failing to take advantage of these contextual changes. We wish to incorporate this changing context into our interaction designs, giving our interfaces a strong location-aware and 'eyes-free' flavour but to make this possible we also need to consider the natural uncertainties these changing contexts bring us.

1.5 Embracing Uncertainty

One of the fundamental problems affecting the development and general acceptance of novel mobile interfaces comes from the omnipresent or 'always present' uncertainties in our sensor measurements and the fact that these sensors are usually indirect proxies for what we really want to measure. Uncertainties can arise from various sources including general internal sensor noise or noise from the outside world. It is essential then that we embrace this uncertainty in a way which makes our interfaces more acceptable. But how do we approach this and what effect will this have on the users of our system?

Location-based games are becoming more prominent as technology develops. Recent examples of location-based games include *AR Quake*(Piekarski and Thomas 2002), *Treasure* (Chalmers *et al.* 2005), *Pirates!* (Björk *et al.* 2001), *Mindwarping* (Starner *et al.* 2000) and *Feeding Yoshi* (Bell *et al.* 2006), demonstrating how handheld computers and see-through head-mounted displays can be combined with sensing systems such as GPS and video-tracking to create experimental gaming experiences. These projects offer glimpses of the potential new applications for location-based technologies. They are especially useful for studying how participants experience location and context-sensing technologies when confronted with considerable technical uncertainty arising from GPS or wireless network coverage. (Benford *et al.* 2005).

Can You See Me Now (Benford *et al.* 2006) is a game where online players are chased through a virtual model of a city by runners equipped with GPS and WiFi technologies. The runners are required to run through the actual city streets in order to catch the online players. They present an ethnographic study of the game, which reveals the diverse ways in which online players experienced the uncertainties inherent in GPS and in WiFi. Mostly the participants were unaware of these uncertainties, but sometimes they saw them as a problem, or treated them as a feature of the game, and even occasionally exploited uncertainty within game play. The authors encourage designers to deal with such uncertainties as a fundamental characteristic of location-based experiences rather than treating them as exceptions or bugs that might be ironed out in the future.

Within this thesis it is one of the aims to examine the effects of uncertainty and explore whether a truthfully uncertain display, which we may also refer to as an *honest* display can actually improve experiences with location-aware applications. Can the proper representation or exposure of uncertainty help improve control performance in an interaction task?

1.6 Modelling Interaction

Many successful computer interfaces have been implemented over the years but there remains a distinct lack of well-founded theoretical principles in their design. There exist many interfaces which have been carefully designed at a high-level, which have then been evaluated with many users with some statistical analysis applied to show the usability of that interface but there are still few solid theoretical frameworks for interaction, which can describe and predict behaviour from the lowest motor-level control actions to the highest level goals and intentions. Thimbleby (1990) notes:

> *I find reports of experiments sometimes related to my particular problem but without some underlying theories, how can I know how safely I can generalise those results to apply in my design, with my users, in my language?*

In this work we consider model-based approaches to our interaction, which are considered to be beneficial for a number of reasons. By basing our interaction on a simulation of a physical model we provide the user with a natural intuitiveness which is not present for non-model based approaches. This approach is also beneficial since it enables a more active exploration of the potential range of interaction with a device, encouraging a user to discover the limits of their interface and maybe use it in unusual ways. Others have illustrated the benefits of this kind of approach. Eslambolchilar *et al.* (2004) describe a working system where they support human behaviour in a document browsing task by adapting the modeled dynamics of the navigation and the visual feedback. Rath and Rocchesso (2005) advocate the use physics-based sound models, which they hypothesise can afford immediate and accurate grasping of everyday phenomena, giving their own example of a rolling ball on a beam, which they say can be used as a metaphor in a variety of interaction tasks. Their initial experiments indicate that the performance in continuous interaction tasks can be improved by carefully designed physics-based sound models. Cook and Lakatos (2003) described a series of studies in human auditory perception based on the use of physical synthesis models. They find that use of realistic physically-based models allows individual parameters to be isolated and tested. Rocchesso *et al.* (2003) describe the use of "cartoonified" physical sound models, where the physics have been exaggerated somewhat in order to increase both computational efficiency and the perceptual sharpness for the user. This is an important engineering benefit, akin to Computer Generated graphics or animation, where the altering of simple parameters allows the whole look and feel of a system to change. This more scientific approach to designing interaction is helpful since it allows us to designed our interaction such that the 'look and feel' of the interaction may be altered or shaped by simply adjusting the parameters of our model.

1.7 Thesis Outline

1.7.1 Our Place

It is important that this thesis has its place. The main aim of this thesis is to demonstrate the potential for rich and highly interactive location-aware applications, which take advantage of newly emerging sensing capabilities. It is hoped that this thesis will act as a bridge between the worlds of human-computer interaction, interaction design and control engineering with the purpose of introducing and demonstrating the use of the kind of tools from the world of control engineering, which have natural applications to interaction design and have a very strong and rich history, which can be of benefit to the HCI community. We also wish to formulate and present some new design ideas and principles from this control theory basis and present them in a way which is useful to researchers from the field of HCI.

1.7.2 Structure

Chapter 2 introduces the theoretical background to the work, expanding on the themes noted above and introducing some tools to be used and challenges to be met in the course of this work. The essential characteristics of interaction and the signals associated with this interaction which we are required to understand are introduced and the use inertial sensing in the domain of mobile computing is discussed.

Chapter 3 introduces *BodySpace*. This is a gesture-based application, which uses inertial sensing and pattern recognition to enable the human body to be used as the egocentric interface for an application, in this example, for a music player. We describe the development of a gesture controlled music player, which functions by placing the device at different parts of the body. We describe a new approach to the segmentation and recognition of gestures for this kind of application and show how simulated physical model-based interaction techniques can shape the gestural interaction. We describe how this is received by users in informal testing.

Chapter 4 introduces the *GpsTunes* application. This is a mobile, GPS-based multimodal navigation system, equipped with inertial control that allows users to explore and navigate through an exocentric augmented physical space, incorporating and displaying the uncertainty resulting from inaccurate sensing and unknown user intentions. The system propagates uncertainty appropriately via Monte Carlo sampling. Control of the Monte Carlo exploration is entirely tilt-based and the system output is displayed both visually and in audio. Audio is rendered via granular synthesis (described in section 4.5.1) to accurately display the probability of the user reaching targets in the space. We also demonstrate the use of uncertain prediction in a trajectory following task, where a section of music is modulated according to the changing predictions of user position with respect to the target trajectory.

Chapter 5 brings together the work conducted in chapters 3 and 4 to introduce a mechanism for providing highly interactive context and location-aware applications and describe as an example the *AirMessages* system. *AirMessages*, allows the dropping and retrieval of text messages in the real-world via a gesture-based interface. An informal user study is conducted, which highlights some potential problems, improvements and provided some interesting insights into the use of this kind of system.

1.8 Thesis Claims

The work in this thesis explores the design space of mobile devices equipped with inertial and location sensing and audio and vibrotactile feedback. From a design-theoretic point of view we have introduced a new style of designing interaction, which considers interaction design from a more theoretical level, starting from the basic notion of treating this kind of continuous interaction as a loop of control and building the application around this principle, thinking carefully about inputs to the system, processing of those inputs and the fedback provided. We have created interfaces using these principles, which use the egocentric body and exocentric real-world environment as interfaces to interact with objects. We have also demonstrated the utility and generality of a model-based approach to the interaction with mobile devices with the aim of allowing other HCI researchers to extract this approach and adapt it to their own interfaces. We have demonstrated the utility of incorporating uncertainty and constraints into the interaction design, which it is hoped can be adopted for the general improvement of interaction with location-aware applications.

From a technical point of view we have developed a new approach to the detection and segmentation of body-based gestures using an end-point or goal state detection approach. We have demonstrated a dynamic systems approach to the representation of gestures. We have shown that the feeding back of uncertainty can improve performance in location-based interaction and that the use of natural constraints in the environment can aid interaction. We have shown that appropriate display of the full distribution of potential future user positions with respect to sites-of-interest using Monte Carlo sampling can improve the quality of interaction over a simplistic interpretation of the sensed data and demonstrated the use of tilt to control the Monte Carlo sampling time horizon and the use of magnetometers to provide rapid bearing updates, enabling the sensing of content in the local environment via multimodal (audio and vibration) sensing. We introduce new metrics for usability analysis, which provide objective measures of the way in which a system was used.

Novel applications have been developed. Applications of the BodySpace system, i.e. *BodyMusic* and *Off-The-Wall Interaction* are described. We also introduce the *gpsTunes* application and a derivation of this known as *airMessages*.

From an empirical point of view we have shown that our *BodyMusic* application could be used by a number of users and demonstrated the initial utility of our proposed *off-the-wall* interaction system. Field trials conducted for the *gpsTunes* application show that users are able to traverse to unknown locations using a basic system and that they can also follow unknown trajectories or paths using this system. We also find that when users are performing the same task in a familiar environment with natural visual distracters, such as people or friends and natural constraints such as paths and buildings, that they perform significantly better than when asked to perform the same task in a wide open featureless playing field. A field trial using the *airMessages* system shows that users are able to probe a local density map or 'context' effectively, using the full functionality of the designed interface, to find messages and drop new messages with the use of gestural interaction. We find that users who utilise the degrees of freedom in the system most effectively are those who also complete the tasks fastest.

Chapter 2

Background and Challenges

2.1 Introduction

In the past, location was not something which had to be considered in interaction design, since most computers were extremely large, desktop-based and definitely not mobile. The Global Positioning System and to a lesser extent the MEMS implementation of inertial sensing have acted as a catalyst in the rise of location-aware mobile computing and we have recently seen an abundance of novel location-aware applications. These applications, though, are limited by the traditional metaphors for interaction with a mobile device, so for this reason there exists a movement towards the development of novel approaches to interaction in this new context. This thesis presents two contrasting location-aware applications, each of which takes a different approach to this location-aware label. So what tools are beneficial for the development of such applications and how should they be used? And how can the addition of inertial sensing influence and enrich this kind of interaction? These are both questions we wish to address in this chapter.

In the remainder of this chapter we will review existing work in this area, consider some of the challenges, introduce some useful tools and show the feasibility of building highly interactive location-aware systems on a mobile device.

2.2 Inertial Sensing

Movement-based interaction based on inertial sensing is still a relatively new paradigm for interacting with mobile devices. As with location-aware computing, in recent years there have been an ever increasing number of applications developed, which take advantage of inertial sensing in the mobile domain. Inertial sensing, though, is by no means a new area of research. For well over 50 years researchers have been developing sensing techniques and sensors for aircraft and military applications (Titterton and Weston 2004), yet research in the mobile devices community has largely failed to take advantage of the tools and algorithms developed in this time. One reason for this may be the undeniable lack of superficial similarity between a large passenger-carrying aircraft and your average handheld mobile device but the underlying principles of inertial sensing apply to both.

So what can inertial sensing bring to our mobile devices? The principal application of inertial sensing has been to the development of more natural and less obtrusive interaction techniques generally but also has a significant influence on context and location-aware systems. A number of commercial products now come equipped with inertial sensors. The Sam-

Figure 2.1: Nokia 5500 sports training phone, Samsung SCH-S310 gesture phone and Sony S2 sports walkman.

sung SCH-S310 comes with a built-in accelerometer used for simple gesture recognition, the Nokia 5500 mobile phone also has a built in accelerometer used for sports training, the iPod nano has a separate wireless sensor which fits into the shoe and is used for calculating run distances and calorie burning and the Sony S2 sports walkman utilises inertial sensing to choose playlists based on the runners current pace. It is envisioned that the use of multi-modal interfaces can expand the input bandwidth for interaction with mobile computers and aid interaction. Hinckley *et al.* (2000) described a mobile device instrumented with a proximity sensor and a two-axis tilt sensor. They demonstrate several new functionalities which make use of the sensors, such as recording memos when the device is held like a cell phone, switching between portrait and landscape display modes by holding the device in the desired orientation and scrolling the display using tilt demonstrating that inertial sensors can provide alternatives to the physical and on-screen buttons in handheld devices. These functionalities are now becoming more common in consumer electronics. There are a number of digital cameras which can orientate a picture depending on how the camera is held and more recently the Apple *iPhone* has been introduced, which can switch between portrait and landscape mode depending on the orientation of the device.

Many researchers have since focused on tilt-based inputs and audio and haptic outputs (Rekimoto 1996, Partridge *et al.* 2002, Wigdor and Balakrishnan 2003, Oakley *et al.* 2004, Hinckley *et al.* 2005) demonstrating the utility of one-handed control of a small screen device. The use of these systems whilst on the move has also been demonstrated by Pirhonen *et al.* (2002) and Crossan *et al.* (2005).

Examples of other kinds of applications include, for example, a dice, which employs inertial sensing and perceives movement and rolls to record what face it lands on. It is thus able to detect bias for unfair behaviour due to its physical imperfections (Laerhoven and Gellersen 2006). Other systems employ inertial sensing for movement based exercise, for example. Foody *et al.* (2006*b*) describe a project where they wish to develop an effective feedback system for a human interface to promote mental and physical exercise and relaxation via therapies such as Yoga or Tai Chi. They describe a prototype sourceless kinematic-feedback based video game, which utilises inertial sensing in the form of an Inertial Measurement Unit, to render and animate a skeleton on screen, which gives participants instructions on which posture to assume next. Development of new interaction techniques specifically designed for mobile scenarios and inertial sensing, it is thought, will help eradicate some of the limitations of current systems including the use of small hardware buttons and stylus keyboards, which can be cumbersome and difficult to use at times, especially when used ‘on the move’.

2.3 Our Sensors

2.3.1 Hardware

The equipment used in the course of this work consists of an HP iPAQ 5550 running WindowsCE equipped with a MESH – Modality Enhancing Sensor-pack for Handhelds (Oakley *et al.* 2004) Inertial Measurement Unit (IMU) backpack consisting of 3 Analog Devices ±2g dual-axis ADXL202JE accelerometers, 3 Analog Devices ±300deg/s single chip gyroscopes, 3 Honeywell devices HMC1053 magnetometers and a vibrotactile device. The main vibrotactile display is a modified VBW32 transducer, originally developed as an aid for hearing impaired people, which resonates at 250Hz and has a dynamic range of 54 dB. A standard orthogonal inertial sensor arrangement is used with the sensitive axis of the respective inertial sensors mounted coincident with the principle device axes providing us with direct measures of lateral accelerations, turn rates and magnetic field strength as well as the current GPS latitude and longitude. Our GPS is a Trimble Lassen Sq module, produced for mobile devices, and is also built-in as part of MESH (see figure 2.2). This module boasts a 9m resolution with up to 6m resolution around 50% of the time it is used (Trimble Navigation Ltd. 2002). It also provides us with velocity resolution of 0.06m/s and an 18m altitude resolution. This module suffers the same problems that most GPS modules suffer, in that there are occasional problems with resolution, latency, slow updates (1Hz update for this module), signal shadowing and noise in the signal, which can be detrimental to a system. It is for these reasons that systems like that described in chapter 4 require further support from other inertial sensors such as accelerometers, gyroscopes and magnetometers, which we have at our disposal with MESH. In our applications, apart from utilising the GPS for positioning, we have also used the accelerometers

Figure 2.2: Left: Mesh device alone and attached to an HP5550 Pocket Pc. Right: The MESH circuit board showing the main components related to the navigation task

to calculate pitch and roll, the magnetometers in conjunction with the accelerometers to achieve tilt-compensated heading and the vibrotactile unit to provide the user with appropriate feedback.

2.3.2 Other Hardware

There has been a significant amount of research conduced with various kinds of sensor pack in recent times. Tuulari and Ylisaukko-oja (2002) describe their sensor pack, *SoapBox* (Sensing, Operating and Activating Peripheral Box). Like MESH, the device contains both accelerometers for measuring the 3 dimensional acceleration or tilt of the device and a magnetic sensor for determining direction or heading. It also contains an Illumination sensor, which measures the intensity of visible light and an optical proximity sensor, which measures the level of reflection from RF pulses, allowing the device to calculate distances. Similarly, Foody *et al.* (2006*a*) have built a USB interfaced motion capture sensor, which contains 3-axis linear accelerometers and a 3-axis magnetometer. Researchers at Microsoft have developed SenseCam (Hodges *et al.* 2006), a sensor augmented wearable stills camera, which is designed to capture a digital record of the wearers day. This camera contains a number of sensors including an accelerometer for sensing movement, a microphone for sensing audio activity, a temperature sensor and a passive infrared sensor. *Smart-Its* (Holmquist *et al.* 2004) are small, self-contained, 'stick-on computers' that users can attach to objects, which were designed to aid researchers and designers in the construction of responsive or intelligent environments. The standard sensor board has five sensors including light, sound, pressure, acceleration, and temperature. Aylward and Paradiso (2006) describe a wireless sensor system for the capture of expressive motion when worn at the wrists and ankles of a dancer. Each sensor node includes a 6-axis inertial measurement unit comprised of three orthogonal gyroscopes and accelerometers, as well as a capacitive sensing to measure close range node-to-node proximity. The *WASP* project (Microsoft Research 2007) is a wearable platform, designed to enable the development of mobile and ubiquitous prototype applications that rely on sensing. Traditionally, wireless sensor networks have relied upon ad-hoc peer to peer networks but *wasp* uses a cellular communications infrastructure enabling a host of new applications beyond the environmental monitoring applications that are typical of sensor networks.

2.4 Mobile Signals

The data we receive from our sensors are reasonable measures of acceleration, angular rate and magnetic field, all with a specified variance and if the variance on these signals was small, we could simply work with the mean value from the sensor. For example, raw data from a mouse has very low variance and is very easy to work with directly but data from an accelerometer has a higher variance and we need to decide exactly what information we wish to infer from this data and treat it in an appropriate way. In appendix B we discuss exactly what we are receiving from our sensors but for now it is beneficial to examine some of the differing signals received from the sensors in different situations. In figure 2.3 we see data from an accelerometer, gyroscope and a magnetometer for the situation where a device is left on a table with no movement at all. This is basic raw data and contains no notable features although we see a downward trend in the accelerometer data, possibly related to a rise in temperature. If we examine the histograms in figure 2.4 we see that even in the case where a device is left untouched on a surface, there is an inherent uncertainty in the data received from the sensors, indicated by the spread of the distribution.

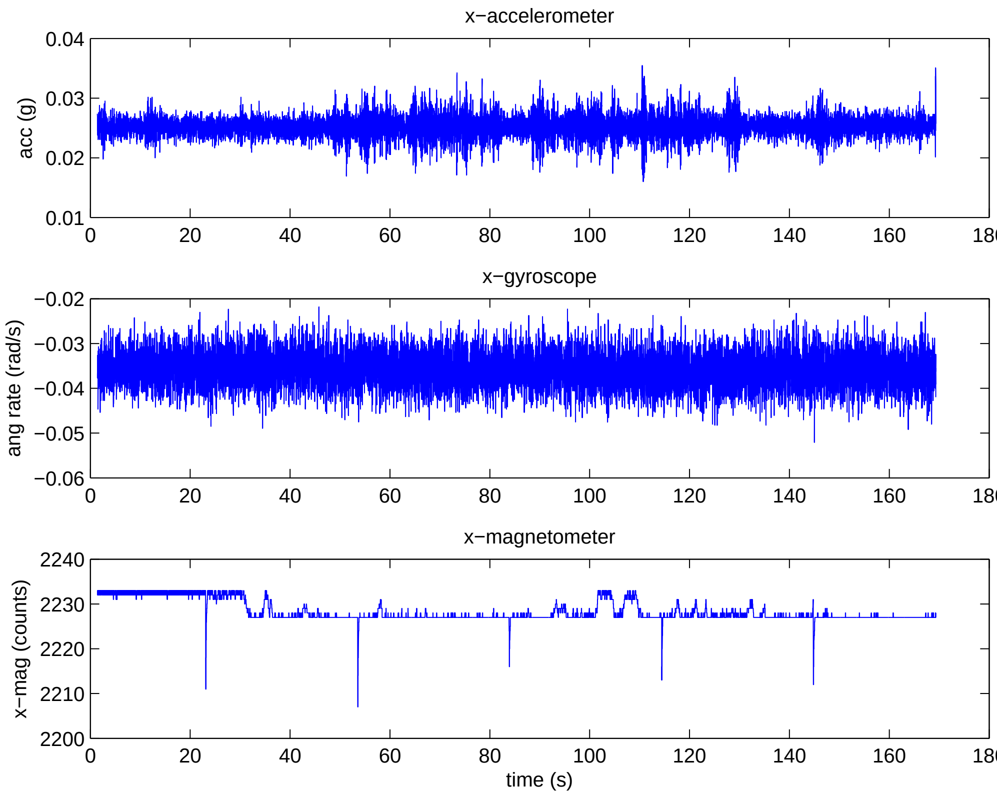


Figure 2.3: Unfiltered accelerometer, gyroscope and magnetometer data for a device motionless on a table.

We can easily observe the difference between the two types of movement in figure 2.5 where the device was placed in the user's pocket as they walked and in figure 2.6 where the device is held in hand and tilted first back and then forward. From figure 2.5 we may make the basic observations that there is a rhythmic structure in the data from the accelerometer and we observe that the magnetic field changes over time as the user walks further away from their starting position. And for figure 2.6 it is clear from all the sensor data where the tilting and stats and stops. Likewise for figure 2.7 it is clear to us from the accelerometer data that the device has been moved and from the gyroscope data that it has also been rotated somewhat. But how can we make more concrete observations from this kind of data? How can we infer the actual acceleration value or tilt of the device from the accelerometers? What is the current angular rate from the gyroscopes? And

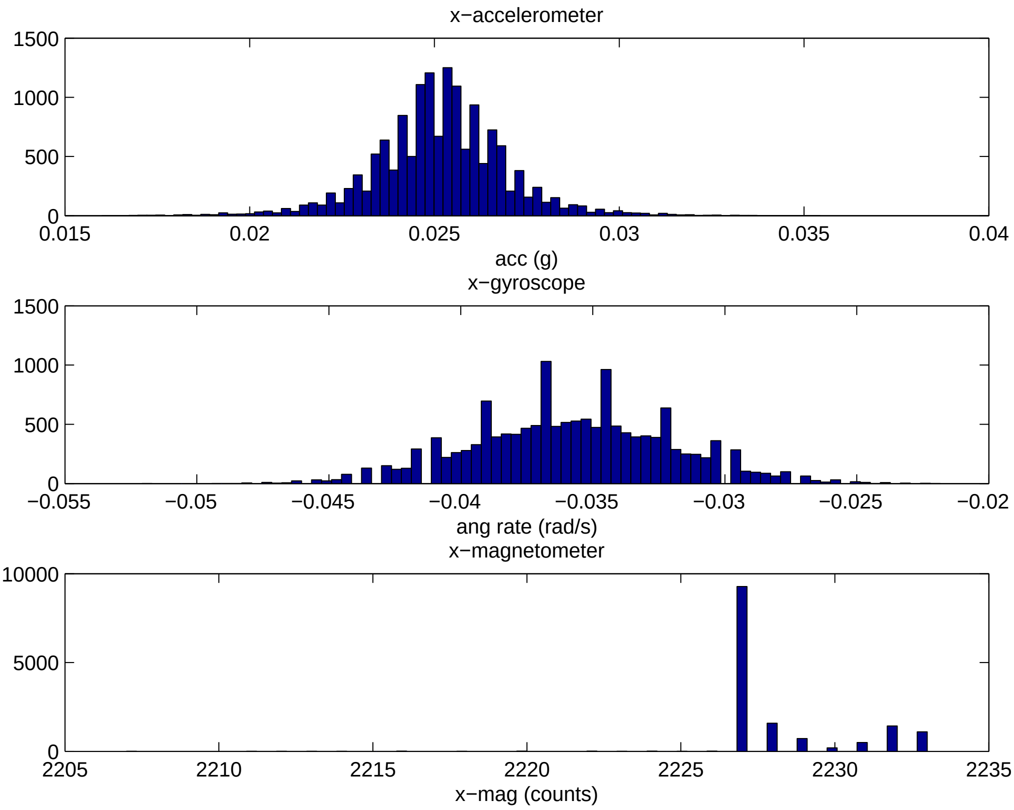


Figure 2.4: Histogram for the data shown in figure 2.3 providing us with a notion of the inherent uncertainty in measurements from these sensors.

what is our precise current heading from the magnetometers? In appendix B we answer these questions and demonstrate standard approaches to the inference of more precise values from these kind of sensors and apply this in a mobile computing context. For now we must consider the limitations placed upon us by the use of this kind of sensor.

2.5 Mobile Limitations

Although at a low level the tasks of determining the position and attitude of a missile and a mobile phone are the same, there do exist some obvious differences which may act to limit us when applying aspects of this mature field of research in our mobile computing context. For example, the sensors we are required to use are the cheapest possible, giving us noisier data than the data from the high-end sensors used in an aircraft. The mechanical gyroscopes used in aircraft applications have biases

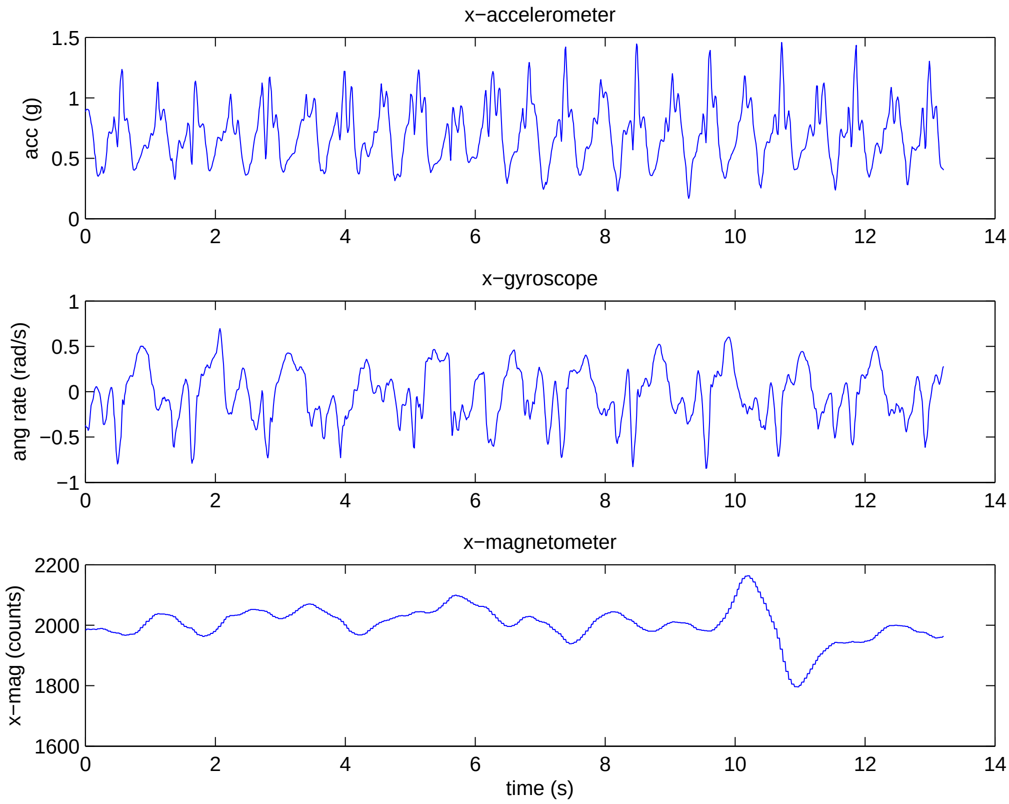


Figure 2.5: Unfiltered accelerometer, gyroscope and magnetometer data from sensors in a user's pocket while they are walking.

of $0.0002 - 0.002°/s/g$ compared to $0.01 - 0.05°/s/g$ in a typical vibrating mass gyroscope as is used in mobile devices (where s is seconds). Likewise for accelerometers, a typical mechanical accelerometer has a bias of $0.0001g - 0.01g$ compared to a value of $\sim 0.025g$ for a MEMS-based accelerometer.

As if to compound the problem, with a mobile device we will generally experience more variable movements and changes of context compared to the slow, smooth and relatively invariant rotations and constant context of a passenger carrying aircraft. For example, we may be walking down the street receiving high amplitude walking data and then enter a car and drive off receiving smoother driving data containing two-dimensional low frequency accelerations combined with high frequency vibrations from the vehicle. With a mobile device we are also subjected to the kind of disturbances that an aircraft will generally not need to deal with. For example, the output from magnetometers is very sensitive to local perturbations in

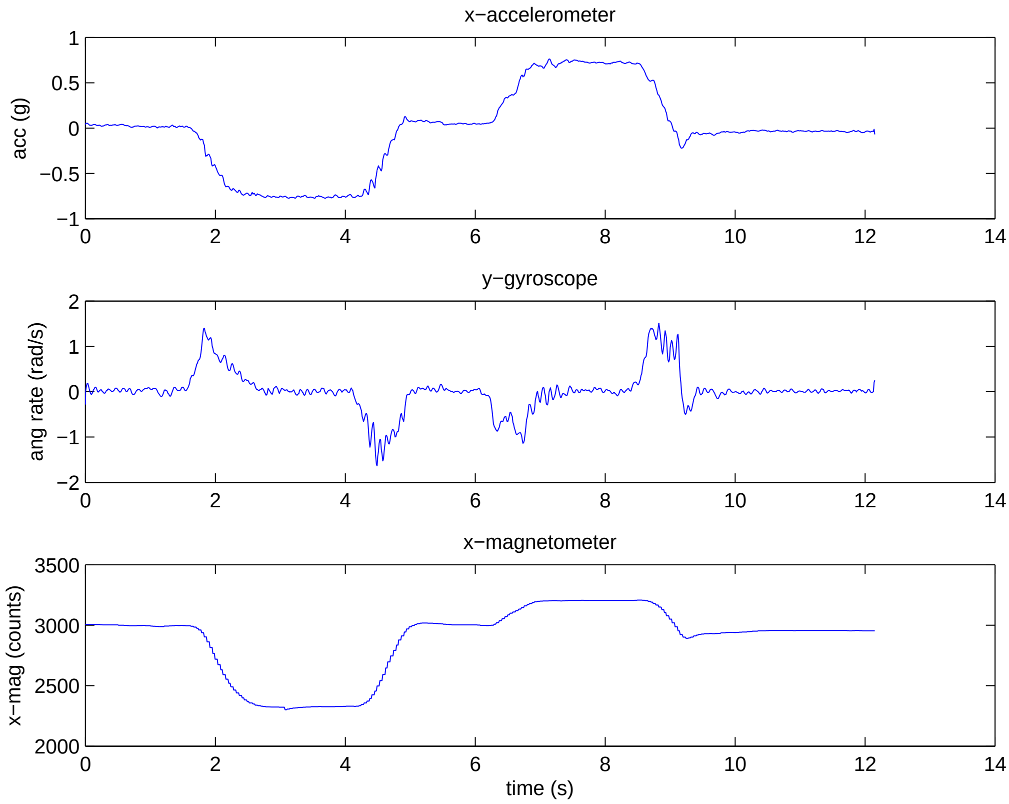


Figure 2.6: Unfiltered accelerometer, gyroscope and magnetometer data from sensors when they are tilted back and forward from horizontal.

the magnetic field from iron framed buildings, tables, filing cabinets, computers etc. that are not a problem for an aircraft traveling at 20000 ft.

It is necessary then that mobile applications are designed to deal with the inherent uncertainty and ambiguity we receive when using these sensors. For this task we are required to introduce a number of tools to aid the development of this interaction.

2.6 Control Theory

Control theory is a broad field which has undergone development for the best part of a century and there exists an enormous literature on control theory for engineering systems. Control theory is concerned with the analysis of closed-loop systems. In a closed-loop system (figure 2.9) feedback is a critical component. Put simply, a “controller” tries to manipulate the inputs of a system to control the desired output relative to the ‘reference

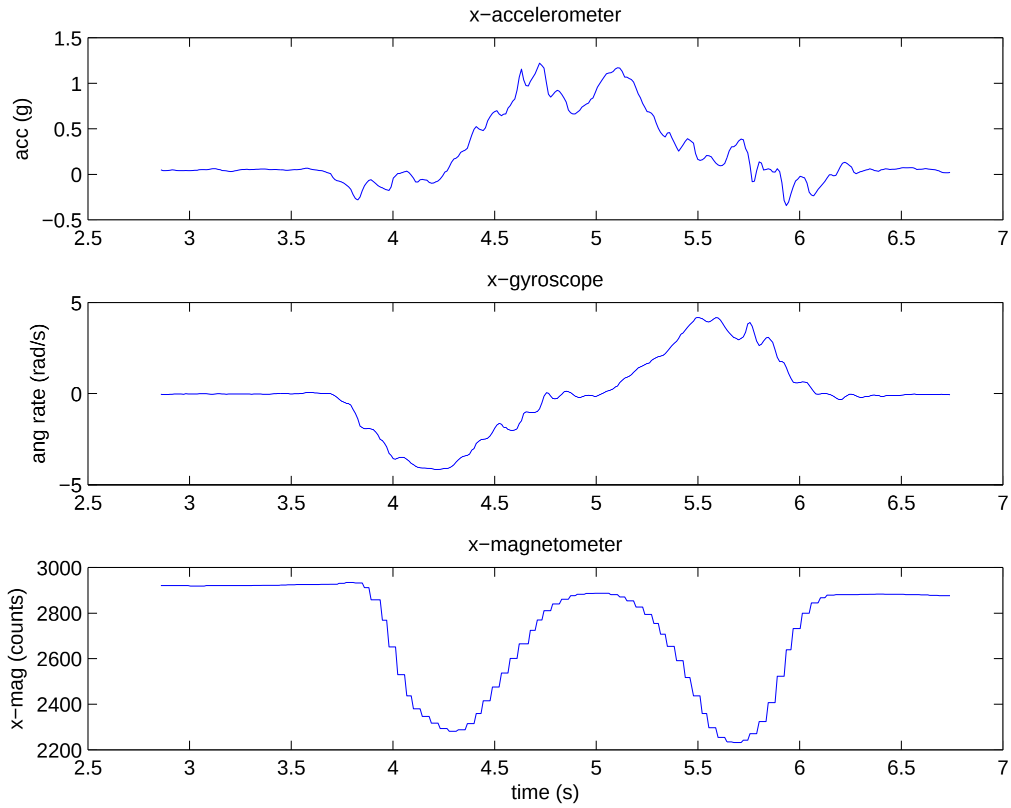


Figure 2.7: Unfiltered accelerometer, gyroscope and magnetometer data from a typical large arm gesture as displayed in figure 2.8.

variable' in order to realise some desired behaviour.

Control theory has traditionally been concerned with the automatic regulation of external systems. The earliest known 'control application' was the control of ancient water clocks around 2000 years ago but the first well documented application was a dynamics analysis of the centrifugal governor, which is a specific type of system that controls the speed of an engine by regulating the amount of steam admitted. Control theory then became an important part of military fire control and guidance applications (Baskett 2000, Blakelock 1991). More recently the use of control theory has become more relevant in other fields including sociology and economics, where, for example, optimal control techniques were used to influence policy involving unemployment-inflation tradeoffs (Athans and Kendrick 1973). Control theory is now also increasingly being applied to the field of Human Computer Interaction (Williamson 2006, Eslambolchilar *et al.* 2004, Eslambolchilar and Murray-Smith 2006, Eslambolchilar 2007).

Figure 2.8: A typical gesture performed to the back of the head.

2.6.1 Manual Control

Manual control is a sub-field of control theory that deals with the human control of dynamic systems. It was originally developed by feedback control engineers for military tasks involving humans, such as the control of aircraft and for tracking tasks or tracking for anti-aircraft gunners. Jagacinski and Flach (2003) gives a modern view of control theory for humans which highlights somewhat the potential application of standard control techniques to the field of Human-Computer Interaction while Kelley (1968), Sheridan and Ferrell (1974) and Poulton (1974) are classical examples of work in the field. There are a number of aspects of manual control theory which are relevant for the design of human-computer interfaces.

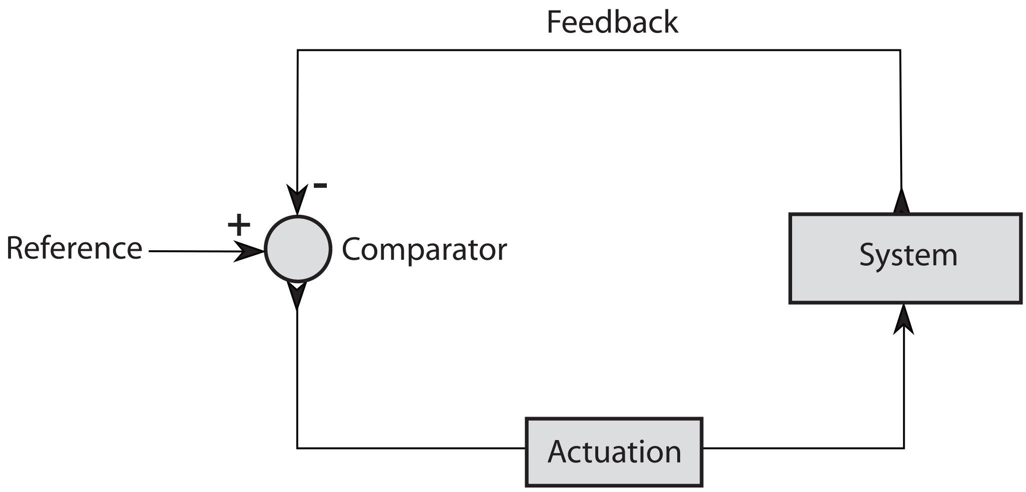


Figure 2.9: Negative feedback control loop with the error signal fedback to maintain control.

Discrete or Continuous Control

It is important that we understand the difference between discrete and continuous control. Schmidt and Lee (2005) describe discrete movements as having a definite beginning and end whereas continuous movements have no recognisable beginning or end and may continue until explicitly stopped. Examples of discrete movements include kicking a ball, turning a key or opening a door whereas continuous movements include swimming, running or steering a car and are typically oscillatory in nature.

Craik (1947) suggested that the output of the human operator performing a perceptual-motor control task consisted of a sequence of discrete, "ballistic" movements. Traditionally user interfaces on mobile devices have utilised a discrete control approach and conventional interfaces use button presses or clicks to navigate in discrete steps through some menu structure. Few interfaces use a continuous control. Schmidt and Lee (2005) describe continuous movements as those that have no recognisable beginning or end. And one important aspect of interaction with a mobile or wearable device especially is that it has the real potential to be continuous, with the user in constant, tightly coupled interaction with the system. Examples of this kind of interaction are given in (Williamson and Murray-Smith 2005*a*, Ward *et al.* 2000, Lantz and Murray-Smith 2004). A positive insight from control theory can enhance the development of these interfaces by providing a quantitative approach to interface design meaning that interaction need no longer consist of an exchange of discrete messages, but can form a rich and continuous dialogue.

Tracking

Poulton (1974), in his book describes in depth the field of tracking and manual control. Tracking is concerned with the execution of accurate movements at the correct time and can involve true motion or relative motion. Tracking is an everyday task for humans. We are constantly tracking with our eyes and hands. When we drive a car what we are doing, in a basic way, is tracking with a control system. So it is natural to extend the concept of tracking to our interface designs and apply some principles from this field to help us understand more the ways in which a user interacts with our interfaces. The aim in any tracking task is to minimise the error between some control object and a target. For example, following a path on screen with a cursor. This task may be performed in two ways. One is known as compensatory tracking; when the error is all that is available to the user and the second method is known as pursuit tracking; where both the target to follow and the current output are available. A pursuit display has an advantage over a compensatory display in that with the pursuit display it is possible to predict future movement of the track more accurately and so reduce any lag present. It is also possible with a pursuit display for the user to learn the control system more easily because the consequences of the user's actions are not compounded with changes in the reference signal (Poulton 1974).

2.7 Inferring Intention

One way in which we can use the well developed tools of control engineering is in the inference of user intention. Inferring the intention of a user is a difficult problem and we need to extract as much information from the *evidence space* (an augmentation of the sensor space which represents the current state of the system in terms of values inferred from sensors over some time window (Williamson 2006)) as we can. The evidence space encodes all of the information necessary to make a decision from our sensors. From a control perspective the intention of an interface can be thought of as the reference value to which it attempts to hold some other system. This implies that we may think of any device which engages in control as having intention. In a usability context we may think of a usable system as one which takes a user's actions and accurately translates them to some action or intention. The detection of user intention has been investigated by Powers (1973) who illustrated examples of intentional behaviour which could be empirically detected using continuous control models. He also demonstrated the extension of control from low level control processes to much higher level processes such as the control of a persons self image. Williamson and Murray-Smith (2004) describe an interface built on this principle and Williamson covers this in depth in his thesis (Williamson 2006). Utilising methods from perceptual control theory and dynamic systems, they present a method for performing selection tasks based on the continuous control of multiple, competing agents who attempt to determine the user's intentions from their control behaviour. Just by analysing the behaviour from a number of variables from some state vector x, user's are able to select their desired target without the use of an explicit visible pointer because the system has successfully determined the user's intention.

So what must we consider when attempting to ascertain the intention of a user? Signal uncertainty must be considered. Uncertain signals are omnipresent for inertial sensing in mobile computing and increase the width of the distribution of potential intentions so it is important that we treat uncertainty in an appropriate manner. 'Constraints' are also something we must consider. Real-world constraints, although not measured directly by our sensors, are something which influence the potential intentions of a user significantly and are something which can simplify the inference task significantly as we show in chapters 3 and 4.

2.7.1 Uncertainty

Our knowledge of the world is uncertain. As mentioned previously, the sensors we are attempting to use in this work are inherently uncertain.

But uncertainty is not just about uncertainty from our sensors, on a more fundamental level it is about uncertainty in the intentions of the user and uncertainty in the system's belief about the user's intention that is fedback to the user. When humans interact with their mobile device, there are a large number of variables which we may not observe directly. Uncertainty can arise from many sources. From sensor noise, from model inaccuracy, from the user's physical state when they are walking or sitting in a vehicle, from the their emotional state, from disturbances in the environment or from variability in the user's general performance of actions as illustrated in figure 2.10, which shows five different attempts at the same gesture from the same user. It may not be possible to directly associate any of these variables with a specific intention with respect to the interface but they do, as a whole, affect the interaction and the communication of intention. Another problem we face comes from the fact that we do not have direct access to what we really wish to measure. We wish to measure exactly how much the phone has been moved but what we get is a complex acceleration trace, so the way we interpret this trace is important. From a theoretical perspective, our system must assume a distribution over all potential states meaning that the system's observations are too abstracted from the actual intentions of the user. It is this abstraction which motivates the incorporation of uncertainty into our interfaces, in both inference and feedback, to aid the negotiation of interaction between user and computer.

2.7.2 Constraints

Part of our evidence space should include real-world constraints, which place limits on the number of possible intentions possessed by a user and the movements they can feasibly make in any given period. By thinking of the natural constraints placed on our system, which limit the range of potential user behaviour and afford us some extra information that our system may use to interpret these intentions, we may narrow down the number of potential goals for that user. So what kind of constraints must we consider? Concrete examples of the use of constraint are demonstrated in chapters 3 and 4. We can consider physical constraints from the world around us or the physiological and cognitive constraints of the user. Even social or environmental constraints can be considered depending on the user's current context. If a user is walking down the street it is more likely that they will wish to call someone or gesture with their device than if they are riding a bike. The potential range of user intentions when a user is riding a bicycle is far more limited because the bicycle *constrains* the potential number of intentions. Likewise, if the user is sitting on a bus the likelihood of them getting off of the bus at a designated stop is far higher than them getting off the bus at random place on the road. The number of potential places they wish to go in the real world, or the distribution of potential intentions, is *constrained* just by the fact that they are on a bus. It is this kind of real world inference that cannot be extracted directly from a stream of sensor data but which is important in the inference of user intention and it is important that this way of thinking is utilised to help us mould and shape our uncertain signals in a constructive manner.

Characteristic and repeated patterns of user behaviour may also be considered as constraints. Krumm and Horvitz (2006) present a system, *Predestination*, which uses a history of a drivers destinations, along with data about driving behaviours, to predict where a driver is going. As time progresses the possible number of potential destinations for that driver becomes more and more constrained. Other work which makes use of typical user behaviour to make predictions includes that of Ashbrook and Starner (2003) who find potential destinations by clustering GPS data, then predict destinations from a number of potential candidates, and Marmasse and Schmandt (2002) who predict a person's destination from a list of previously visited destinations. All these systems are, in one form or another, exploiting constraints in the user's behaviour, where behaviour is a result of constraints and desires, exploiting biases in the physical environment, cognitive constraints or even fatigue, to deduce a user's intention.

2.8 Feedback

Feedback is essential for the control of any system subject to uncertainty. If we do not receive feedback on our actions, in any context, we are unaware of the effect those actions had, we have an open-loop system. This is particularly true when it comes to interaction with our mobile devices. If we press a button on our mobile phone or PDA and nothing happens this can be very disconcerting. If your system takes fifteen seconds to boot and there is no progress bar this is also disconcerting, especially if you did not anticipate this delay. So it is important that appropriate feedback is designed to enhance and allow the user to engage in a positive relationship with their device. Feedback can come from a number of channels or modalities so we need to consider the correct choice of modality, or combination of modalities, when designing feedback for our system.

2.8.1 Multiple-Modalities

The most obvious feedback modality for humans is visual. Our eyes are our most dominant sense so it makes sense that visual feedback is used in most interfaces. Visual displays are the most common kind of display, from the earliest days of computing information was being displayed visually on a monitor or paper trail. Our visual channel has very high bandwidth and a vast amount of work over the years has gone into the use of visual feedback to convey a very wide range of information. But the visual channel can become inefficient in situations where a user is mobile, as they may be paying more attention at this point to their surroundings, such as passing cars, lamp posts and other pedestrians. For this reason we attempt to develop an 'eyes free' interaction, focussing principally on the auditory and haptic modalities. There are significant advantages to non-visual display in a mobile context since when mobile, a user is likely to allocate the majority of his visual attention to tasks not involving the device display.

The audio channel is the natural second choice for displaying feedback and is already used extensively for augmenting our visual modality; we frequently use our ears to tell our eyes where to look. The vast majority of visual displays also utilise some form of audio feedback to augment the visual data. The majority of auditory interfaces are limited to discrete sounds, emitted after certain events such as button presses or used as warning sounds. This kind of discrete summative feedback has been extensively investigated (Brewster *et al.* 1993, Brewster 1997, Gaver 1986) but there has been much less work conducted on continuous feedback in continuous control contexts. The use of continuous control allows the provision of more formative feedback. Poulton (1974) describes some very early experiments that were based principally on pitch modulation (e.g. (Milnes-Walker 1971)) or interruption rate (for example, for error display as in (Ellis *et al.* 1953)) and Williamson and Murray-Smith (2005*b*) describe a general framework for producing formative audio feedback for gesture recognition, where granular synthesis is used to present the audio display of the changing probabilities and observed states of the performed gesture. There are a number of issues with audio feedback, which we must consider. There are potential accessibility issues with depending solely on audio feedback in an application. Some users may not be able to hear well and others may be working or passing through a noisy environment. Audio feedback can also be considered annoying to users at times, especially when this involves, for example, altering or interrupting a user's personal choice of music. There can also exist issues with the discreetness of audio feedback and privacy issues may arise if anyone other than the user of the system can hear the feedback. In these situations it may be beneficial to consider the uses of haptic feedback. It is possible to use haptic feedback alone, which could potentially change the nature of the interaction but this may also be coupled to audio feedback. Linjama and Kaaresoja (2004) describe the implementation of gesture input supported by haptic feedback. They describe a bouncing ball game, where tapping the device in horizontal or vertical directions controls ball motion and Chang and O'Sullivan (2005) describe an evaluation of audio-haptic feedback on a mobile phone, comparing audio-based haptic user interface (UI) feedback with audio-only feedback and found that users were receptive to audio-haptic UI feedback. The results also suggest that the combined audio and haptic feedback seemed to enhance the perception of audio quality.

2.9 Sensor Fusion

The information we gather to help us with the inference of a user's intention comes from a number of distinct sources. So we need methods for combining these sources appropriately. We are utilising a number of different sensors, each with their own characteristic strengths and weaknesses. As was mentioned before, noise is also a significant problem with most sensors that can make the output unreliable at times, especially when used alone, making it difficult to draw any real meaning or intention from an individual sensor. The use of sensor fusion seeks to overcome some of the drawbacks associated with inertial sensing and is concerned with the synergistic use of multiple sources of information i.e. multiple sensors, to provide a more complete and more usable view of the device context. This is something we humans are very good at. We are constantly taking information from our multiple sensors (eyes, ears, nose etc.) and fusing this information to give ourselves a more complete view of the world. The human vestibular system is a good example of a natural sensor fusion. It is essential for stable posture control, taking inertial information from our ear canal to enable us to move freely. To keep track of our orientation in space, we constantly update our mental egocentric representation of our surroundings, matching it to our motion. Researchers have attempted to mimic this system in their work on self motion, finding that providing consistent cues about self-motion to multiple sensory modalities can enhance the perception of self-motion, even if physical motion cues are absent (Riecke *et al.* 2005).

Brooks and Iyengar (1998) divide sensor fusion into three categories. The first is *complementary sensor fusion*. Complementary sensors do not depend on each other directly but can be merged to form a more complete picture of the environment, for example, a set of radar stations converging non-overlapping geographic regions. Complementary fusion is easily implemented since no conflicting information is present. The second class is *competitive sensor fusion*. Competitive sensors each provide equivalent information about the environment. For example, a configuration with three identical radar units can tolerate the failure of one unit. The third class is *cooperative sensor fusion*. Cooperative sensors work together to drive information that neither sensor alone could provide. An example of cooperative sensing would be using two video cameras in stereo for 3D vision.

From our point of view we would wish to implement a cooperative sensor fusion system in our mobile devices since we have a number of different sensors which when used alone may not provide us with the information that we need but when combined with information from other sensors can provide us with a more complete picture of our environment. For example, it is not possible to use acceleration data alone to achieve an accurate value for velocity due to integration drift and sensor noise but fusing this information with the less frequent but more accurate velocity information from a GPS, using a standard sensor fusion algorithm, it may be possible to achieve accurate realtime velocity information.

Traditional approaches to sensor fusion are usually very rigorous. Accurate position and orientation may be achieved with a fully fused missile application. But is this possible or even necessary in a mobile device? The sensors used in a typical mobile IMU are cheaper, less accurate and more noisy than those used in classic large-scale applications. For these reasons it is necessary to 'settle for less' or accept imperfection in our mobile applications. It may be difficult to accurately determine an absolute position but it is possible to determine the device orientation with reasonable accuracy after suitable alignments and calibrations.

Probabilistic approaches, which we use in the course of this work are another approach to sensor fusion. Probabilistic approaches provide a natural way to handle uncertainty and errors in sensed data and can integrate new sensor data in a consistent way. So if we are using data from multiple sensors, one of which contains a lot of noise, the data from that sensor will be naturally downweighted by basic probabilistic inference, thus implicitly performing sensor fusion on that data.

2.10 Location-Aware Technologies

There are a number of different technologies available at this time for use in the world of location-aware computing. The United States funded Global Positioning System (GPS) is probably the most familiar technology to be used in the field. Thirty one satellites orbiting in geosynchronous orbits above the Earth are utilised in various ways, providing absolute position estimates to help car drivers find their way. They help to stop hikers getting lost in the mountains. They can even be used for fun (Chalmers *et al.* 2005). The introduction of *Galileo* in Europe will vastly improve the accuracy, integrity, reliability and availability of satellite navigation. The Galileo 'Open Service' will be free for anyone to access and signals will be broadcast in two bands. Receivers will achieve an accuracy of less than 4 m horizontally and 8 m vertically if they use both bands or less than 15 m horizontally and 35 m vertically if they utilise one band only. This is comparable to the current service with the GPS. Galileo will also provide an encrypted service, the 'Commercial Service', which will be available for a fee and will provide an accuracy of better than 1m. If the Commercial service is combined with data from ground stations, accuracy will be increased to less than 0.1m. This is expected to increase significantly the number of satellite location-based applications by 2010.

GSM cell references are another more lower resolution way of determining the position of your mobile computer. Enhanced 911 (or E911) was introduced by the Federal Communications Commission (FCC) in the United States and required that, by the end of December 2005, wireless phone providers develop a way to locate any phone which makes an emergency 911 call to within 150 meters 95% of the time. This has inevitably lead to a significant amount of research in this particular area of location-aware systems (Sayed *et al.* 2005, Gustafsson and Gunnarsson 2005, Patwari *et al.* 2005). The main method used to take a location value from these GSM references is triangulation. This is a simple and powerful method of local locationing. The method revolves around the solutions of a set of linear equations involving the coordinates of multiple reference points. With this method, given three or more reference points with known coordinates in range, a node can estimate its own position, limited only by the precision of distance measurements and the accuracy of the reference point measurements. This method can also be used to estimate locations from angles instead of distances, by using the sine and cosine rules for planar triangles. Wireless networks are also used for indoor locationing and commonly take advantage of triangulation. There has also been a significant amount of research into hybrid location-aware systems which take advantage of a mixture of GPS, GSM and wireless. Gwon *et al.* (2004) describe algorithms for estimating the location of stationary and mobile users based on heterogeneous indoor RF technologies. They propose two location algorithms, Selective Fusion Location Estimation (SELFLOC) and Region of Confidence (RoC), which can be used in conjunction with triangulation, or with third party commercial location estimation systems. Similarly, Randell and Muller (2001) describe a low-cost indoor positioning system which utilises a combination of radio frequency and ultrasonics. And *SpotON* (Hightower *et al.* 2001) is a system created to investigate flexible location sensor deployments in small-scale environments and uses Radio Signal Strength Information (RSSI) as a distance estimator to perform ad-hoc lateration.

Relative positioning can also be useful when there is no explicit infrastructure available to devices for absolute positioning. Occasionally there may be no access to GPS in places where the signal is jammed or occluded. There may also be no wireless infrastructure available. Certain indoor location systems are capable of providing fine-grained location and orientation information sufficient for relative positioning tasks (Addlesee *et al.* 2001, Priyantha *et al.* 2001, Patten *et al.* 2001). Čapkun *et al.* (2001) describe the problem of node positioning in mobile ad-hoc networks and propose a distributed, infrastructure-free positioning algorithm that does not rely on the Global Positioning System. Their algorithm uses the distances between the nodes to build a relative coordinate system in which the node positions are computed in two dimensions. Hazas *et al.* (2005) likewise describes a system, *Relate*, which provides fine-grained relative position information to co-located devices on the basis of peer-to-peer sensing. Kontkanen *et al.* (2004) describes a probabilistic approach to locationing in wireless radio networks. They demonstrate the usefulness of the a probabilistic modelling framework in solving location estimation (positioning) problem. They also discuss some of the links between positioning research done in the area of robotics and in the area of wireless radio networks. An example use of this kind of approach is given in (Hermersdorf *et al.* 2006), who show that it is possible to derive complex behavioral patterns and device location from collected Bluetooth data.

2.11 Location-Aware Applications

Perhaps the first implementation of a location-aware system was by Want *et al.* (1992) who implemented 'The Active Badge Location System'. The badge, which uses diffuse infrared signals to provide information about their location for a central computer, was worn by members of staff in an office setting and was used to modify the behaviour of programs running on near-by computers. At the time this system was implemented, mobile computing was new and GPS was not operational and the technologies we take for granted today such as cellular phone networks and wireless computing were not available. The *Cricket Location-Support System* (Priyantha *et al.* 2000) uses ultrasound emitters and receivers embedded in the object they wish to locate and the RADAR (Bahl and Padmanabhan 2000) system uses wireless networking technology to compute the 2D position of objects in a building.

Even low resolution location information can be used for practical purposes. For example, The ContextPhone (Raento *et al.* 2005) uses simple GSM cell references to infer information regarding a user's context. Drozd *et al.* (2006) describe a game for mobile phones, Hitchers, that uses cellular positioning. Players create digital hitch hikers, giving them names, destinations and questions to ask other players, and then drop them into their current phone cell. Players then search their current cell for hitchers, pick them up, answer their questions, carry them to new locations and drop them again.

The potential range of applications for location-aware computing is vast. 'Smart Dust' (Pister *et al.* 1999) is a location-aware project involving the combination microelectromechanical sensors with wireless communication into a one cubic millimeter sized package and may be spread out along remote roads or mountain ranges to determine, amongst other things, the velocity and direction of passing vehicles or animals.

2.12 Location-Aware Audio

Since we are interested primarily in developing 'eyes-free' interaction with our location-aware systems, it is necessary to review some of the previous work conducted in this area. Location-aware audio systems are not new and many standard GPS applications come with some form of audio feedback. Car navigation systems are the most common form of satellite navigation system in use today and the majority of them utilise at least some form of combined audio and visual feedback, with direct voice commands being the most popular mechanism for influencing drivers. This kind of speech based feedback is also popular in pedestrian based GPS applications, especially for visually-impaired users (Makino *et al.* 1996). And more recently Apple and Nike produced a system for the mass market, which utilises accelerometers in the user's shoe to keep track of distance and pace information, which is fed-back to the user via voice commands through their iPod, although there is no absolute location information recorded in this case.

There is also a significant amount of work conducted with the use of non-speech based audio cues. Loomis *et al.* (1998) describe an experiment where users are guided along a route of predefined way-points using their back-pack based system, developed for use by blind users, which uses spatialised audio (either speech or sound) from a virtual acoustic display in order to convey information about the surrounding virtual environment to the user. Their system uses information from a GPS combined with heading information from a fluxgate compass to achieve accurate location. They also have extensive GIS information on their local area including all buildings, roads, walkways, bikeways, trees and other details, which is used along with the heading and GPS information to relate the user to their surrounding environment. Their experiment was designed to determine whether spatialised audio from a virtual acoustic display resulted in better or worse route-following performance than verbal cues and they found that the virtual display mode fared best both in terms of guidance performance and user preference. Other work in the area of pedestrian navigation includes that of Holland *et al.* (2002) who describe their prototype spatial audio user interface, *AudioGPS*. Their interface is designed to allow mobile users to perform location tasks while their eyes, hands or general attention are otherwise engaged. They found with the use of spatial, non-speech audio and a prototype back-pack based system that very simple and computationally inexpensive spatial mappings are effective for helping users to find specific locations.

Using music as the mechanism for guiding users has also been previously investigated. Nemirovsky and Davenport (1999) describe *GuideShoes*, a shoe-based GPS navigation system, which consists of a pair of shoes, equipped with a GPS, wireless modem, MIDI synthesiser, CPU, and a base station used to perform all processing. They describe the use of *emons*, short musical "emotional cues", to guide a user to their desired location. Other work on music-based guidance includes our *gpsTunes* system (Strachan *et al.* 2005), where initial testing of a prototypical system had shown, with the use of a small field study, that it was possible to allow users to navigate in the real world, using a combined Audio/GPS player to aid navigation. Similarly, Etter (2005) describes a system known as *Melodious Walkabout*, which again utilises a user's music to guide them to their desired location. A study was conducted which concluded that it was possible, after users had gained some initial experience, to guide people by adapting their own music in a spatial way. Warren *et al.* (2005) have conducted a similar study with their *OnTrack* system both in a VRML simulation and in the real world. They show that it is possible to guide a user through a number of audio beacons to a desired location using continuously adapted music. Jones *et al* (Jones *et al.* 2006, Jones and Jones 2006) present a more complete system for audio trajectory following with a modulated music approach. Other work which utilises music as a tool for influencing a user in this mobile domain was conducted by Oliver and Flores-Mangas (2006) who constructed a system that takes advantage of the influence of music in exercise performance, enabling users to more easily achieve their exercise goals. It works by selecting music depending on a user's jogging speed and on their current heart rate. Likewise, (Elliott and Tomlinson 2006) describes a context aware music player, which makes real time choices of music based on user pace. Although these systems do not take into account the location of the user it is still a demonstration of the increasing convergence of mobile devices and audio, particularly music, for the emergence of a more embodied interaction with these devices.

Probabilistic approaches to the display of information in mobile GPS remain largely uninvestigated. We described an approach to location-aware audio computing in (Williamson *et al.* 2006). The *gpsTunes* system described uses a probabilistic approach to presenting auditory information about the user's current state. The display fully represents all estimated uncertainty in the prediction of which targets the user may be interested in, and where and how they are likely to move to them. This provides the interactor with clear, direct feedback when entropy (the measure of the "spread" of a distribution; see (MacKay 2003)) is low, and appropriately diffuse feedback as predicted entropy rises. An experiment was conducted which was designed to test the hypothesis that a probabilistic approach with an appropriate display of uncertainty could increase the usability and acceptance of mobile GPS systems, the results of which are described in chapter 4.

2.13 Gesture Recognition

Gesture Recognition refers to the area with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state, most commonly from the hand. The information contained in a typical gesture from the hand through time is far richer than the information provided by a general pointing task, the metaphor that still dominates desktop computing. Gestural approaches show much potential in the domain of mobile computing due to the lack of screen-space and the desire for eyes-free displays.

The learnability of gestures is one issue that exists with this kind of interface. How do users learn what is the correct gesture to perform and how can this be presented to the user? There is a natural large variation in the performance of gestures by humans. A large variation in a gesture from what a system was expecting can result in a misclassification and an annoyed or frustrated user, which can seriously affect the adoption of such systems. Appropriate feedback is one potential solution to aid users 'in gesture' but some form of guidance in the actual learning phase is essential. Kallio *et al.* (2006) present a visualisation method as an additional feature for accelerometer-based gesture control and illustrate how this could be utilised in providing essential feedback and act as a tutor for new users in gesture controlled applications.

The field of gesture recognition is broad and there exists well established gesture recognition techniques, each with their own strengths and weaknesses. Methods commonly employed in gesture recognition studies include simple template matching (Kramer and Leifer 1988), statistical analysis, neural networks, particle filtering or Hidden Markov Models. Template matching is well renowned to be the simplest method of recognising gestures but the pattern recognition approach has been the dominant method of creating gesture recognition systems. Thus far there have not been any real successful systems produced. The gesture functionality on some web browsers has been successful to a point, but these implement simplistic two-dimensional gestures performed with a mouse, which provides clean and precise data with a clear beginning and end to each gesture. Recently we have seen the introduction of commercial products, which incorporate gesture recognition such as the Samsung SCH-S310 mobile phone and the Nintendo Wii.

Hidden Markov Models's (HMM's) are one of the most popular methods used for temporal classification and have been particularly popular in the field of speech recognition. One of the most important advantages of HMMs is that they can be easily extended to deal with complex gesture recognition tasks. Another advantage of HMMs is that they remove details of the time evolution of the gesture while keeping the information about the trajectory which was formed. The condensation algorithm is also used for gesture recognition. It uses random sampling techniques to simply and elegantly search a multi-variate parameter space that is changing over time. The algorithm was proposed by Isard and Blake (1998) for tracking objects in clutter and has been extended to the field of gesture recognition. Neural Networks are another choice for use in gesture recognition since they afford the ability to derive meaning from complicated, imprecise or variable data.

The multilayer perceptron (MLP) is the simplest form of feedforward neural network (Bishop 1995). The MLP's simplicity makes it suited to the type of problem we are tackling in this work due to its desirable properties such as its compact parametric model making it suitable for low memory mobile devices.

2.14 Gesture Variability

Variability is an obvious problem when it comes to the recognition of gestures. Human beings are very good at distinguishing different kinds of gesture. We can easily recognise if a gesture performed by two different people is the same but it is highly unlikely that those two gestures will look the same when we come to examine the sensed data. Variability can come from a number of sources. General uncertainty is a contributing factor as well as the natural variations between humans. The two most common kinds of variability we experience from gestures are *spatial variability* and *temporal variability*, where spatial variability is the variation in the actual performed gesture and temporal variability is the variation in the timing of the performance of the gesture (Schmidt and Lee 2005). Both forms of variability are illustrated in figure 2.10. It is important that a recognition system is designed to deal with these kinds of variability. What we desire is a gesture recognition system which is able to estimate from this variable data exactly what gesture the user means to perform. Obviously, incorporating some degree of flexibility into a recognition system is no problem and potentially increases the usability of that system but in most cases this implies that we would need to accept a corresponding decrease in accuracy. Additionally, another big factor hampering the general acceptance of gesture recognition systems is the need to train the system for individual users, something which is exacerbated by this natural variation from person to person.

So how may we approach this problem? One way is to examine variability at different parts of a gesture. It is likely that some parts of a gesture will not vary much at all while at other parts the variability will be large. For example, in figure 2.10 we see that the end point of the gestures don't

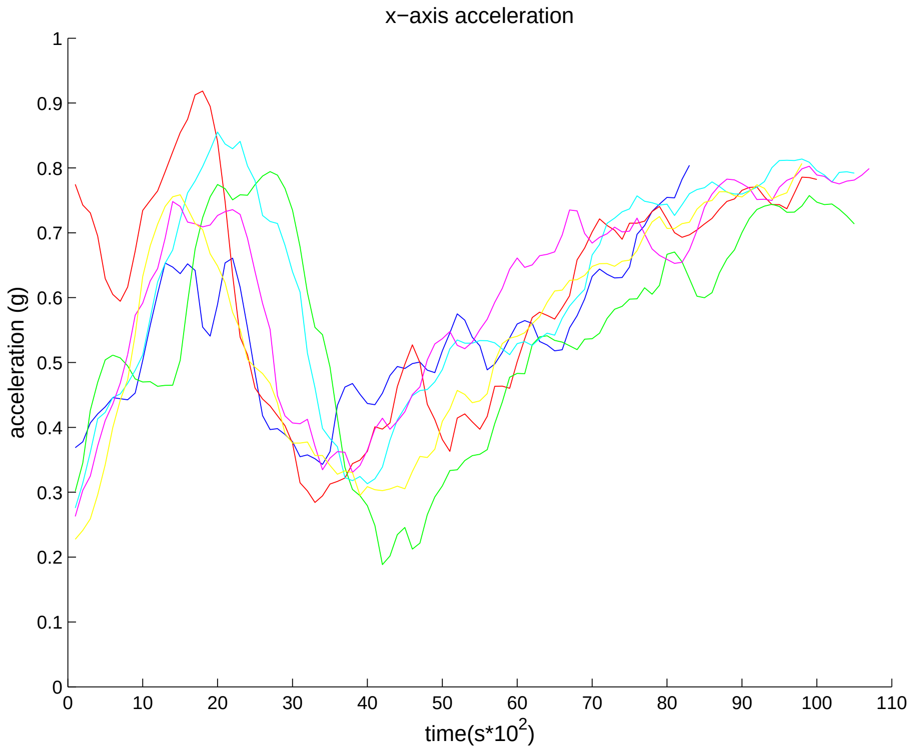


Figure 2.10: Unfiltered accelerometer data for the same gesture performed five times by the same person.

vary much at all whereas the beginning and middle parts vary substantially.

The measurement of variability also depends on the domain we are measuring in. A perceived difference in one domain may disappear in another. For example, if we look at the temporal acceleration signal for a device held motionless in the hand and the signal for a device laying motionless on a table, in the temporal domain the signals will look similar but if we examine both signals in the frequency domain we would see that one of the signals has information in the 8-12 Hz range from tremor in the muscles.

2.15 Gesture Controlled Applications

There has been significant work conducted on gesture controlled applications in the past. Some defining work on gesture recognition systems includes that of (Rubine 1991, Lipscomb 1991, Fels and Hinton 1990, Zimmerman *et al.* 1987). Some later work focussed on virtual reality systems and virtual environments using a camera in combination with an image recognition system (Kjeldsen and Kender 1996). The Opera web browser is the most well known application, which incorporates gesture recognition into the interaction to perform actions such as page reloading or moving back and forward a page with discrete mouse gesture. Inertial sensing though, has emerged as a viable technique for sensing movement in gestural interaction with mobile devices. Rekimoto (2001) describes his GestureWrist system which is a wrist band that recognises hand and forearm movements and uses these movements to communicate with a computer. Perng *et al.* (1999) describe a glove augmented with 2-axis accelerometers on each finger and one on the back of the hand which makes it possible to detect the angle of each finger and hence to detect static hand gestures. Using this approach they developed software that allows the glove to be used as a mouse pointing device. Ubi-finger (Tsukada and Yasamura 2002) is a system which uses acceleration and touch sensors to detect a fixed set of hand gestures. Using this system they may select target appliances by simply pointing at the device then control this with simple finger gestures. Kela *et al.* (2006) describe the use of a matchbox sized sensor pack, SoapBox, described in section 2.3.2, which they use to control the functionality of different appliances in their design studio. They describe a study which aimed to find the most natural types of gesture for controlling different appliances, such as a VCR. They also describe a study designed to compare the usefulness of the gesture modality compared to other modalities for control such as RFID objects or PDA and stylus finding that gestures are a natural modality for certain tasks and can help to augment other modalities. This reflects the conclusions of Pirhonen *et al.* (2002) who previously investigated the use of gesture and non-speech based audio as a way to improve the interface on a mobile music player. The key advantage of this gestural approach is that it enables eyes-free interaction with a music player which is advantageous, especially when the user is ‘on the move’. Cho *et al.* (2004) present a gesture input device known as the *Magic Wand*, equipped with inertial sensors, which enables a user to perform gestures in 3-D space. They employ a trajectory estimation algorithm to convert the gestures into a trajectory on 2-D plane and use a recognition algorithm based on Bayesian networks to achieve a recognition rate of 92.2%. Choi *et al.* (2005) describe a gesture-based interaction using a tri-axis accelerometer on a mobile phone (the Samsung SCH-S310). The mobile phone recognises digits from 1 to 9 and five symbols written in the air and uses a shaking motion to initiate commands for interactive games and musical instrument applications. Williamson *et al.* (2007) describe a completely eyes-free system, *Shoogle*, which uses inertial sensing and is used single handed for sensing data within a mobile device, such as the presence of text messages, via the use of simple 'shaking' gestures, which reveal the contents rattling around "inside" the device. Cho *et al.* (2007) describe a tilt controlled photo browsing method for small screen mobile devices. They describe their implementation on a mobile phone and an interaction based on a simple physical model, with its characteristics shaped to enhance usability. Informally comparing their photo browsing system to that on an iPod, they found that the tilt based approach performed slightly better.

2.16 Social Issues

The social acceptability of such systems is important and must be considered at the design stage. It is generally accepted that input devices should be as discrete and natural as possible and this has been a significant problem with previous gesture-based systems in that they were considered too obtrusive or too obvious. Costanza *et al.* (2005) discuss the use of more subtle interaction with 'Intimate Interfaces' and they argue that the use of a mobile device in a social context should not cause embarrassment and disruption to the immediate environment. It should be noted though, that 'subtle' should not necessarily mean 'no movement' and the use of subtle movements is something essential for the general acceptance of gestural interfaces. The recent introduction of the Nintendo Wii looks set to bring the concept of using gesture for interaction to the mass market, eliminating some of the social inhibitions which affected the use of gesture previously and making the use of gesture in public more socially acceptable. This should hopefully aid the development of more gesture-based applications in the near future.

Chapter 3

Bodyspace

3.1 Summary

This chapter demonstrates the construction of a novel egocentric location-aware system and presents a new approach to the segmentation and recognition of gestures. A model-based approach to this kind of interaction is demonstrated and it is shown that this kind of approach to interaction can enable the easy provision and adjustment of feedback. A small user study is conducted, which shows that this model based approach to interaction can be both intuitive and can be learned quickly. The use of real world constraints is demonstrated and an example is provided which shows that this may be used for inferring user intention.

3.2 Introduction

The Body Mnemonics project (Ängeslevä *et al.* 2003*a*, Ängeslevä *et al.* 2003*b*) developed a new concept in interaction design. Essentially, it explored the idea of allowing users to store and retrieve information and computational functionality on different parts of their body as illustrated in figure 3.1. In this design, information can be stored and subsequently accessed by moving a handheld device to different locations around the body. Moving the device to the back pocket, for example, may open a user's personal finances application on the mobile device. From a technical point of view we see this as a gesture to the back pocket but the user may think of the back pocket as the actual physical location of their personal finances.The *method of loci* is a technique for remembering, which has been

Figure 3.1: Examples of what we may store on different parts of the body

practiced since the time of the ancient Greeks when orators would use the method to help them memorise long narratives. The 'loci' were physical locations, usually in a large public area or building, such as a market place or a church. Practicing this method involved walking through this familiar place a number of times, viewing distinct places in the same order each time. After this was repeated a number of times, it was possible to remember and visualise each of the places in order reliably. This physical space within a room was used as a mental model and different parts of their narrative would be placed into the loci where they could be recalled in order by imagining the same route through the building, visiting each of the loci. In medieval Europe the method was adapted to include the space around a persons body or their 'body space'. Different body positions were used as markers to remember chants, lists or even as a computational system.

Previous work on this concept focussed mainly on the basic ideas and requirements for the project without a working, mobile implementation. Ängeslevä conducted surveys and interviews and found from potential users that:

> *the body view is a very personal artefact and that it is rich in meaning. It therefore has the potential to serve as a powerful memory aid.*

In this chapter we describe the first implementation of a completely handheld and fully functioning 'BodySpace' system which utilises inertial and magnetic sensing to recognise when it is placed at different areas of a user's body, to control a music player, essentially using the human body as the mnemonic device. A user may place the device at their hip to control the volume of their current song or at their ear to switch tracks, as illustrated in figure 3.15. They may also use their right shoulder to browse through a set of playlists and their left shoulder to start and stop the current track. The system uses a combination of pattern recognition and orientation sensing in order to recognise that the device is placed at the different parts of the body.

Our system differs from other gesture controlled systems in that we are not required to explicitly design a lexicon of gestures. The range of gestures we use is constrained by the limits of the human body, in that the arm can only move to certain locations around the body and for comfortable movements, has a constrained range of velocities. And since we are required to gesture to certain parts of the body we already have an obvious, perfectly natural and easily generated set of gestures at our disposal. Another difference is that we do not use any buttons at all in our interface, making the interaction more fluid and natural than a system which requires an explicit button press or release at the beginning and end of each gesture. This opens up the opportunity for the design of potentially tiny devices as the need for a set of buttons is removed. Additionally, we use a model-based approach to our interaction design which provides us with a real physical basis for the interaction and allows us to alter the interaction simply by varying the parameters of our model. Many wearable computers involve some kind of extra equipment which can be detrimental to normal social interactions. We feel that the inclusion of our system in a normal mobile device with natural gestures to different parts of the body is a step in the correct direction towards the acceptability of gesture based systems.

3.3 Gesture Recognition and Segmentation

As mentioned in chapter 2, one of the major challenges for any continuously sensing system is how do we know when to activate our system? How do we detect what is meaningful? And how do we detect user intention? If we construct some state vector x, which represents the state of our system and analyse the elements it should be possible to detect our user's intention, i.e. detect when the user intends the system to activate when it is placed at a specific part of the body. The state vector should contain any information relevant to the action to be inferred. In our case we can use information from any of our sensors. The state can contain information from the accelerometers or from the gyroscopes, which allow us to monitor the general movement of the device, be that rotation or larger translational movements. We may even include tremor from our muscles, which is observed in accelerometer signals while we hold the device (Strachan and Murray-Smith 2004). For gestural interaction the main inference we are required to make is to segment the gestures, i.e. where does one gesture end and another begin (referred to as 'the segmentation problem'), and how certain are we about that gesture.

One of the main problems with any gesture recognition system is the segmentation problem. One popular approach is to this problem is to use a hardware button to delineate the ends of the gesture, or use long periods of inactivity between gestures. This is obviously not desirable as it conflicts with the desired natural, free and easy interaction we wish to produce and this is one of the main reasons for the failure of gesture recognition systems to date. Harling and Edwards (1997), for example, describe a method for segmenting gestures using hand tension in their work with the recognition of sign language but obviously another more appropriate approach is required for use with our mobile devices.

3.3.1 Our Approach

Our approach to the recognition of when the device is placed at different body parts is a two stage process. The first stage involves identifying if the device *may be* at a certain part of the body, which we refer to as the *Segmentation Stage* and the second stage involves checking back through recent accelerometer data and classifying this using a simple multi-layer perceptron (Bishop 1995), the *Recognition Stage*.

Segmentation

One of the aims of this project is to avoid the use of explicit button presses in our gestural interfaces. In previous gesture recognition systems button presses have been used to segment or separate one gesture from another but this can have a detrimental effect on the system since the button press itself may affect the actual gesture being performed and can interrupt the natural fluidity and the desired free and easy interaction with the system.

There are two main ways in which this problem can be approached. The first approach (which we utilised in (Strachan *et al.* September 13-16, 2004)) would be to set some initial condition for the start of every gesture, which, for example, could be at the hip or stomach, and work from there to the end of the gesture, classifying the transient accelerometer data on the way. But this has problems in that users may not wish to be constrained to this initial starting condition at the start of every gesture and segmentation of the data becomes problematic since we are never sure of exactly where the gesture should end. The second approach, which is novel to this field, is to work on goal state/end-point identification and subsequently classify the accelerometer data prior to that. Finally, we chose the second approach since it aided the construction of a reliable gesture segmentation scheme, allowing us to differentiate between when a user has performed a gesture and when they were simply moving the device around in a general way. Another benefit of this approach is that it allows the rapid generation of data, which is beneficial for training purposes. This approach does have its limitations though, since it can not support formative feedback during (i.e. feedback at every part of the gesture as opposed to summative feedback, which only delivers feedback at the end of the gesture) the gesture, but since we are constrained to gesturing around the body our gestures are small and well defined enough that users are already naturally familiar with these kinds of gesture, reducing the need for a more formative 'in gesture' feedback.

We chose, in this set-up, to represent the state of our device in a simple way using its orientation, represented by pitch and roll pairs observed as it was placed at different parts of the body. By examining pitch and roll pairs from histograms of previously recorded end-point data, deduced from accelerometer measurements, is possible to signal that the device may potentially be placed at a certain part of the body, as illustrated in figures 3.2 and 3.3 allowing us to move into the second stage of recognition to confirm that this potential location is the true location of the device.

We formalise the problem by first stating that the prior probability of our orientation training set, ω, which holds our values for pitch θ and roll ϕ, originating from an area of the body (or class) C_b is $P(C_b)$. This corresponds to the fraction of our samples in each class, in the limit of an infinite number of observations with an initial assumption that each class contains the same fraction of samples. When we make a new observation, ω, we may assume that this observation belongs to one of a discrete set of values Ω^l corresponding to one of the bins in a histogram from our generative training set as illustrated in figures 3.2 and 3.3. The joint probability $P(\Omega^l, C_b)$ is defined as the probability that the observation ω belongs to a body area, C_b. The conditional probability $P(\Omega^l|C_b)$ specifies the probability that the observation falls in bin Ω^l of our histogram given that it belongs to class C_b. In other words, it specifies the fraction of our observations which fall into bin Ω^l for the histogram representing C_b. We may now see that the fraction of the total number of observations over all classes, which fall into bin Ω^l on histogram C_b is given by the fraction of the number of observations in

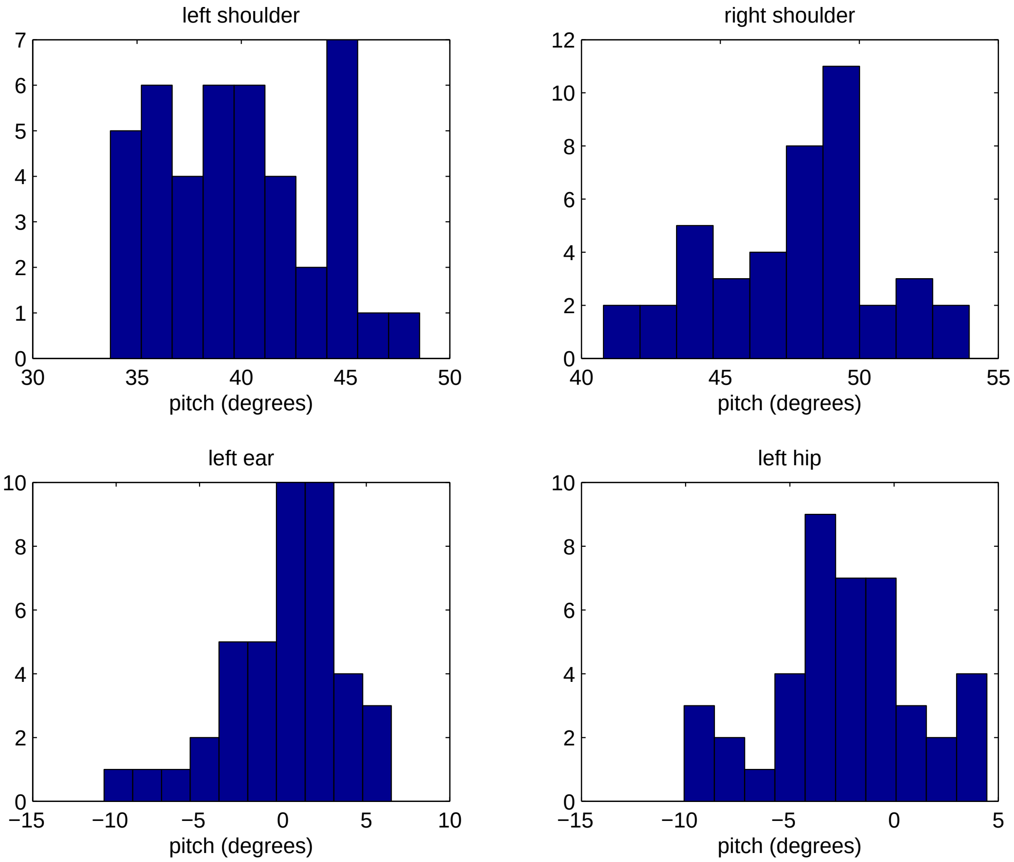


Figure 3.2: Histograms of pitch data, θ, for four positions around the body (42 samples)

histogram C_b which fall in bin Ω^l multiplied by the prior probability for that histogram $P(C_b)$ which is equivalent to

$$P(C_b, \Omega^l) = P(\Omega^l|C_b)P(C_b) = P(C_b|\Omega^l)P(\Omega^l), \tag{3.1}$$

where $P(C_b|\Omega^l)$ is the probability that the class is C_b given that the observation falls in the bin Ω^l and $P(\Omega^l)$ is the probability of observing a value from Ω^l with respect to the whole data set, irrespective of class, and is therefore given by the fraction of the total number of observations which fall into bin Ω^l over all classes. If we equate the last two expressions we produce Bayes' theorem which gives us a measure for the posterior probability that our observation belongs to a specific class given the prior probability $P(C_b)$ of our observation belonging to that class and the class conditional probability $P(\Omega^l|C_b)$.

$$P(C_b|\Omega^l) = \frac{P(\Omega^l|C_b)P(C_b)}{P(\Omega^l)}, \tag{3.2}$$

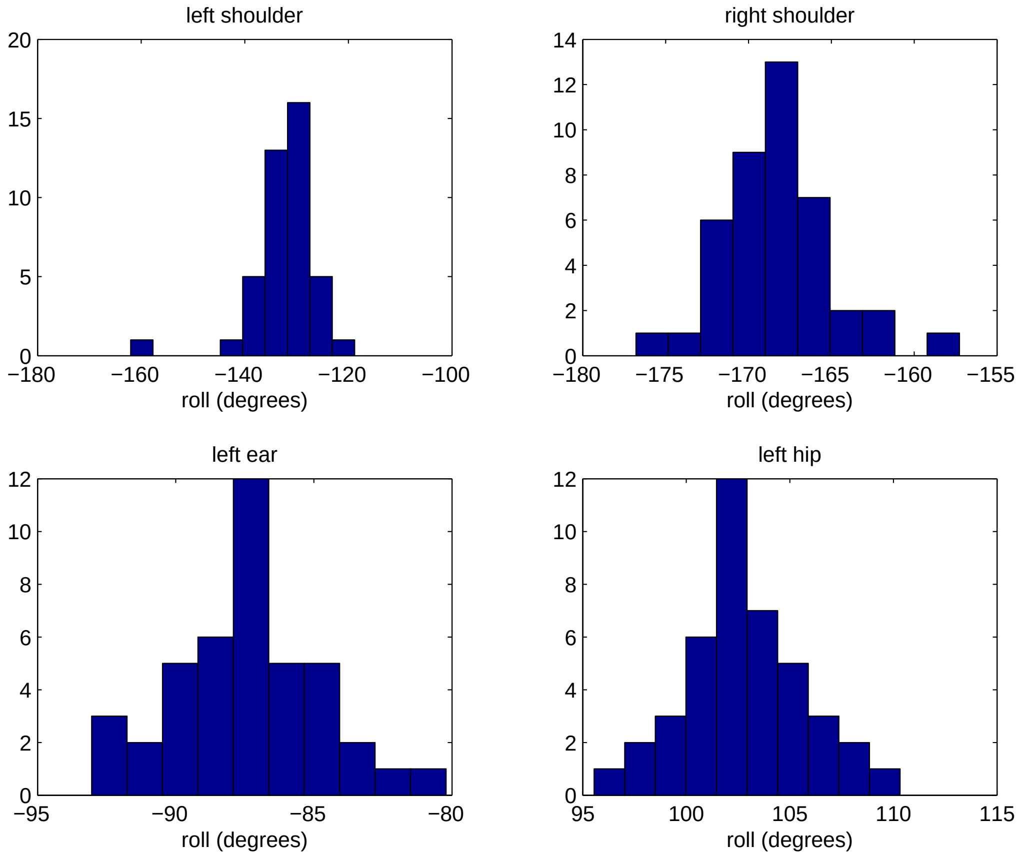


Figure 3.3: Histograms of roll data, ϕ, for four positions around the body (42 samples)

where $P(C_b|\Omega^l)$ is the class conditional probability that an observation Ω (containing values for the pitch θ and roll ϕ) belongs to an area of the body or class C_b, $P(\Omega^l|C_b)$ specifies the probability that the observation Ω originates from the ellipse corresponding to class C_b in figure 3.4, $P(C_b)$ is the prior probability of our orientation training set, Ω, originating from class C_b and $P(\Omega)$ is the probability of observing a value Ω with respect to the whole data set, irrespective of class (Bishop 1995).

Recognition

The recognition stage of classification first waits for a notification of a potential classification from the segmentation stage. If a possible measured orientation falls into one of the ellipses in figure 3.4, accelerometer data for the last second of motion is taken and classified using a simple Multi-Layer Perceptron to classify one of four body positions, left shoulder, right

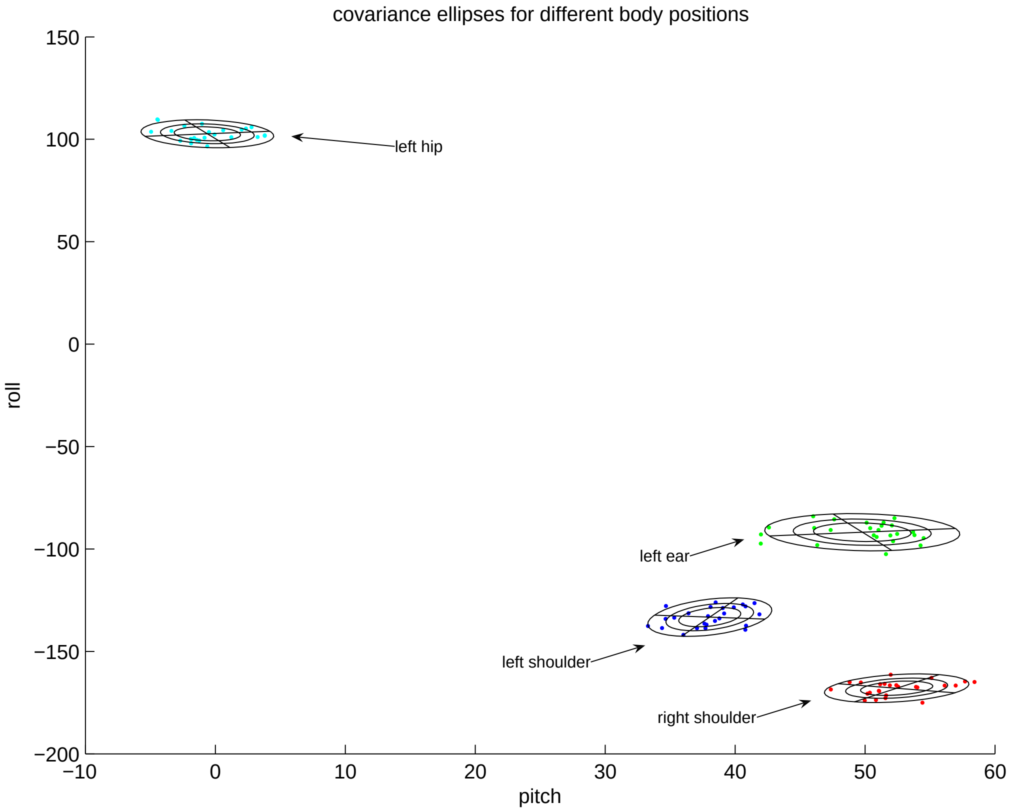


Figure 3.4: the four covariance ellipses corresponding to each part of the body where data was measured. Cyan - Hip. Green - Left Ear. Red - Right Shoulder. Blue - Left Shoulder

shoulder, left ear and left hip.

The system uses raw accelerometer data for classification at this stage and classifies this using a simple Multi-Layer Perceptron. The use of a Multi-Layer Perceptron shows the generality of the approach and its compact final form makes it suitable for low memory mobile devices. The fact that the accelerometer data for each kind of gesture used in this configuration is distinct enough to be classified on its own without any pre-processing is also a good reason to keep the pattern recognition mechanism simple although it is possible in future applications to use more advanced methods of recognition as more parts of the body are ‘added’ to the system making the recognition task more complex. Example gesture data for movement to each part of the body classified in this configuration are shown in figure 3.5.

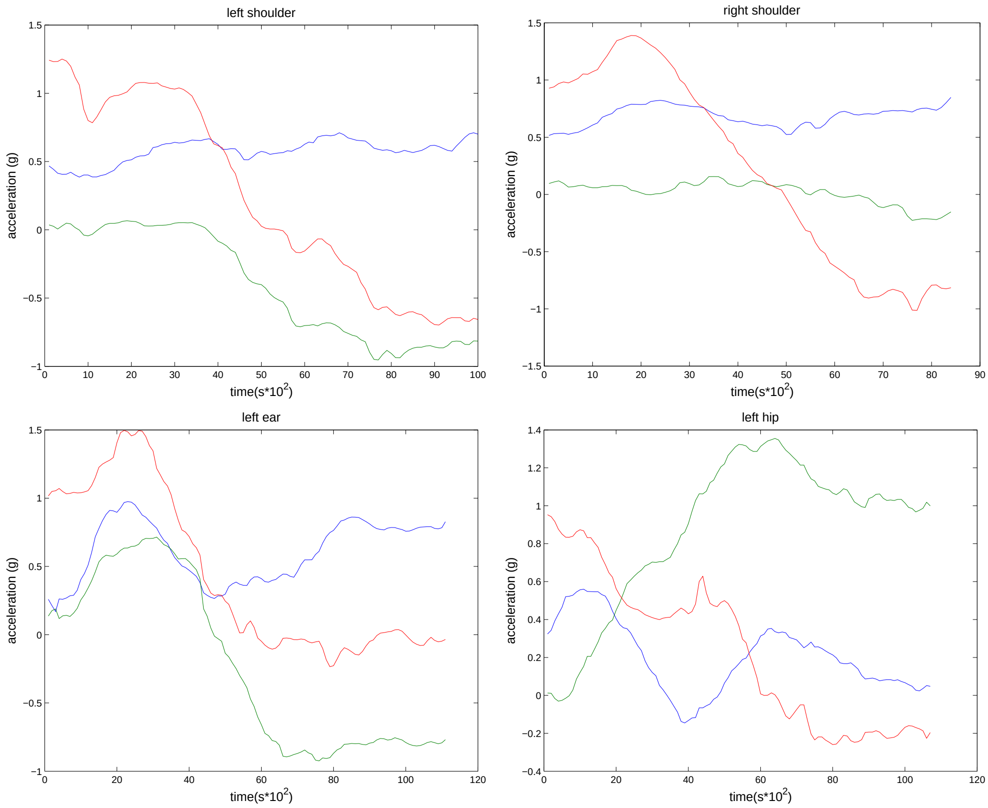


Figure 3.5: Example three-dimensional accelerometer data for movements to the four parts of the body we wish to recognise.

Training

Training of this system involves repeated gestures to selected parts of the body which we wish to classify as they are requested by the system via voice commands. Three gestures per location were required to achieve adequate training in this configuration and data from the x and y accelerometers only was used. The decision was made to leave out the z accelerometer data as previous experimentation with gesture data showed this to be the least useful for these kind of gestures and also to limit the amount of processing required to be performed by the pocket PC. Example data used to train the system for three different users is shown in figures 3.6 to 3.7.

The training set-up in this configuration is basic, although it was sufficient for the recognition of four locations on the body and performed well in initial trials. The addition of a 'noise' class could facilitate the elimination of false positives and an investigation into the best configuration of input data to use could facilitate the expansion of the system to include more recognisable body locations.

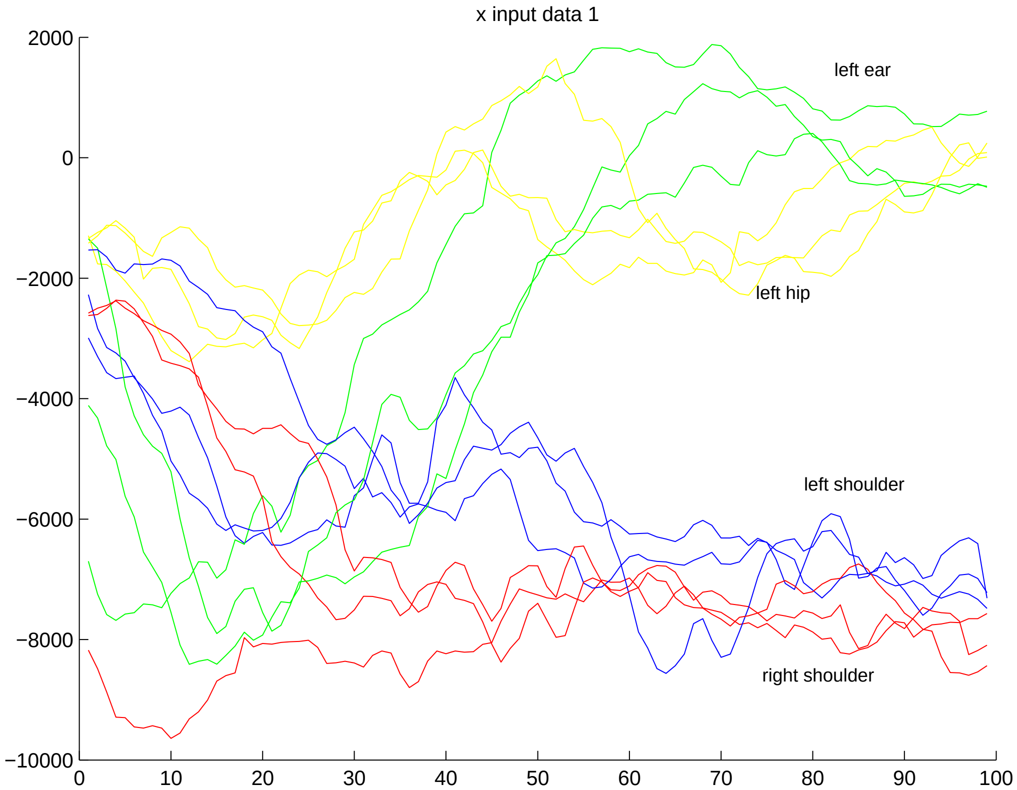


Figure 3.6: The x accelerometer input data used to train the MLP for user 1. There are three examples for each of the four different classes of gesture. This data shows the last 100 samples logged from the end point of the gesture, sampled at 100Hz.

3.4 Utilising Constraints

As mentioned in chapter 2 one of the fundamental problems affecting the development and general acceptance of gestural interfaces comes from the omnipresent uncertainty in our sensor measurements and general behaviour with our mobile devices. So the accurate interpretation of a user's intention from these noisy observations can be a difficult problem.

We may take a step forward by thinking of the natural constraints placed on our system which limit in some way, the range of potential user intentions and provide us with some extra information, which our system may utilise in the interpretation of these intentions. In our body-based application, one constraint on the potential range of intention is the physical limits of the user since the range of potential gestures which may be performed is constrained by the human body in normal and comfortable use. For

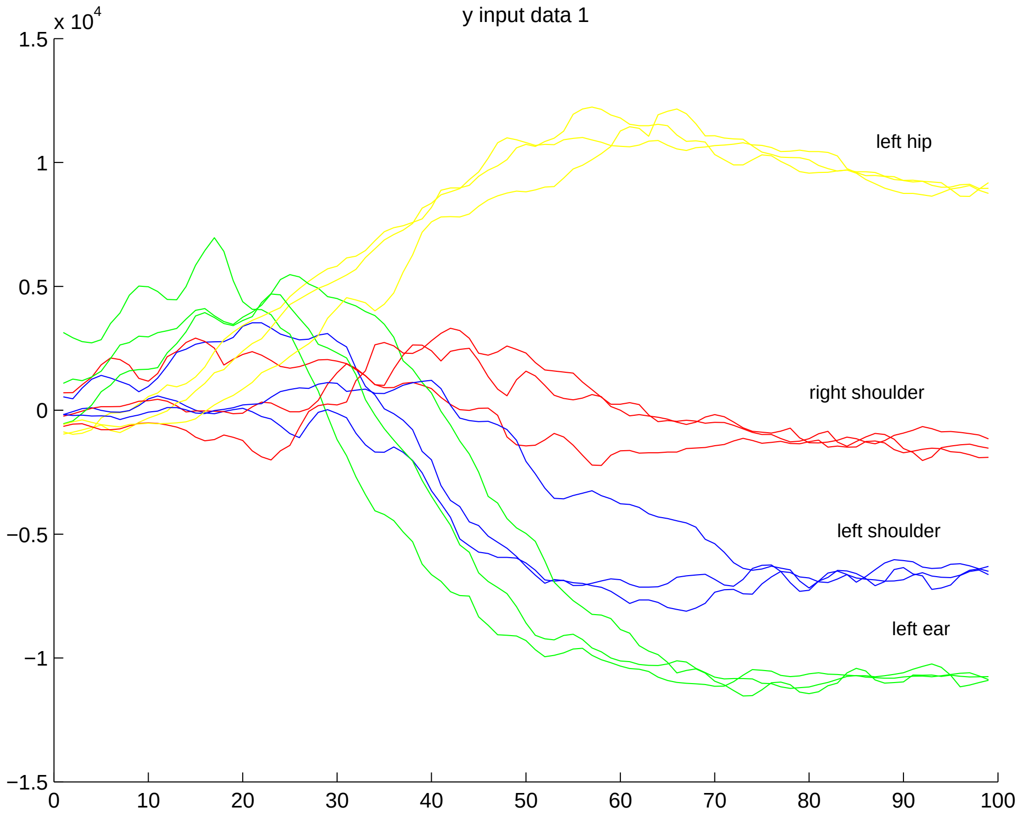
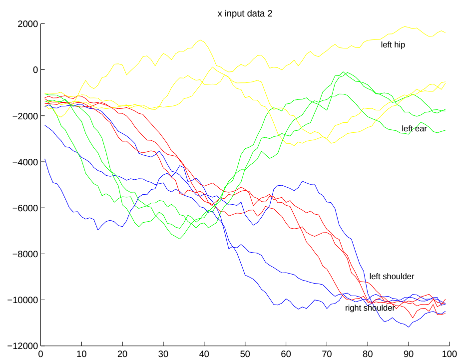
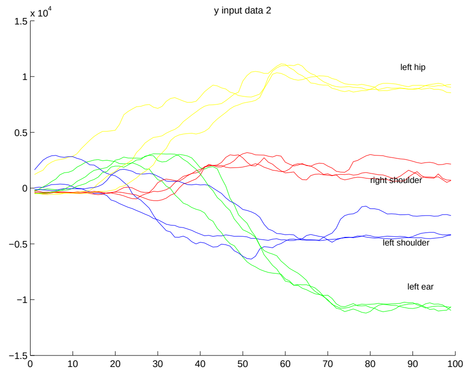


Figure 3.7: The y accelerometer input data used to train the MLP for user 1. There are three examples for each of the four different classes of gesture. This data shows the last 100 samples logged from the end point of the gesture, sampled at 100Hz.

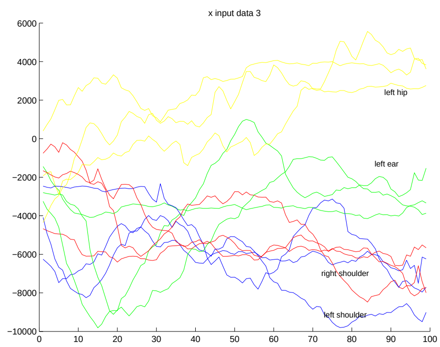
example, it is highly unlikely that a user would turn their device through a full 360 degrees with one hand. The form factor of the device may also act to place constraints on the interaction. A very small device held in the hand may produce slightly different data to larger device and could act to shape the interaction in some way and must be considered. These kind of constraints also exist in other contexts. If this system was, for example, used on a wall as illustrated in figure 3.30, there are constraints imposed from the combined physical length of the users arms and the fact that they are constrained to the surface of the wall, so it is important that we embrace these natural constraints, drawing them into our interaction design when attempting to interpret user intention.



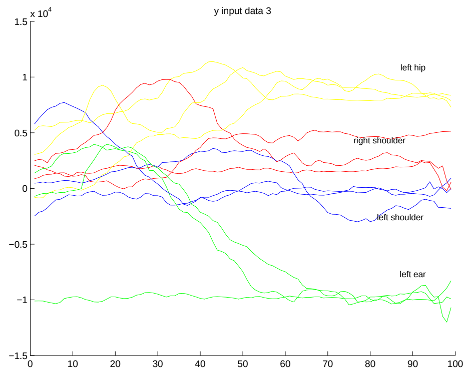
(a) The x accelerometer input data used to train the MLP for user 2.



(b) The y accelerometer input data used to train the MLP for user 2.



(c) The x accelerometer input data used to train the MLP for user 3.



(d) The y accelerometer input data used to train the MLP for user 3.

Figure 3.8: MLP training data for two different users. There are three examples for each of the four different classes of gesture. This data shows the last 100 samples logged from the end point of the gesture, sampled at 100Hz.

3.4.1 Varying Planes

Although the principle problem for our system is to recognise when the device is placed at a certain part of the body there exists another significant problem in that when the device reaches a relevant part of the body, it is required to switch to a new mode of control. But how does the system know that it is no longer constantly checking for gestures and that it must now work on a different functionality? For example, in our *BodyMusic* application, described later, how does the system know that it must switch mode to control the volume or track switching functionality? And how does it know when to switch back? This is a good example of how we utilise natu-

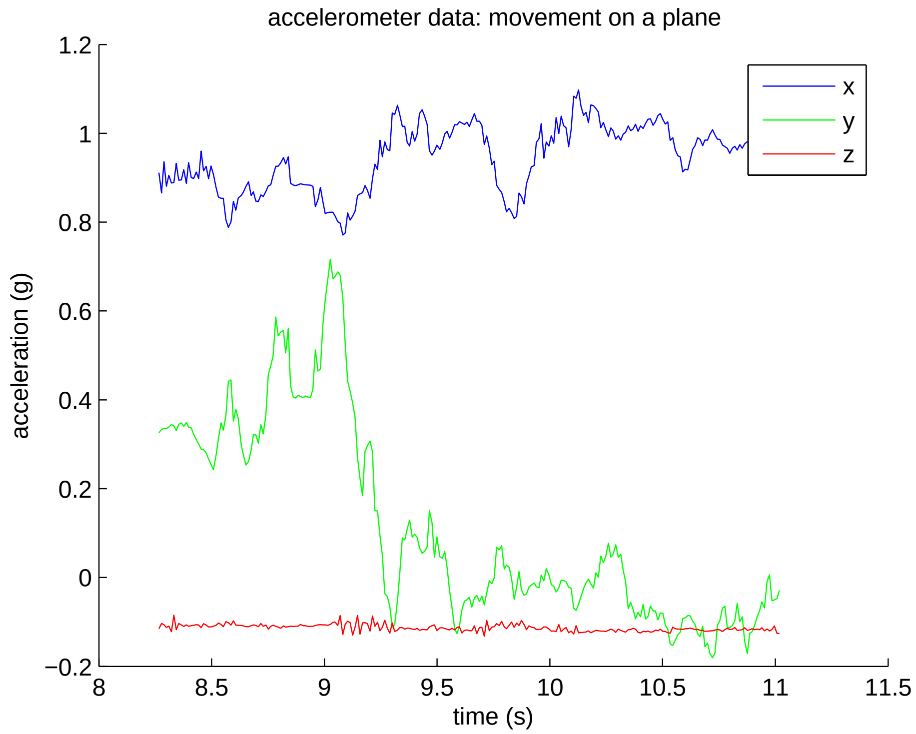


Figure 3.9: accelerometer data recorded from the movement of our mobile device on a well defined plane. We see that there is little activity in on the z-axis in this example indicating planar movement.

ral constraints to aid our interaction. If we consider a device which is being moved around on the surface of a wall as in figure 3.30, we would expect the data from our accelerometers to look a lot like that in figure 3.9. If we decompose this data using a singular value decomposition we observe that the eigenvectors of the first two eigenvalues define the plane of motion as illustrated in figure 3.10, with anything projected onto the third eigenvector indicating non-planar motion, i.e. motion off of the wall. We should expect then that any localised movement around the shoulders, head, hip or any other part of the body should display similar planar behaviour. We use

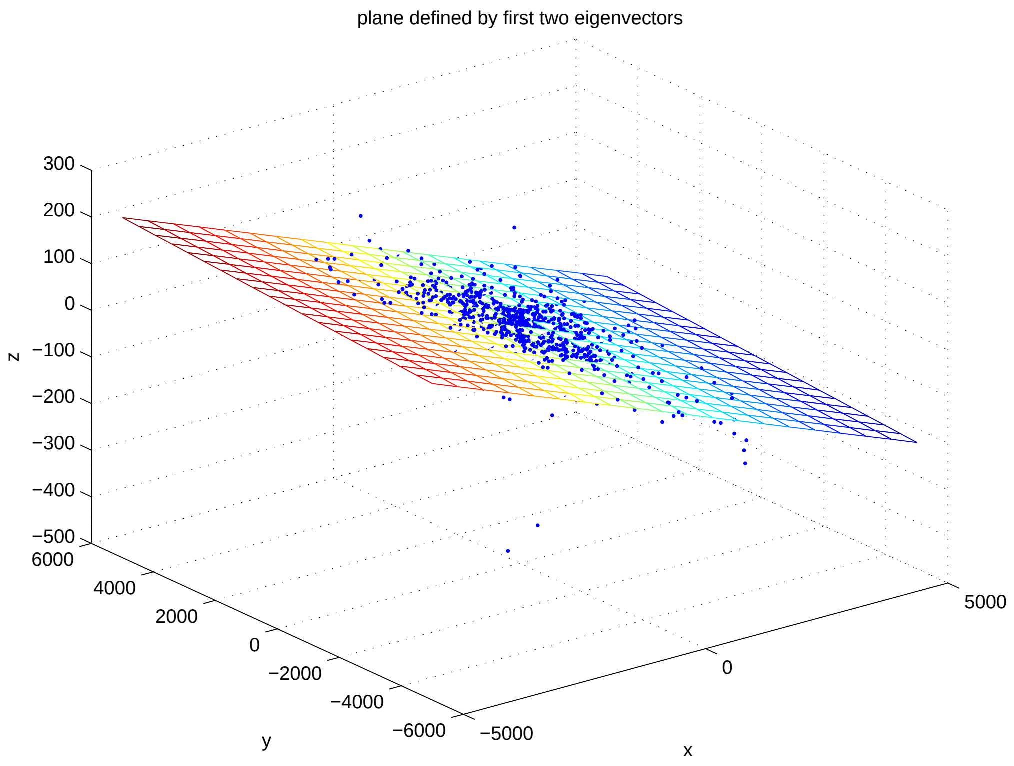


Figure 3.10: The plane created from the first two eigenvectors from the singular value decomposition of the data shown in blue.

this property in our application to control the mode switching functionality. When it is recognised that the device is at the ear, for example, the application switches to the 'track switching' mode where it remains until there is significant projection onto the third eigenvector for that plane or motion at the ear. If it is detected that the device is no longer on that plane the application switches back to the a general recognition mode.

3.5 Isomorphism

One important issue which needs to be addressed and which is a problem for any system is the mismatch between what a user perceives the system to be doing and what the system is actually doing, referred to as an *isomorphism error* (Williamson 2006). Problems can arise when some activity is sensed by the system but does not communicate any intention, such as disturbances in motion caused by external forces such as a moving vehicle. All external disturbances are sensed in the same way as intentional movements but communicate nothing about the intention of the user. This issue occurs because of the mismatch between the sensing hardware and the users expectations of the sensing hardware. In our case the user perceives the system to be checking the position of the device with respect to the location on the body but in reality what the device is doing is simply monitoring angles and pattern matching accelerometer data. Any external disturbance to the device, when looked at as nothing by the user, is seen as more data to monitor by the system. It is important that we reduce the effect of this isomorphism error on our system. Of course we can reduce the effect of this error by creating a better match between the system inference and user belief or we may use more sophisticated sensing but this is not always practical due to the limited range of sensing available, the increased financial cost of more sophisticated sensors or simply the added hassle to the user.

3.6 Approaches to Feedback

As was mentioned previously in chapter 2, feedback is required for the control of any system subject to uncertainty. To provide continuous formative feedback with our interaction it is necessary that we provide some kind of mechanism for achieving this. In previously described systems, simulated physical models were used to achieve this. Others use a state space approach where a number of densities can be placed in the space, and the trajectories in that state space can then be sonified using granular synthesis (Williamson and Murray-Smith 2005*b*). We describe here two approaches to providing feedback. One provides feedback from a simulated physical model and the other uses a dynamic systems approach to gestural interaction using Dynamic Movement Primitives, which model a gesture as a second order dynamic system followed by a learned nonlinear transformation.

3.6.1 Dynamic Movement Primitives

The *Dynamic Movement Primitives* (DMP) algorithm proposed by Schaal *et al.*, is

> *"a formulation of movement primitives with autonomous non-linear differential equations whose time evolution creates smooth kinematic control policies"* (Schaal *et al.* 2004).

The idea was developed for imitation-based learning in robotics and is a natural candidate for the provision of feedback in continuous control problems. If we take the generation of body-based gestures as our example, these dynamic movement primitives allow us to model each gesture trajectory as the unfolding of a dynamic system, and is better able to account for the normal variability of such gestures. Importantly, the primitives approach models from a specific origin to a specific goal as opposed to the traditional one point to another point gestures used in other systems. This, along with the compact and well-suited model structure enables us to train a system with very few examples, with a minimal amount of user training and provides us with the opportunity to add rich continuous formative feedback to the interaction during the gesture. Dynamic Movement Primitives also posses another advantage in that they have guaranteed stability and so they can perform the control task and predict what kind of behaviour to expect from the user throughout the gesture.

DMP's are linearly parameterised enabling a natural application to supervised learning from demonstration. Gesture recognition is made possible by the temporal, scale and translational invariance of the differential equations with respect to the model parameters.

A Dynamic Movement Primitive consists of two sets of differential equations, namely a canonical system, $\tau\dot{x} = h(x)$ and a transformation system, $\tau\dot{y} = g(y, f(x))$. A point attractive system is instantiated by the second order dynamics

$$\tau\dot{z} = \alpha_z(\beta_z(g - y) - z) \tag{3.3}$$

$$\tau\dot{y} = z + f \tag{3.4}$$

where g is a known goal state (the left shoulder, for example), α_z and β_z are time constants, τ is a temporal scaling factor, y and $\dot{y}$ are the desired position and velocity of the movement and f is a linear function approximator. In the case of a non-linear discrete movement or gesture the linear function is converted to a non-linear deforming function

$$f(x,v,g) = \frac{\sum_{i=1}^{N} \psi_i w_i v}{\sum_{i=1}^{N} \psi_i}, \text{where} \quad \psi_i = e^{\left(-h_i\left(\frac{x}{g}-c_i\right)^2\right)} \tag{3.5}$$

These equations allow us to represent characteristic non-linear behaviour that defines the gesture, while maintaining the simplicity of the canonical 2nd order system driving it from start to goal. The transformation system for these discrete gestures is

$$\tau\dot{z} = \alpha_z(\beta_z(r-y)-z)+f \tag{3.6}$$

$$\tau\dot{y} = z,\ \tau\dot{r} = \alpha_g(g-r) \tag{3.7}$$

where $\dot{z}$, z and y represent the desired acceleration, velocity and position respectively.

Movement Primitive Example

We take a simple example of a gesture to the back of the head. Accelerometer data is recorded for one example of this gesture. The approach to learning and predicting the dynamic movement primitive is to provide a step change in reference and pass this through the non-linear deforming function described above. Values for the f's can be calculated along with sets of x's and v's from the canonical system. The attractor landscape is then learned, in this case, by a Locally Weighted Projection Regression (LWPR) algorithm (Vijaykumar and Schaal 2000) (although alternatives such as Gaussian Process regression (Rasmussen and Williams 2005) may also be used) allowing us to make predictions of the function f given values for x and v. So if we were to train the system with our desired gesture, any further performance of the gesture could be compared to the learned dynamic system and feedback provided which was proportional to any deviation from that dynamic system. Figure 3.11 shows us an example of 'real data' in red along with the learned dynamic system representation in blue. One thing to notice from figure 3.11 is a higher frequency remnant in the real data, which is not meant to be modeled by the movement primitive, but is a remnant from the tremor in the performer's muscles.

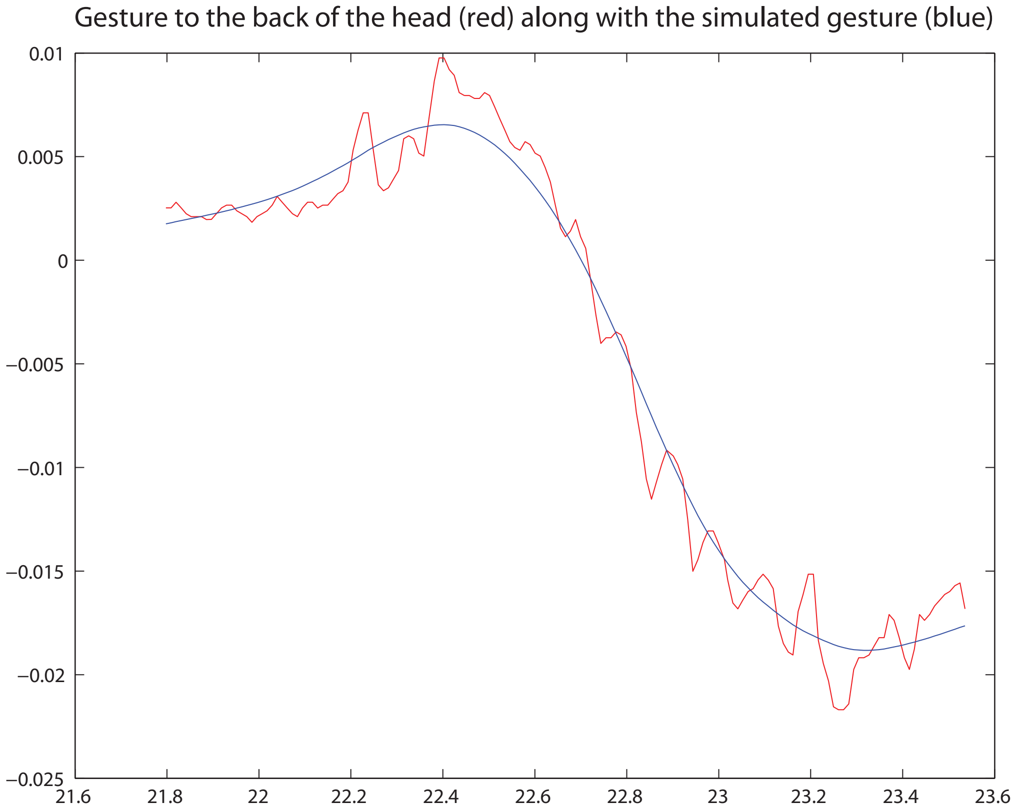


Figure 3.11: Example of a simulated gesture alongside the real x-axis gesture data from the chest area to the back of the head

Muscle Tremor

'Muscle Tremor' is present in everyone. In fact oscillatory behaviour is a common form of normal biological function and is described by Beuter *et al.* (2003) as "an approximately rhythmical movement of a body part". The aspect of muscle tremor we wish to exploit is often referred to as a person's 'physiological tremor', which is part of a category of tremor referred to as 'postural tremor'. The investigation of muscle tremor can be very complex and there are many differing forms of tremor studied. There are two main classifications of tremor in use. The first is based on the state activity of a body part when tremor is observed and the second is based on the etiology of an underlying disease or condition (Beuter *et al.* 2003). The classification of tremor by state activity includes: (Bain 1993, Deuschl *et al.* 1998)

- *Rest Tremor* occurring when relevant muscles are not activated and the body part is fully supported against gravity.

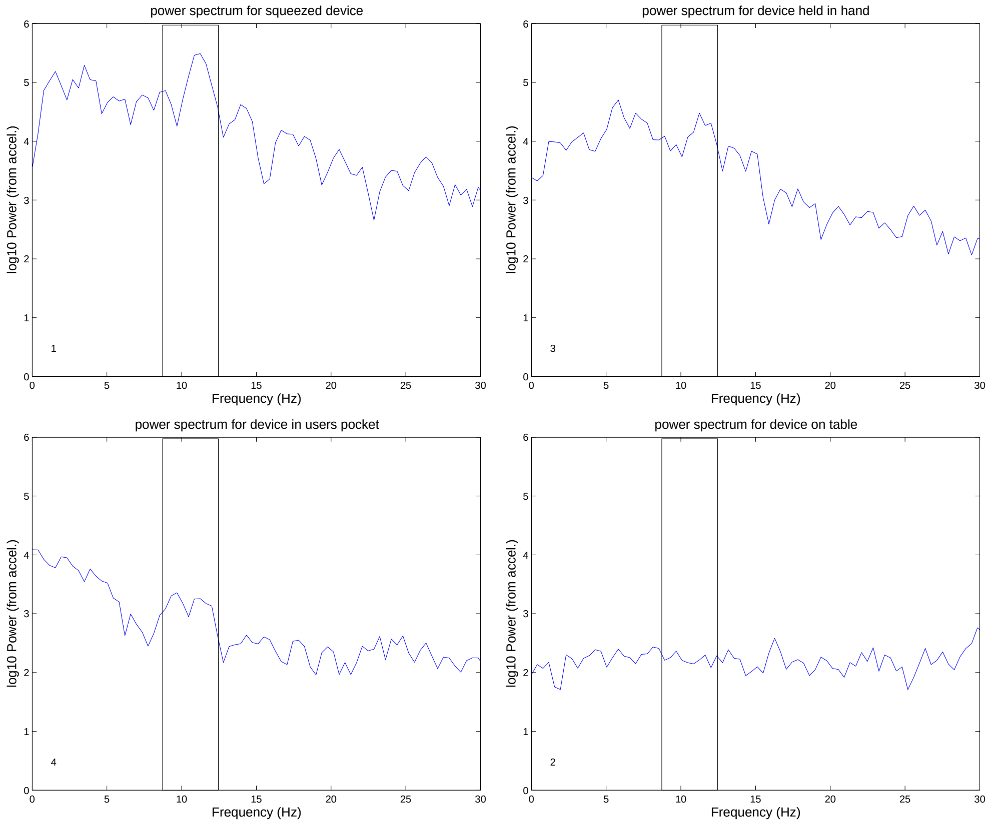


Figure 3.12: Power spectra of accelerometer data in four differing situations. Large activity can be seen in the 8-12 Hz range in the case where the device is squeezed and smaller activity can be seen in the cases where the device is held in the hand and where the device is in the user's pocket. For the case where the device is left on a table there is no 8-12 Hz tremor activity.

- *Action Tremor* occurring when relevant muscles are activated, which includes postural, kinetic, isometric, intention, and task-specific tremors.

There are two separate oscillatory components apparent in 'normal' physiological tremor. The first component is produced by the physiology of the arm and has a frequency determined by the mass and stiffness of a persons limb. This is due to the passive mechanical properties of body parts that are a source of oscillation when they are perturbed by external or internal forces. The second component of muscle tremor is referred to as the 8 to 12Hz component. As opposed to the first component, the 8-12Hz component is resistant to frequency changes. Its amplitude, however, can be modified by manipulating limb mechanics (Beuter *et al.* 2003) and it is this characteristic of muscle tremor that we can potentially incorporate into our interaction design.

The ‘Action Tremor’ category and more specifically, isometric tremor, which occurs when a voluntary muscle activity is opposed by a rigid stationary object, is interesting because it has potential use as an input mechanism in our interfaces (Strachan and Murray-Smith 2004). Figure 3.12-1 shows the power spectra of accelerometer data for a PDA held in the hand and squeezed. We observe here a peak in the 8-12Hz range which is still there, but to a lesser extent, when we examine the power spectrums from a device held normally in the hand as in figure 3.12-2 and in the user’s pocket as in figure 3.12-3 and not observed at all for the device left motionless on a table as in figure 3.12-4. Figure 3.13 shows a spectrogram for a device, which is held in the hand and repeatedly squeezed. It is clear where this squeezing is happening due to the increased activity in the 8-12Hz range, indicated by the strong red colours. This highlights the potential for use of this information in our interfaces even for the simple example of ‘device in hand’ and ‘device not in hand’. There are also possibilities for the use of the tremor signal as part of the state vector for inferring our current body pose, as described in section 3.3.1, since a device held at different parts of the body gives a slightly different characteristic power spectrum.

3.7 BodyMusic: Gesture Controlled MP3 Player

In this example our BodySpace system utilises our body as the interface for a music player. By placing the device at different parts of the body we may control the different functionalities of the music player, such as the play/stop functionality, volume control and track switching.

A model based approach to this kind of interaction has a number of advantages. As discussed in chapter 2, a model based approach to interaction with the simulation of a physical model provides an immediate intuition to the user. This kind of approach also allows us to alter our feedback easily by simply changing some value or parameter associated with the model, such as, in the following example, the friction on the surface of the ball in the bowl, the height of the bowl or the mass of the ball. This kind of approach is also useful in cases where there may be increased general movement, such as noise from walking movements or movement from being inside a vehicle.

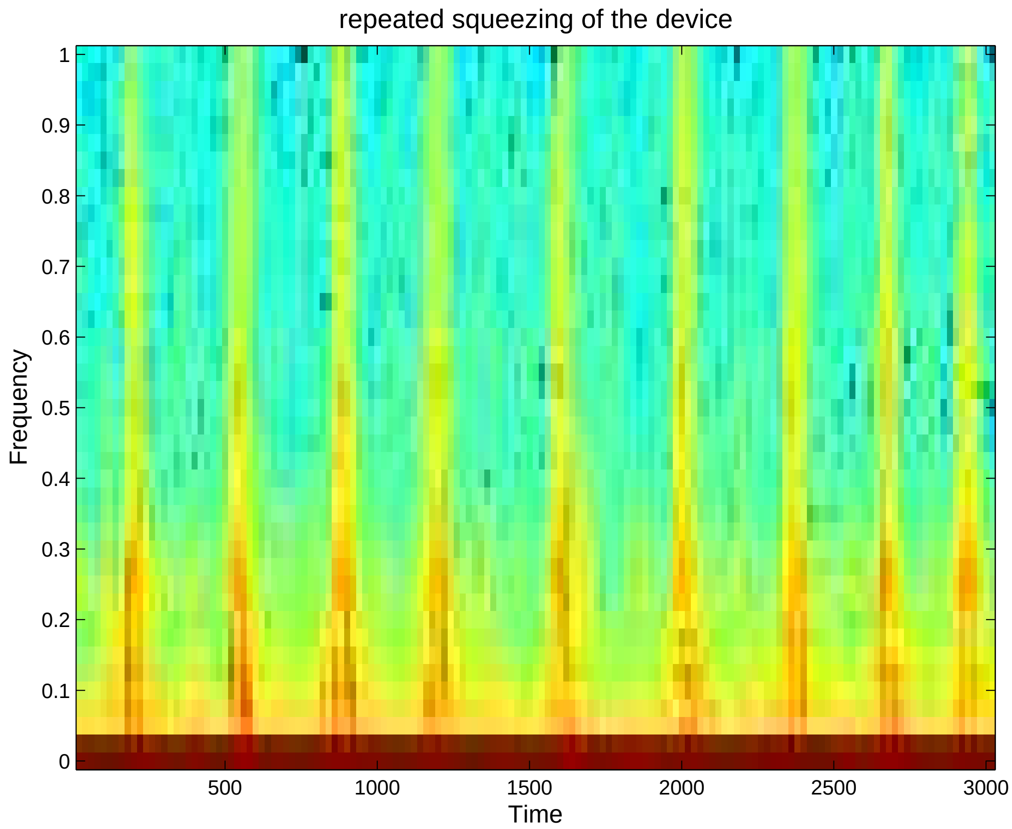


Figure 3.13: A spectrogram for the repeated squeezing of a device. We see increased activity in the 8-12Hz region at the points when the device is squeezed.

3.7.1 Model: Ball in Bowl

Our system uses a ‘ball in a bowl’ physical model to represent interaction with this system. We can imagine a ball placed in a bowl or concavity as shown in figure 3.14. Intuitively, if we tilt this bowl the ball will roll to the side. If we tilt the bowl over a certain point the ball will roll over the edge into the next bowl. Similarly, if we give the bowl a sharp flick we may propel the ball into the next bowl. We use this simulated model to control the track switching and volume control functionalities of our music player. When it is recognised that the device is at a certain part of the body corresponding to that functionality, the system switches to the correct mode and model associated with that part of the body. So for example, when we wish to switch tracks, the device is first moved to the left ear where recognition occurs. A mode switch then happens, allowing the device to be tilted back or forward at the ear in order to switch tracks, as in figure 3.14, where each bowl or concavity represents a different track. With a row

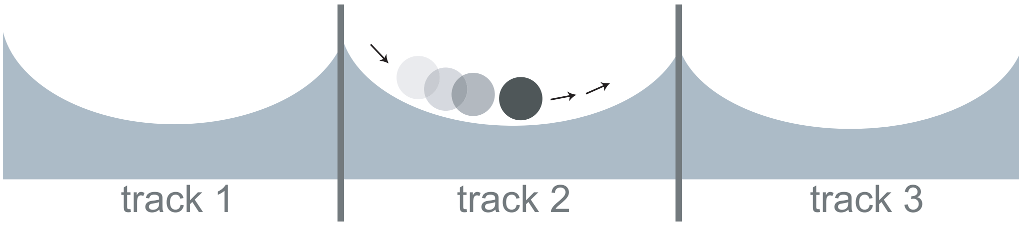


Figure 3.14: Combination of bowls which the user must navigate the ball through in order to switch tracks.

of bowls representing a list of tracks, it is possible to simulate the task of transferring a ball from one bowl to the next by providing an external force from the movement of the device. In this case the external force comes from a flick of the device, as shown in figure 3.15. Increased velocity and momentum of the flick would allow users to reach the peak, and effectively fall into the next track. We may model the surface friction and the effort required to overcome the peak of the bowl with some simple physics. Each bowl is represented by a simple parabola, with a certain height, y, used to calculate angle of slope: $\theta = tan^{-1}(\frac{x}{y})$ and the force: $F = mg\sin\theta$ (where mg is mass $\times$ gravity) minus surface friction (Kelly 2004). This interaction is also augmented with vibrotactile feedback allowing the user to feel when the track switch has happened, where the level of feedback presented to the user is associated with a parameter of the physical model.

A similar mechanism is used to control the volume of a track. The volume control is located, in this set-up, at the left hip. So when the device is placed at the left hip the mode switches to a volume control mode. This mode is represented by one bowl only as shown in figure 3.16 so that when the device is held level there is no change in the volume but when the device is tilted the ball rolls to one end of the bowl over a number of lines, each representing a vibrational pulse, which in this instance consists of a square wave with a frequency of 250Hz and amplitude of 45dB. At the end of the bowl the ball is stopped and a larger vibrational pulse is felt by the user. One end of the bowl represents the volume-up functionality and one end represents the volume-down functionality. Why can we not represent

Figure 3.15: An Illustration of the BodyMusic functionality

this functionality with a simple threshold crossing in the accelerometer signal? This would have been simple to implement, but what the model based approach allows us to do is provide feedback at all stages of the balls movement in an intuitive manner by simply linking the vibrotactile feedback, in this case, to the current position of the ball within the bowl. This context

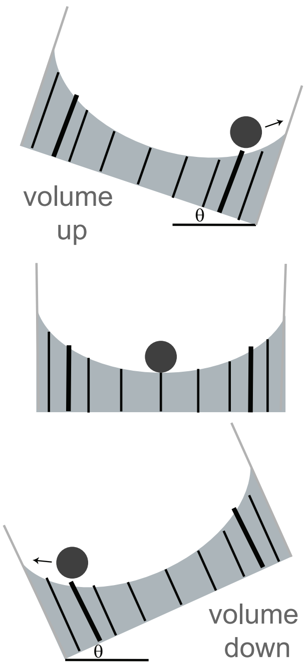


Figure 3.16: When the ball rolls into the left side of the bowl the volume decreases. When it is rolled into the right side of the bowl the volume increases

could be detected by the system which could then alter the dynamics of the model. For example, the bowl could become much larger when the user is walking or the movement of the ball on the surface of the bowl could become more viscous making false positives much less likely to occur.

Figures 3.17 and 3.18 show examples of how the accelerometer data interacts with our simulated physical model. Figure 3.17 shows how accelerometer data provides the energy to the model which switches the current track by causing the ball to roll into the next bowl as shown in figure 3.14. Here the track is moved forward five times then back again five times. Figure 3.18 shows that as the device is tilted, the ball in the physical model is pushed to the edge and passes a threshold which causes the volume to change. Here the volume is first decreased then increased again.

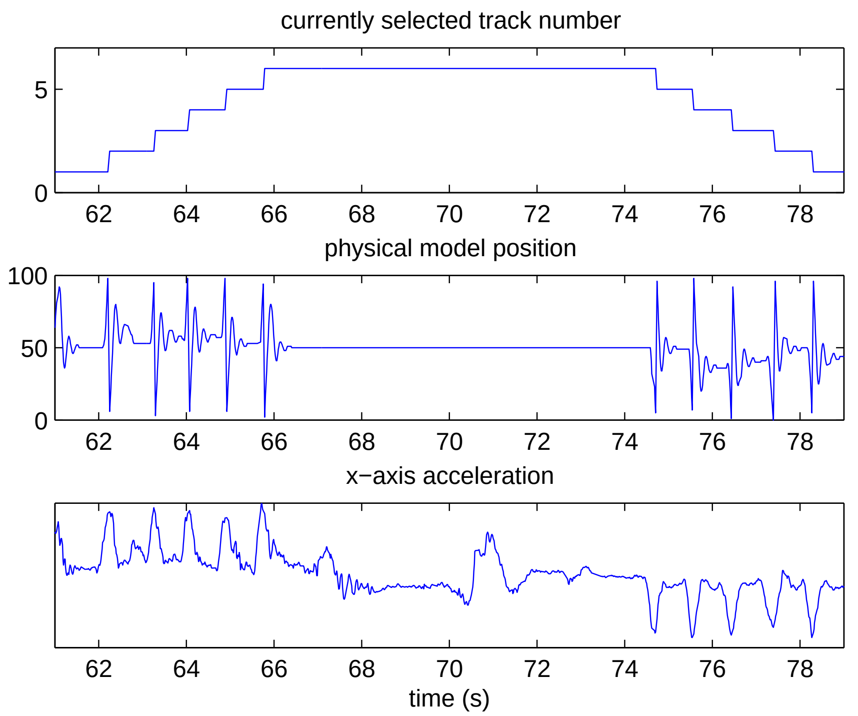


Figure 3.17: Example of the data observed for a track switching task.

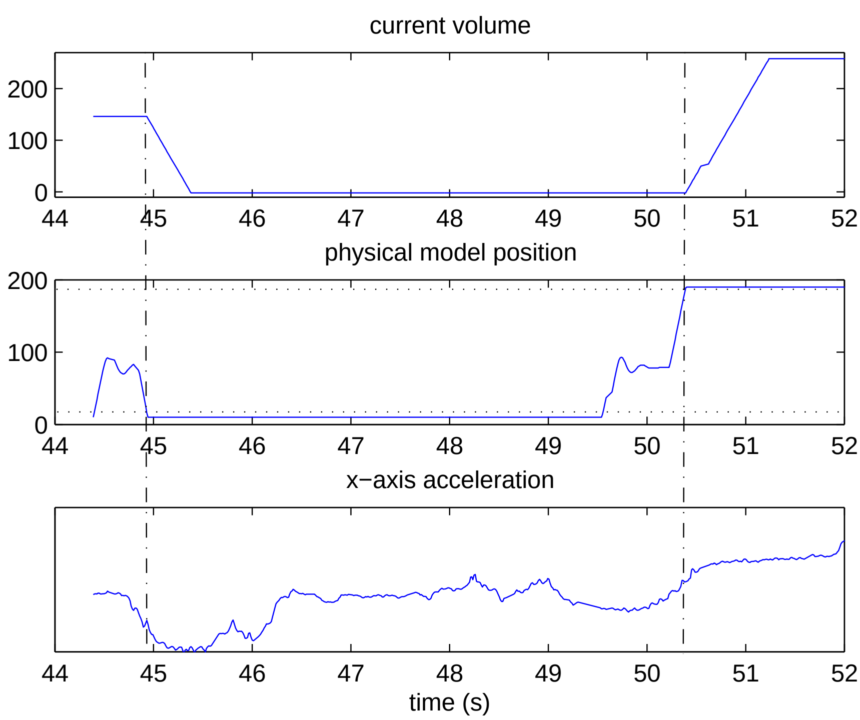


Figure 3.18: Example of the data observed for a volume control task.

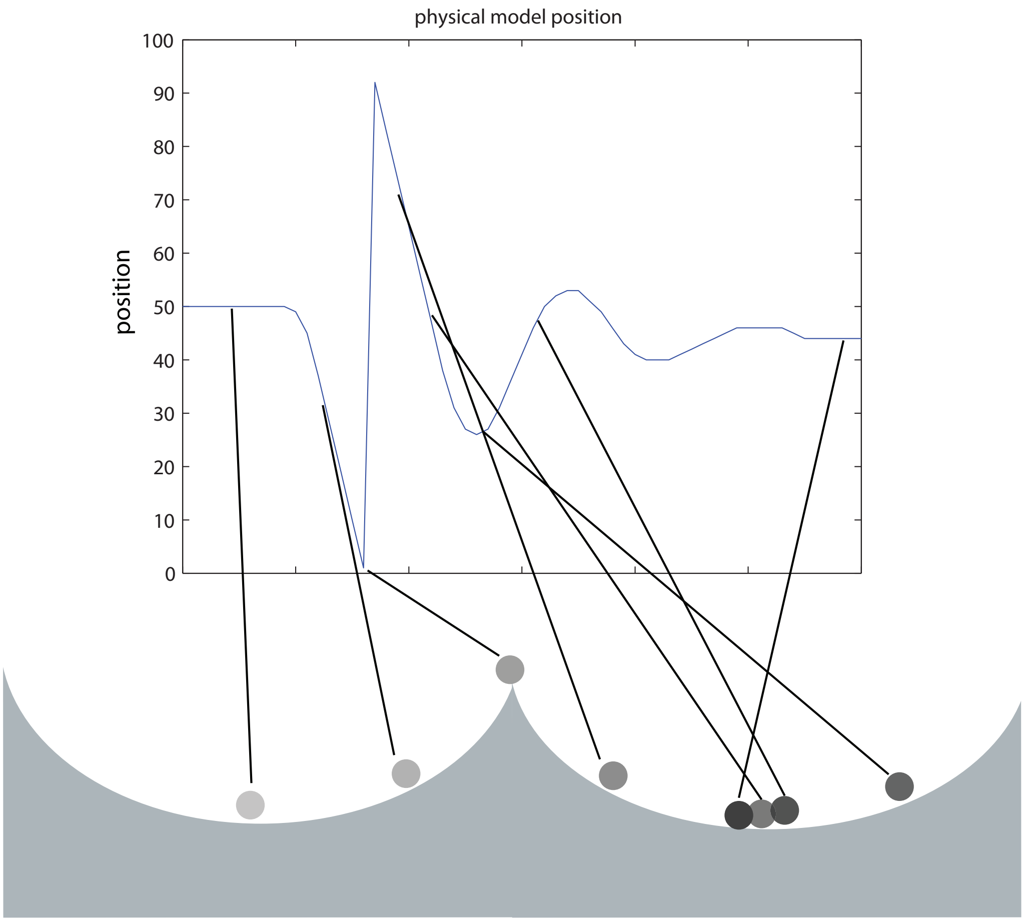


Figure 3.19: Visualisation of how the ball moving between bowls corresponds to the model data

3.7.2 System Testing

Testing was conducted with a number of participants, principally to demonstrate the that the system worked for more than one user and to explore how users interacted with the simulated physical model highlighting any weaknesses and potential improvements in this basic configuration. Six participants were used, all aged between 21 and 30, with 5 male and 1 female.

Method

Five different tasks were performed by the participants. These tasks were:

1. Navigate to Track A
2. Decrease the volume on Track A
3. Increase the volume on Track A
4. Move Forward 5 Tracks
5. Move Back 5 Tracks

Each participant was given a brief introduction to and demonstration of the system before being allowed to practice and develop a feel for the interaction. They were then asked to perform all 5 tasks twice using their left hand at their left ear, for track switching and at their left hip for volume control. All data from our accelerometers and our physical model was recorded.

Results

The logged data (logged using a built in data logger within out application) shows that after only 3-4 minutes of practice, the system could be used effectively by all the participants. All participants performed better in the second run than the first but still none were without mistake. Figures 3.20-3.24 shows the successful performance of each task 1 to 5. One significant problem, especially in the participants' initial attempts at using the

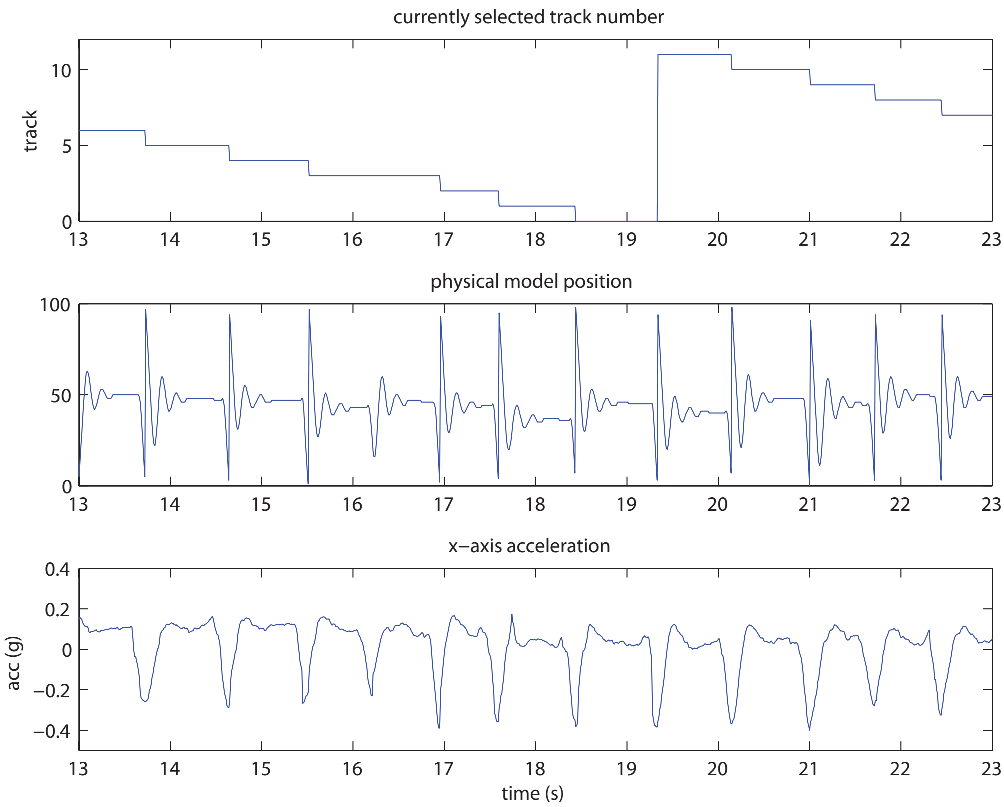


Figure 3.20: Example of a completed traversal to track 7.

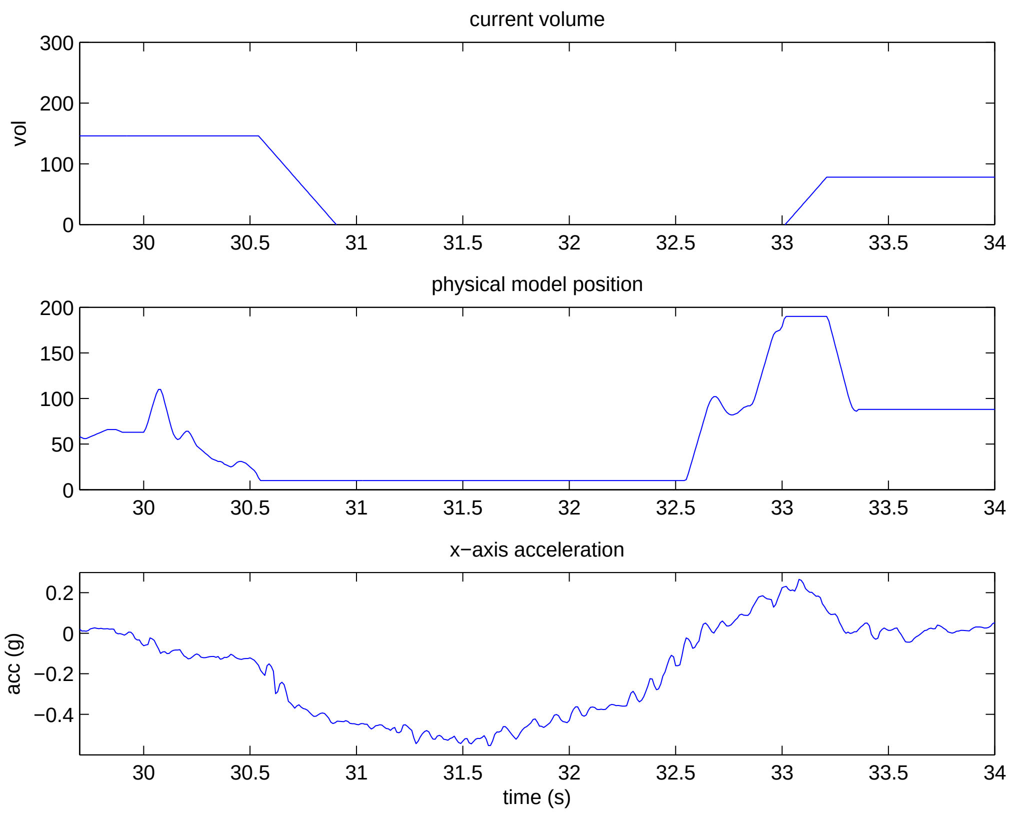


Figure 3.21: Example of volume decrease to zero with false positive increase at the end.

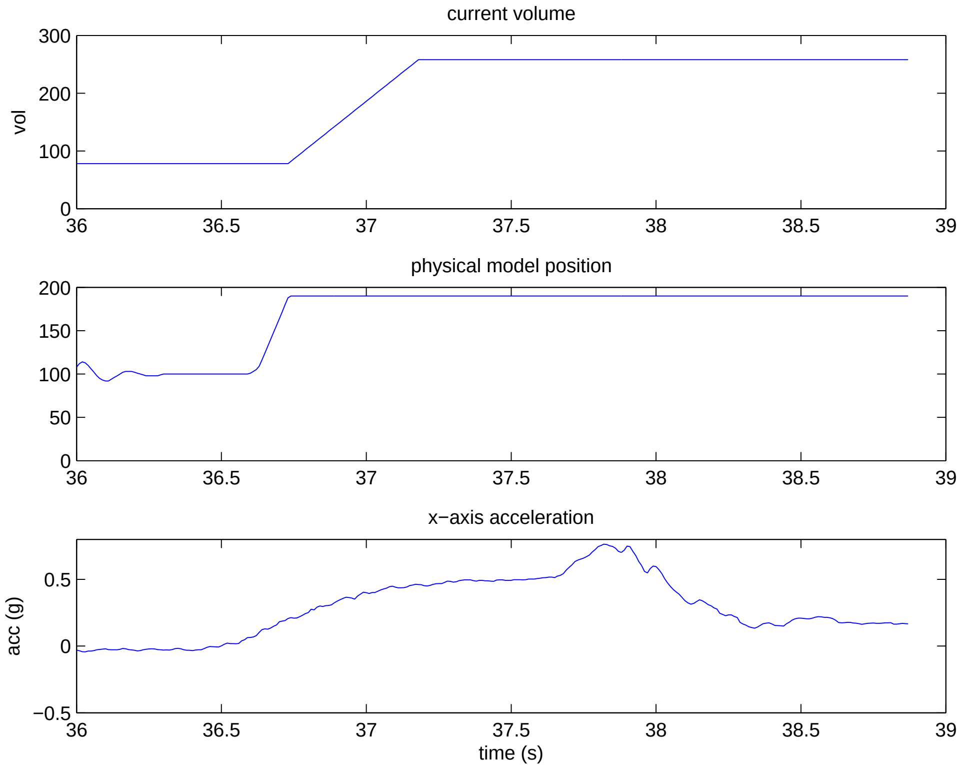


Figure 3.22: Example of a volume increase to the maximum value.

system was with track switching. Initial attempts at switching the track, either backwards or forwards, generally ended with multiple track changes as shown in figure 3.25 or complete loss of control of the system as displayed in figure 3.26. Another common problem observed was the movement of the device away from the left ear which would occasionally cause an extra track switch as displayed in figure 3.27. The volume control tasks were more successful with most users successfully increasing and decreasing the volume as illustrated in figure 3.21-3.22. One frequently occurring problem, as illustrated in figure 3.21 and as mentioned for the track switching tasks was the unwanted volume change as the device was taken away from the hip. This is a good example of the general segmentation problem affecting gestural interaction based systems generally.

Observations

Interesting variations in behaviour were observed for these tasks. Each user had their own comfortable posture when performing the tasks and this

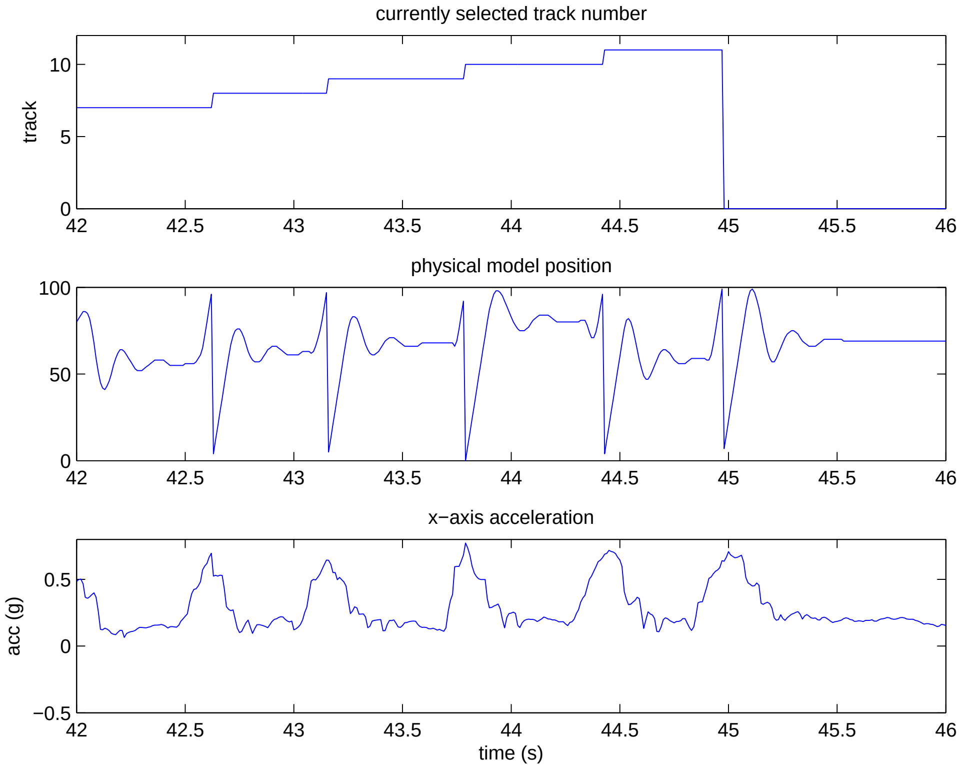


Figure 3.23: Example of track being moved five forward.

posture usually affected the rest position of their hand when placed at the ear, although the varying posture did not affect the placement of the device at the hip to the same extent, indicating a potential reason for the greater success observed in the volume switching tasks. Other observed behaviour included the drifting of the hand position as shown in figure 3.28 where the 'horizontal' level of the hand gradually changes causing the position of the physical model to drift, also without the user being aware. This tended to cause a loss of control, as also shown in figure 3.28. It was also apparent while observing the participants that forward flicks of the wrist were easier to perform than backward flicks as illustrated if we compare figure 3.29 for the backwards playlist traversal and figure 3.23 for forward traversal. Crossan and Murray-Smith (2004) describe a study that examines human performance in a tilt control targeting task on a PDA with a similar result in that there is an increase in variability of motions upwards from the centre, compared to downwards motions of the PDA.

It would be possible to remedy these problems with more careful con-

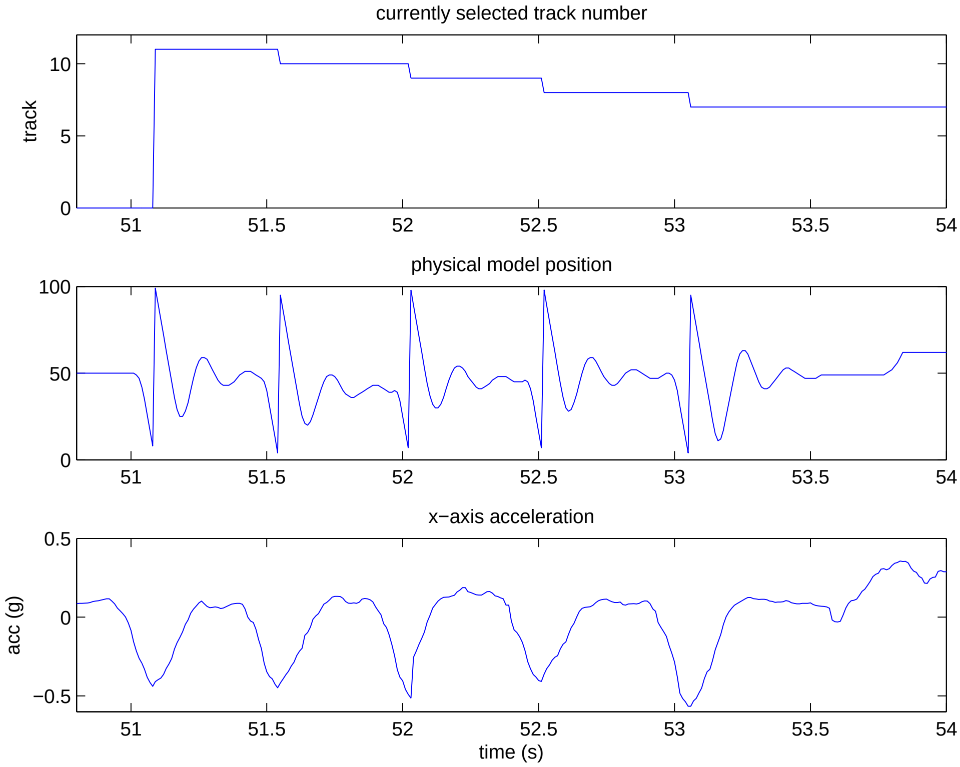


Figure 3.24: Example of track being moved five back.

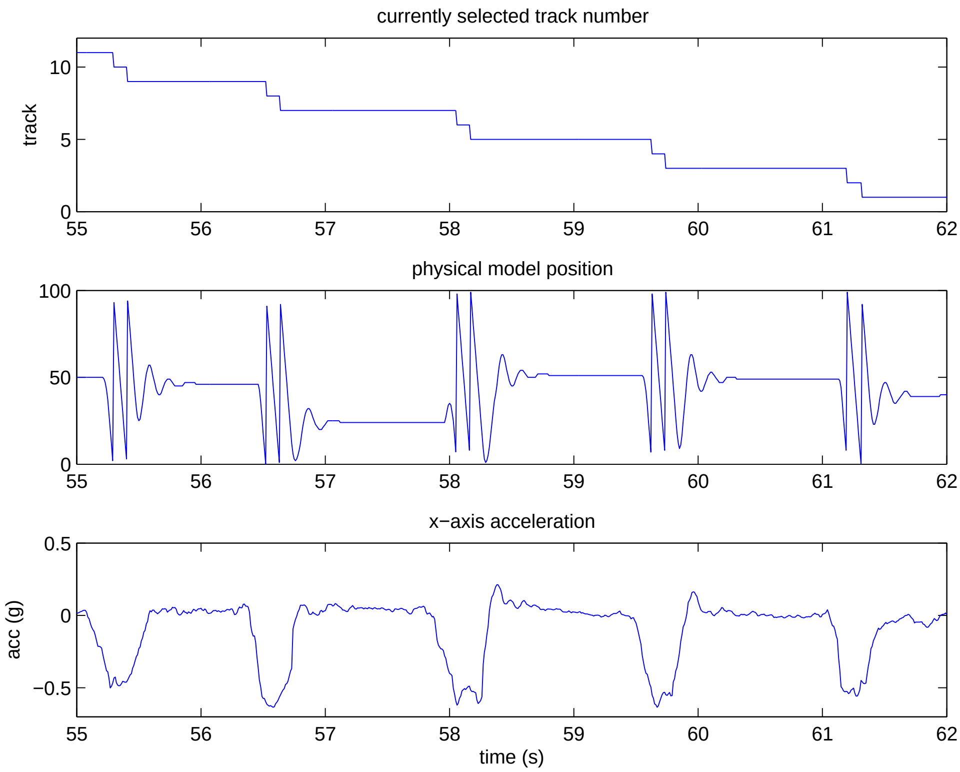


Figure 3.25: With each flick of the device, here for task 5, the track switches two times indicating that the ball has skipped two bowls.

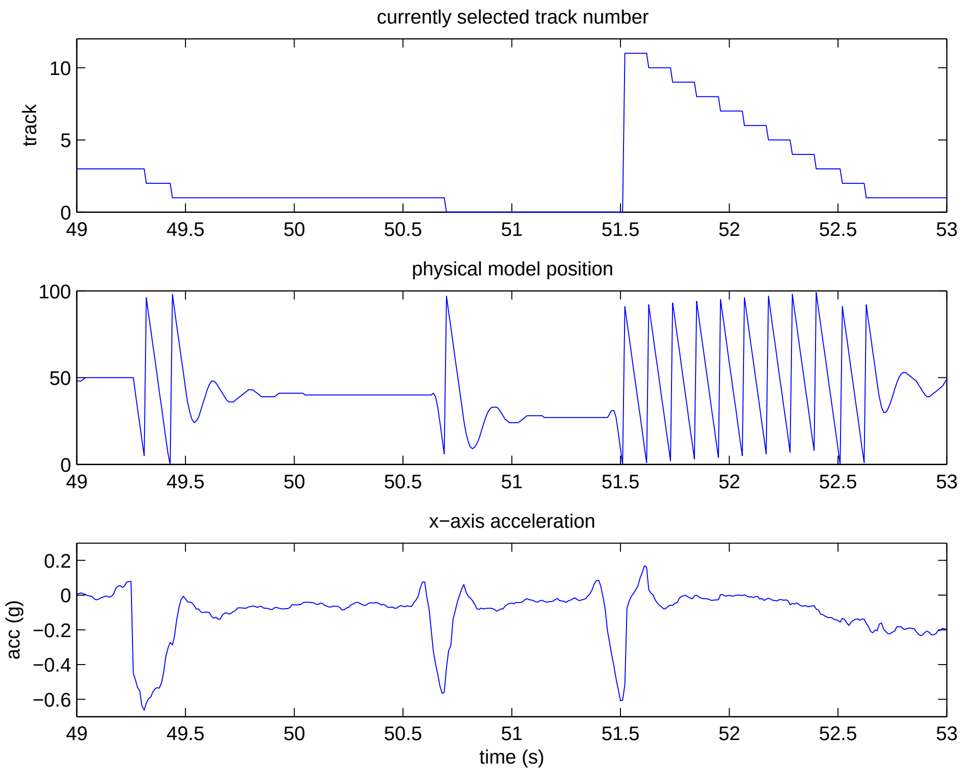


Figure 3.26: User loses control of the system at the end of this task.

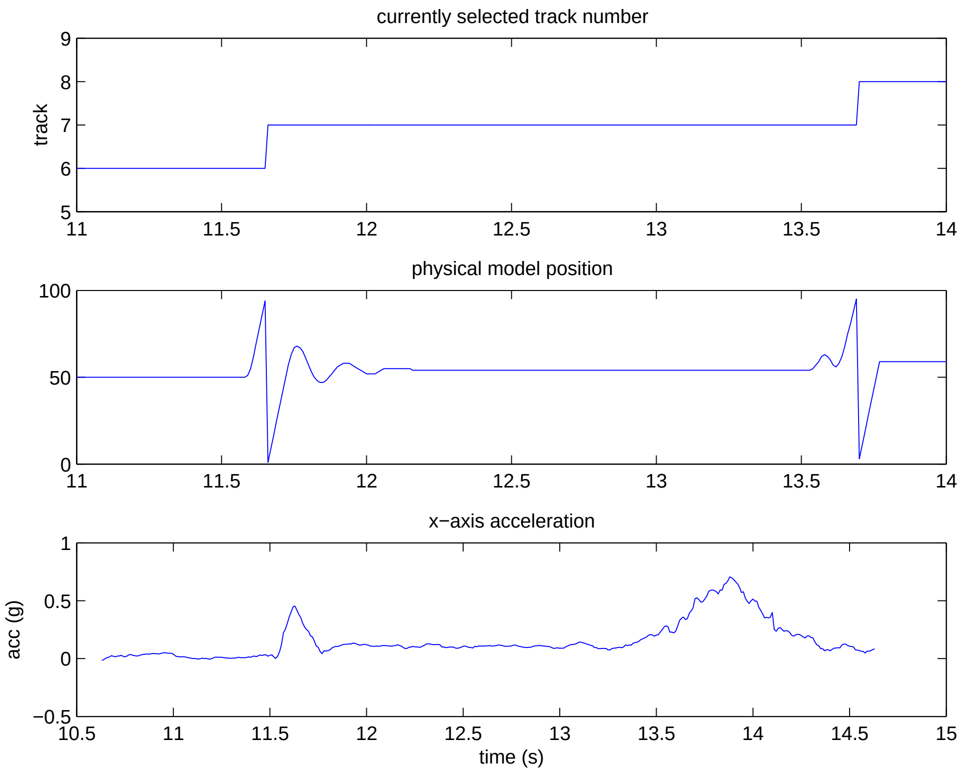


Figure 3.27: For task one switching to track 7 was completed in one step but when the device was moved away from the ear, as indicated, the track switches again by mistake.

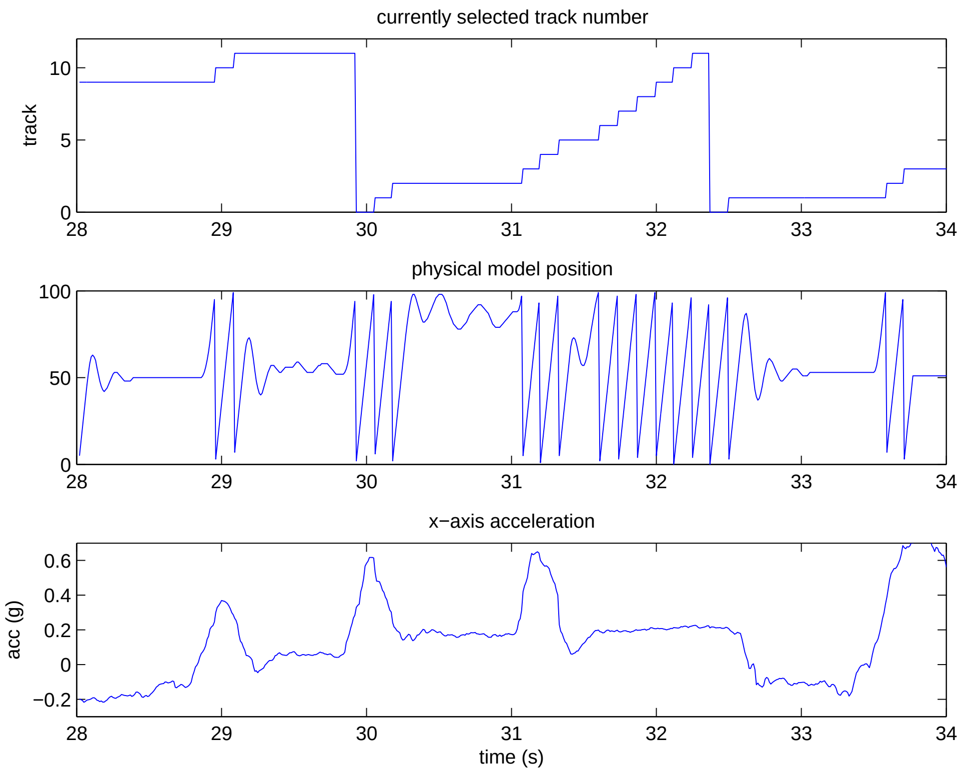


Figure 3.28: The accelerometer data at the bottom shows a gradual drift of the 'rest position' of the hand which isn't noticed by the user.

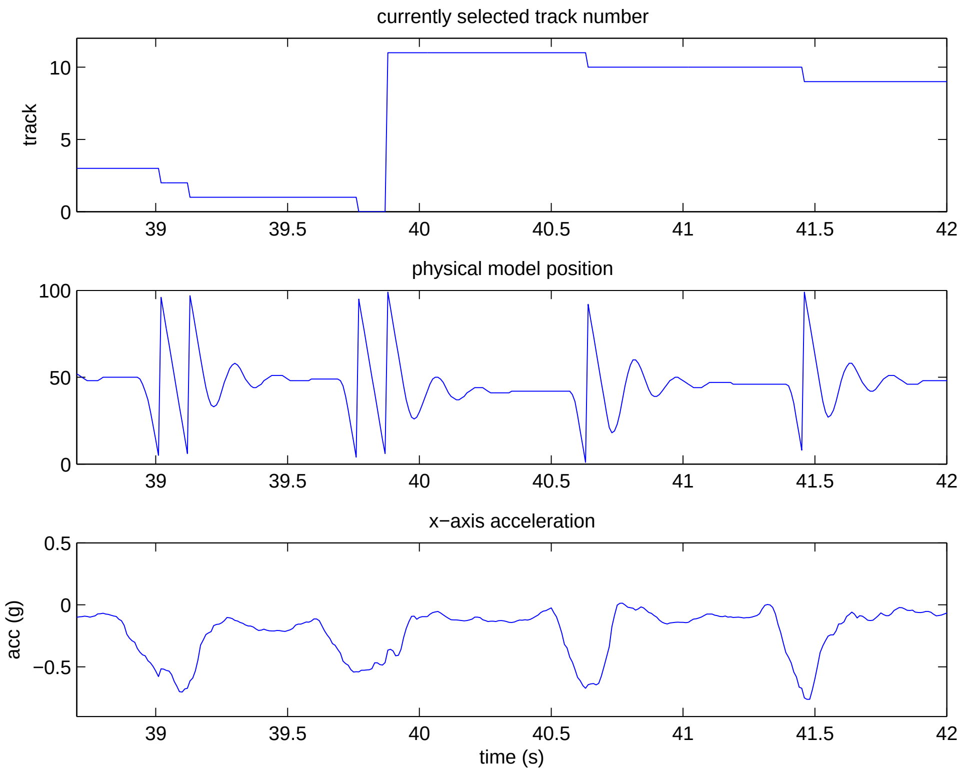


Figure 3.29: This user had a slight problem with the first two track switches in task 5 but then soon learns to move one track at a time. The large jump we at t=39.58 is caused by the wrap around of the track list number.

struction of the physical model to reflect more the strengths and weaknesses of the system user. For example, to cut down on the number of false positive track switches it is possible to simply increase the size of the bowl or to add more friction to the model, likewise, to aid the forward track mechanism it would be beneficial to construct a bowl with the 'forward side' side higher than the 'reverse side', for example.

3.8 Other Potential Applications

The control of a music player is just one application for this system but it may also be used for other tasks such as the retrieval or storing of files around the body or the activation of different functionalities at different parts of the body, for example, the activation of your to-do list when the device is placed at your head. You may also wish to call your girl/boyfriend just by placing the device at your heart or answer the phone by placing the device at your ear. But this system is not confined to the body, as the body is simply being used in this example as the mnemonic device or the interface.

With this kind of application it is possible to 'interact' with objects of interest in our general environment. An object could for example have a gesture hand-drawn on a 'PostIt' sticker attached to the object; if the user performs that gesture, the software can automatically adapt to the appropriate context, point out a location on a map, or start a particular application.

3.8.1 Off-The-Wall Interaction

It is possible to use other interfaces and in this case we have chosen to use the wall, as in the 'off the wall' interaction described earlier or on a poster.

We produced a prototype application which utilises the wall, or any poster placed on that wall as illustrated in figure 3.30, as the mnemonic device. One possible use for a system such as this includes data entry in a 'wet lab', for example as suggested by Phil Gray at the University Of Glas-

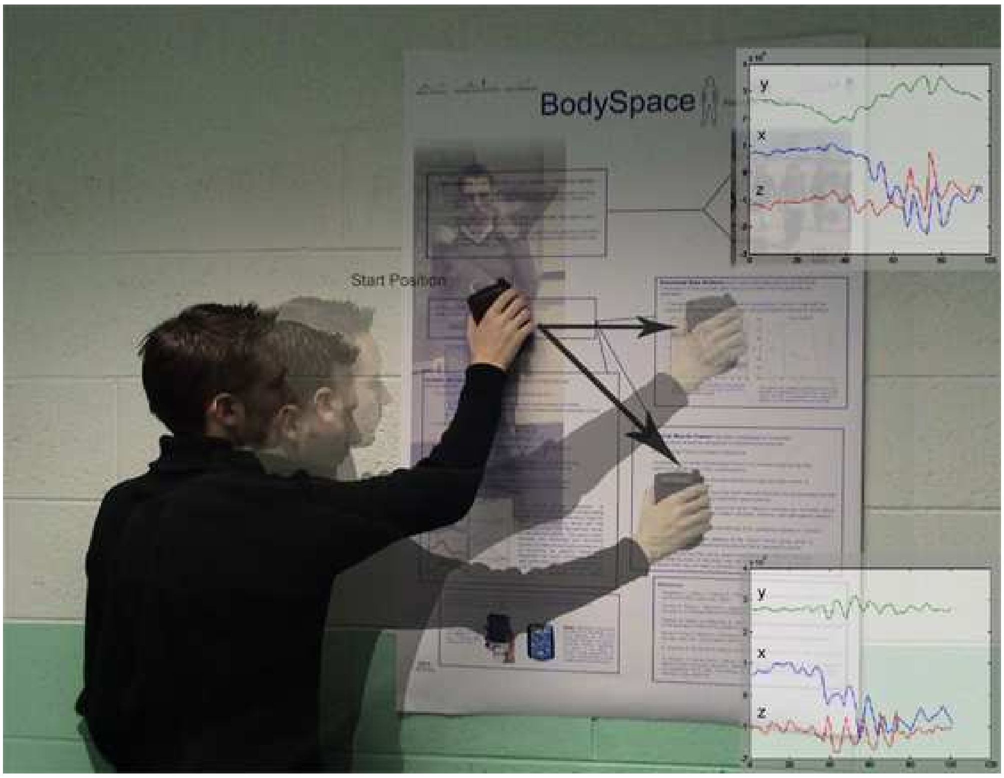


Figure 3.30: Another example application for this system. Gestures to different parts of the poster may display different information.

gow. A molecular genetics laboratory includes various hostile environments in which a researcher cannot easily enter data via normal mobile input methods such as laptop PCs and Personal Digital Assistants. Traditionally, experimental data is stored in hard-backed paper notebooks since PCs are not normally located in wet research laboratories due to various environmental risks, such as liquid spillage. Thus, data is often duplicated by first recording it into a laboratory notebook, then inputting it into the PC elsewhere. This duplication increases both the chances of erroneous data and the overall time spent in the data capture process (McLeigh 2007). It would be useful therefore if researchers in these wet labs had access to a system which allowed them to input data directly to the system but removed any risks from working with electronic equipment in this kind of environment. It would be considered beneficial then if a researcher could simply gesture on an annotated wall with a simple gesturing device in their hand as in figure 3.31, with each gesture activating a different functionality associated with the data entry task.

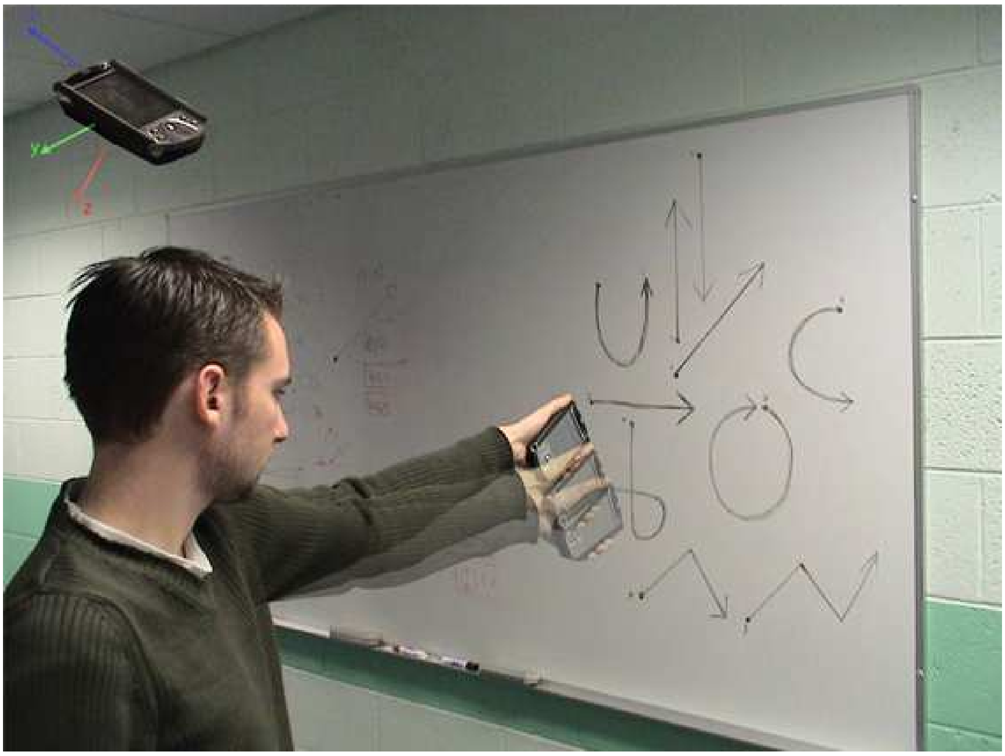


Figure 3.31: One gesture being performed on a whiteboard

In order to demonstrate the utility of this approach data was logged for the 10 gestures shown in Figure 3.32, which shows the x, y and z coordinate accelerations for gestures 5, 6 and 8 from the list, respectively, and it is observed that these three rather distinct gestures also have three distinct acceleration traces, with reasonably good repeatability among gestures, although there is some timing variability. To examine the recognition performance of the prototype application, 10 example gestures were performed for each class of gesture by the author and the results recorded. We achieved 95% successful recognition with only 5 misclassifications.

One of the most attractive features of this kind of approach is the possibility of creating new objects for interaction by simply scribbling on a piece of paper, sketching on a whiteboard as in figure 3.31 or arranging shapes on a table or other surface. In the case of the molecular genetics wet lab, during the handling of gels a researcher can potentially get their hands wet, so the use of PDAs or digital pens is problematic. Vision-based alternatives to measure gesture movement are more difficult to implement, due to the variety of hostile environments in which data may be recorded. Interfaces

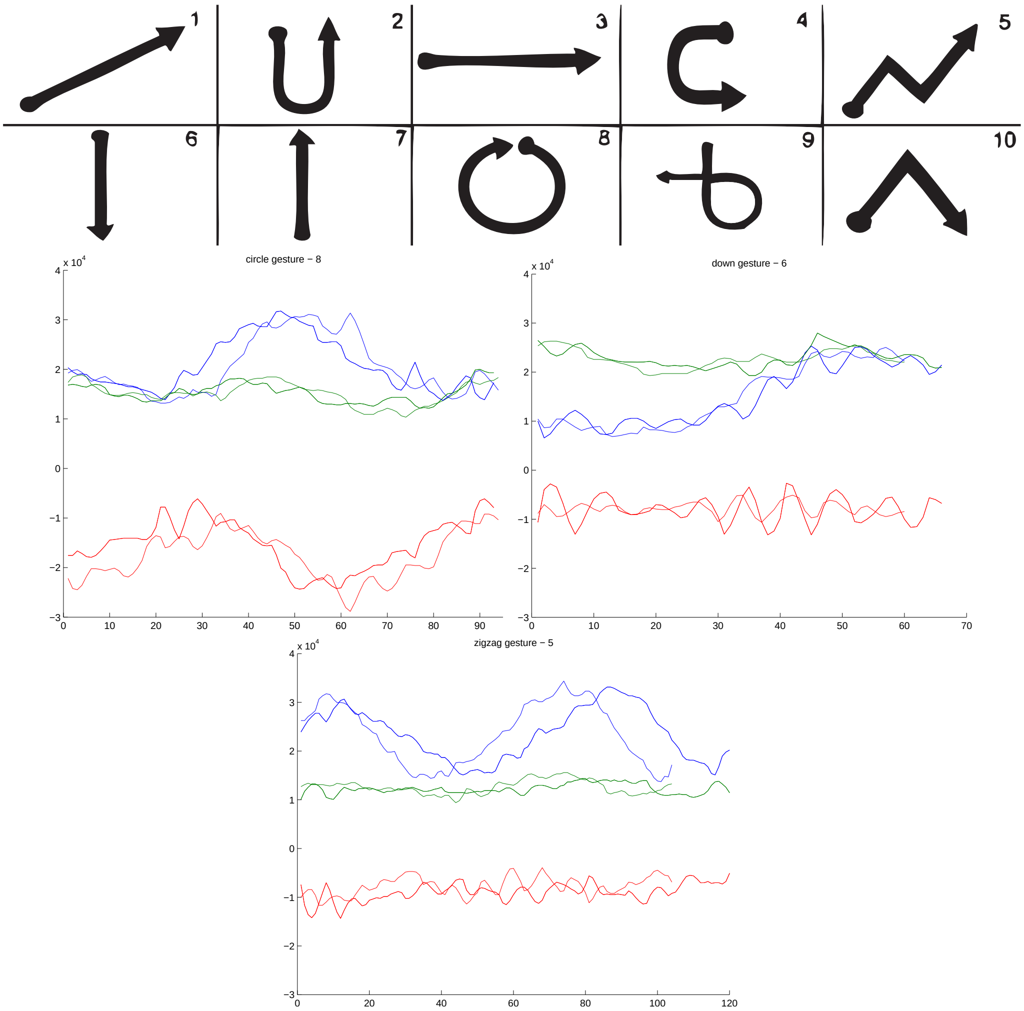


Figure 3.32: Acceleration traces for 3 examples of 3 classes of gesture.

drawn on paper offer a satisfactory solution to data capture in the laboratory. Paper is cheap and can be easily replaced if damaged. It can be shared between researchers and copied easily. A library of different paper interfaces can also be created that address different data capture needs. The "post-it" location-aware approach provides an extremely flexible way for users to define their own location-aware context, using simple hand-sketched wall gestures, or in some cases basing the gesture on features of e.g. the wallpaper, the view from the window, or other features in a room.

3.9 Discussion and Conclusions

In the preceding sections we described a handheld system that utilises inertial sensing and basic pattern recognition to allow the gestural control of a music player by simply placing the device at different parts of the body, rather than having to press buttons, or dial wheels or wear instrumented clothing. We have demonstrated a new approach to the segmentation and recognition of gestures for this kind of application and shown that a model-based approach to this kind of interaction can be both intuitive and enables the easy provision and adjustment of feedback and provides a constructive basis for arbitrarily rich multi-modal interaction depending on the complexity of the model and the quality of the sensors used. We also emphasised the benefits of taking into account constraints in our interaction design and gave a suitable example.

Initial system testing showed us that the system was intuitive enough to be learned quickly. All participants understood and could use the system with only a few minutes practice, which is encouraging. Although the use was not perfect for any of the users these tests provided us with some very interesting and intriguing usability and physiological insights as to how our model based approach to this kind of interaction actually coped with real people and provided an insight as to how this kind of approach should be modified to minimise these problems. We found that each user tended to have their own comfortable posture, which emerged after only a few minutes of practice, indicating that any system adopting this kind of approach would need to be personalised somewhat. We also found that users were particularly susceptible to hand drift, which tended to cause a number of false positives and we found that the participants were less adept with forward flicks than backward flicks, which is perhaps intuitive given our understanding of human physiology.

In the following chapter we expand our location-aware interfaces away from the egocentric body space and into the broader spaces of the exocentric world.

Chapter 4

Whereable Computing

4.1 Summary

This chapter demonstrates the construction of a novel exocentric location-aware, eyes free system. The use of Monte Carlo propagation for browsing virtual environments is introduced, which takes the natural constraints of the local environment and uses them to aid navigation. It is shown that the feeding back of uncertainty to users during a target acquisition task can improve their performance. The guidance of users around a predetermined density or trajectory is demonstrated and finally we show that it is possible to produce a simple model of human behaviour, which can mimic behaviour in this trajectory following task.

4.2 Introduction

Global Positioning Systems are increasingly being integrated into standard mobile devices such as PDAs, handheld gaming machines, watches, and mobile phones. The Nokia N95 and Samsung SCH-V850 already come with built-in GPS. GPS can be unreliable at times since there are frequent problems with spatial resolution, latency and signal shadowing, which may all be detrimental to navigation systems. This, coupled with the user's lack of knowledge of an area in which they are navigating may, in the worst case, render their system unusable. There are a number factors which contribute to the inaccuracy of a GPS, including atmospheric effects, ephemeris errors and multi-path effects. For a description of these errors and the GPS system in general the reader may refer to appendix A.

As an example we may think of a user equipped with a handheld satellite navigation system. When the user enters a built-up area with high buildings the GPS signal becomes increasingly shadowed, the number of satellites visible is reduced and the system becomes increasingly inaccurate. At some point the user will be given misleading info and lose confidence in the system. This problem arises because the system did not convey its increasing uncertainty to the user and instead presented this confused information as *fact* meaning that the user magnifies the systems uncertainty. One real example of this failure to display uncertainty is the grounding of the Royal Majesty on the 9th of June 1995. This was a direct result of the crew relying on a GPS navigation system which showed apparently accurate information despite not having accurate measurements ((National Transportation Safety Board 1995), (Degani 2004)). What the crew did not know is that their GPS system was actually operating in dead-reckoning mode, and was accumulating error rapidly as the ship travelled. However, the GPS continued to display the position as *fact*. As a result, the ship ran aground on rocks. It is for this reason that we seek to introduce probabilistic, multimodal displays with the appropriate display of uncertainty and with the user engaged in a continuous negotiation with the system.

We apply these ideas to the GPS navigation problem on our mobile device, demonstrating a probabilistic approach to navigation using a combination of GPS and general inertial sensing. The incorporation of techniques from control and probability theory allows us to embrace the omnipresent uncertainty, providing a more flexible and usable system. It has been shown that the introduction of goal-focused predictive displays to an interface, with appropriate calculation and display of the outcomes, may actually improve control of the system. Smith (1997) gives a rigorous explanation of the importance of maintaining uncertainty in nonlinear prediction problems and examines methods which aim to maintain uncertainty rather than adopt unsubstantiated conclusions. This is not just of interest to technical systems. There is significant, well-controlled experimental evidence (for example, the work of Körding and Wolpert. (2004)) that display of uncertainty leads to regularised control behaviour in human motor control, in reaching actions and targeting actions. If this can be generalised to broader interaction scenarios then it suggests that uncertain displays have the potential to 'smooth out' the interaction process and make use of a inherently uncertain system less frustrating. In this chapter one of the aims is to investigate this hypothesis using a location-aware audio system, which uses an implementation based on Monte Carlo propagation for browsing a virtual environment.

Since the inclusion of GPS in hand-held mobile computers and mobile phones is a relatively new phenomenon, the ways in which we may use this new functionality have not yet been fully explored. Is the kind of guidance with discrete and precise instructions we see in a motor vehicle really appropriate for a person navigating by foot in a much more open and less constrained world where it is easy for a user to stop an browse to regain their bearings? In this situation we feel it is much more appropriate to *persuade* the user that they should move in a certain direction using subtle cues and alterations to their comfortable state rather than force them in certain directions with obtrusive and unsettling commands. The second part of this chapter therefore, will focus user traversal around a set trajectory using the same notion of uncertain display described previously. We show that it is possible to persuade a user around a set path by simply adapting the music they are listening to. We investigate the limits to which this is possible and how varying the width of a trajectory affects user behaviour. Finally we demonstrate that it is possible to build a simple model of this observed behaviour and use this model to mimic the behaviour observed in our experiments.

4.3 Monte Carlo Propagation For Browsing Virtual Environments

The[1] novel interaction feature of our *gpsTunes* system is the browsing interface which allows us to actively probe the locality. This is achieved by

[1] The work in this section was conducted in conjunction with John Williamson at the University of Glasgow and appears primarily in (Williamson *et al.* 2006).

projecting possible paths into the future from our current location along the current heading. Of course, since the sensed state is noisy, and any prediction introduces further uncertainty, the eventual outcomes form a density over the area we are exploring.

Ideally, an estimate of the user's potential future locations would be represented as a probability density function over the navigable space, taking into account likely movement areas, sensor noise and obstructions. This function, however, is extremely complex for non-trivial, i.e. real-life, landscapes, and no solution of a simple form is available. Instead, it is possible to approximate using a set of samples drawn from the density. It is much simpler to draw such approximating samples than it is to directly evaluate it, and the technique lends itself well to the subsequent display of the probabilistic information in a particulate form, such as granular synthesis. Details of Monte Carlo methods can be found in Chapter 29 of (MacKay 2003). For example a visual display may consist of a point cloud overlaid on a map; goal-directed auditory analogues of this process are described later in this chapter.

For the *gpsTunes* browsing task, a simple algorithm for sampling future possible trajectories is as follows:

- Draw samples $x^0 \dots x^S$ from a distribution ϵ around the current state. This distribution represents the sensor uncertainty at the initial position (e.g. from the shadow maps described later).

- For each step t until some horizon T:

- $x_t^s = x_{t-1}^s + h + l(x_t^s) + \sigma(x_t^s)$ where $\sigma(x_t^s)$ represents the model noise at the new point x_t^s (Gaussian, in our examples), and $l(x_t^s)$ represents the derivative of the likelihood map at that point. h is heading specified by the user. $\sigma(x_t^s)$ can be a constant value or a more complex function; e.g. from a map indicating the resolution or quality of the likelihood map.

- Display the samples x_T^s

This is somewhat similar to the *Hamiltonian* (or *hybrid*) Monte Carlo sampling process; Chapter 30 of (MacKay 2003) has further details. In our implementation, our inertial sensing platform is used to control this scanning, obtaining a heading from the magnetometers to produce h and controlling t via vertical tilt, as measured by our accelerometers. Physical location is estimated via the GPS. Intuitively, this process can be imagined as a beam of particles flowing out from around the initial state, probing into likely destinations as in figure 4.1.

4.3.1 Likelihood Maps

If we were to perform a straightforward propagation of particles through a featureless space we would create a fairly simple distribution of points at the time horizon, which would be unlikely to model likely possible user destinations effectively. It is extremely unlikely, for example, that a user will be inside the wall of a building at any point in the future. To represent these varying positional likelihoods we use a simple likelihood map, giving a probability p of being in a particular position (as measured by the sensors) in the mapped area. An example of such a map is shown in Figure 4.1; in this example the buildings have very low likelihood and there is increased likelihood around pathways on the map. In this case, the map is generated by hand from an existing map, but such likelihood maps may also be generated automatically from digital photogrammetry maps, for example.

In the simplest case the propagation algorithm can be modified to take account of this likelihood map simply by removing particles at a rate inversely proportional to their likelihood given their position. However, our implementation modifies the dynamics of the particles such that they are deflected away from regions which are less likely. This causes the samples to "flow" across the surface by following the derivatives of the likelihood map producing a browsing system that channels Monte Carlo samples towards regions of increased likelihood, following traversable paths and avoiding obstacles in a natural manner.

It is obviously simple to extend this technique to multiple likelihood maps which may be combined based on context variables. We can imagine the scenario where a user of the system has two different behaviours, one walking and one riding a bicycle. The likelihood map for a user walking

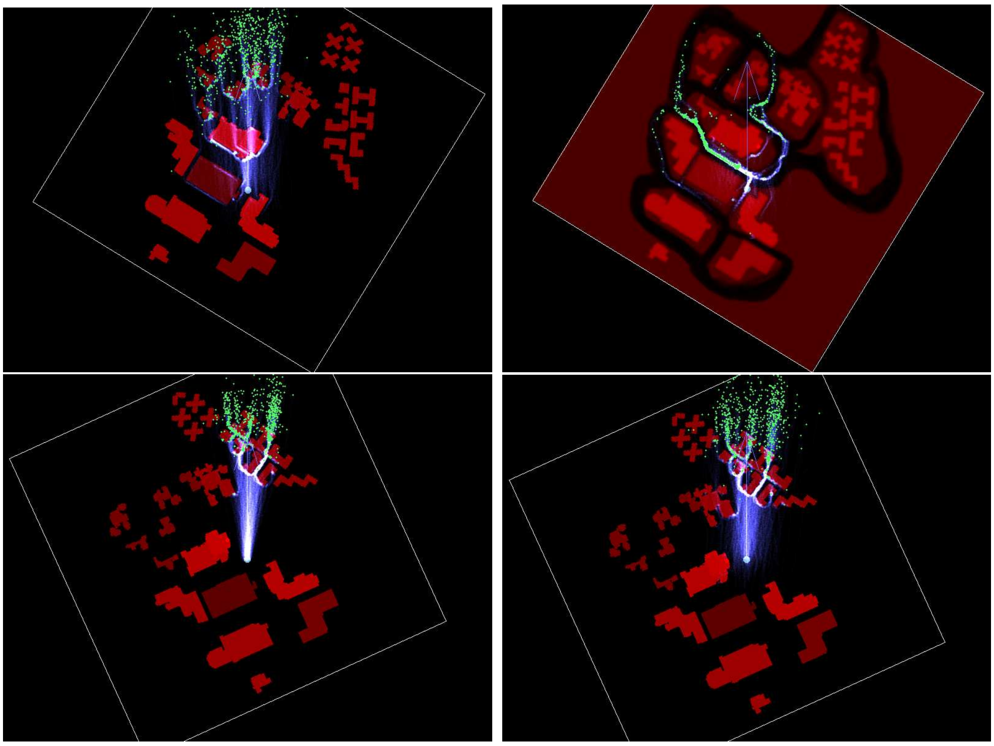


Figure 4.1: Particles flowing around the campus likelihood map. Higher red values indicate lower probability. The particle paths are illustrated in blue; the samples at the time horizon are highlighted as bright green dots. From top left to bottom right: Top Left: shows particles on likelihood map which is a model for walking behaviour. Top Right: shows the effect of a more constrained map which models a user on a bike, where particles tend to flow along available paths. Bottom Left and Bottom Right: show the effect of the GPS shadow map on the propagation; Bottom Left is a point outside of shadow, while Bottom Right is a nearby point with heavy shadowing. The increased dispersion is obvious.

around an area would be much less constrained as they are far more likely to walk off of the main paths in that situation where as the likelihood map for a user riding a bicycle would be much more constrained as they are far more likely to stick to the main roads and paths in this case. Figure 4.1 shows an example, where suitable likelihood maps for walking and cycling behaviour are shown. A relatively simple context detection method, using our system's sensors, can then estimate the probabilities of these possible alternatives, and combine these maps to produce a single output map incorporating context information.

4.3.2 A Priori Sensor Uncertainty Maps

One further problem with our naïve propagation algorithm is that it takes no account of the varying uncertainty in sensor measurements, especially the previously mentioned atmospheric effects, ephemeris errors and multi-path effects affecting the GPS signal compounded by spatially varying uncertainty arising from shadowing in our local environment. Such maps can be constructed ahead of time given knowledge of the geometry of potential occlusions (for example see (Steed 2004)).

In our working system we simply used a static map of the local area where buildings are given a low probability and everywhere else given a high probability since we did not have a detailed knowledge of satellite positions at this point. But it is possible to construct static occlusion maps for use in our platform with a raytracing technique based on currently locked satellite positions, which provide us with the knowledge about the potential shadow positions. The resulting sensor uncertainty map for our test region is shown in Figure 4.2.This map may be included in the sampling algorithm by modulating the diffusion parameter ϵ at each time step by the calculated sensor uncertainty at the point. The total sensor uncertainty is then a combination of the map input and accuracy in the reading produced by the GPS device itself.

The accuracy of a GPS fix is also computed in the sensor hardware in real-time. This includes the number of satellites which have locks and other data giving the fix quality and the "horizontal dilution of precision". This horizontal dilution of position gives a scaling factor for the current uncertainty from 1–50. These may be combined with the *a priori* sensor maps to obtain a certainty estimate for the current location.

It would theoretically be possible to improve the accuracy of these maps by comparing GPS readings with the likelihood maps described in the previous section; readings suggesting positions of low likelihood decrease confidence in the current veracity of the sensors. Additionally, we have assumed simple Gaussian diffusion in our spread model, which while a reasonable approximation, could be improved by diffusing particles proportional to the likelihood at their *new* positions (effectively Metropolis-Hastings sampling

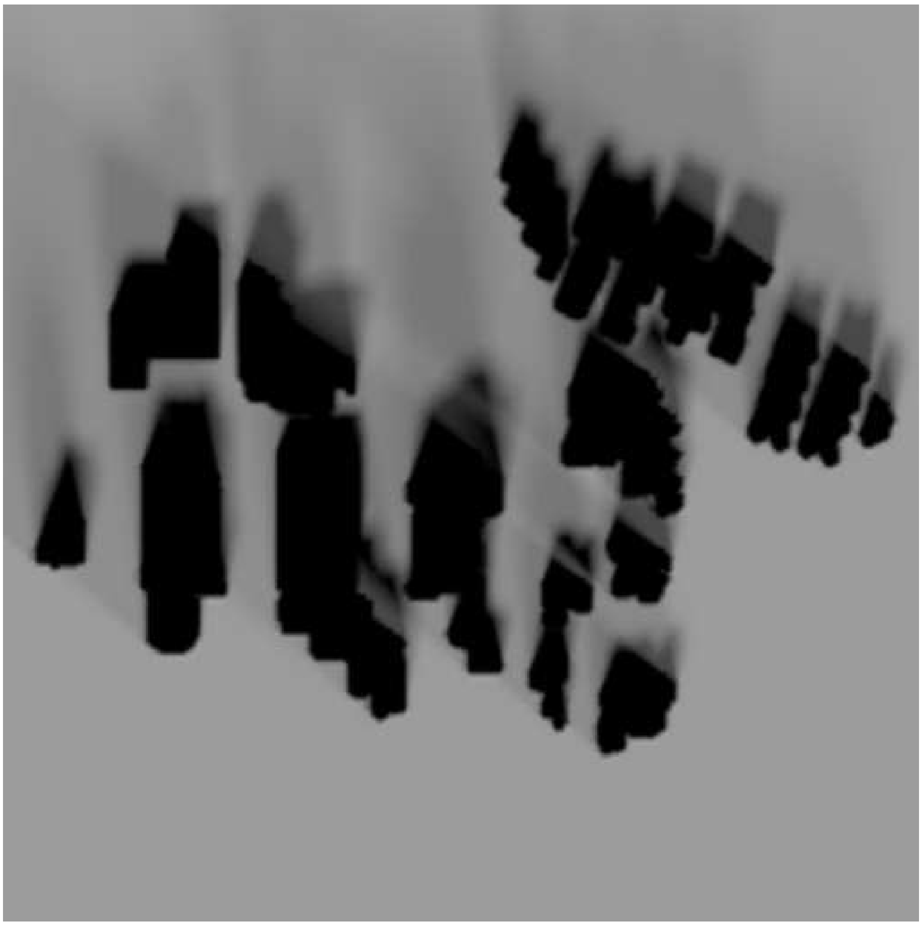


Figure 4.2: Estimated GPS shadows for the test campus region. Shadows are computed via a raytracing algorithm, based on satellite azimuth/elevation and an estimated height map for buildings in the area. Darker regions have less satellite coverage.

(MacKay 2003)).

4.3.3 Variable Time Horizon Prediction

One way in which a user may interact with our navigation system is via the direct manipulation of the prediction time horizon. The interactor can use this to probe further into the future or bring their particle probe in close to examine nearby objects. In particular, this allows the user to experience how the uncertainty in the potential goal space changes. It provides an answer to the question: do all possible movements in this direction inevitably converge to some likely goal or do they spread out rapidly to a multitude of potential targets? This feedback directly informs the user as to how much effort they will have to expend in scanning the space in the future.

In our implementation the Monte Carlo time horizon is controlled via vertical tilt (sensed by the accelerometers in the MESH hardware), by analogy to artillery fire illustrated in figure 4.3. Higher tilt levels project the

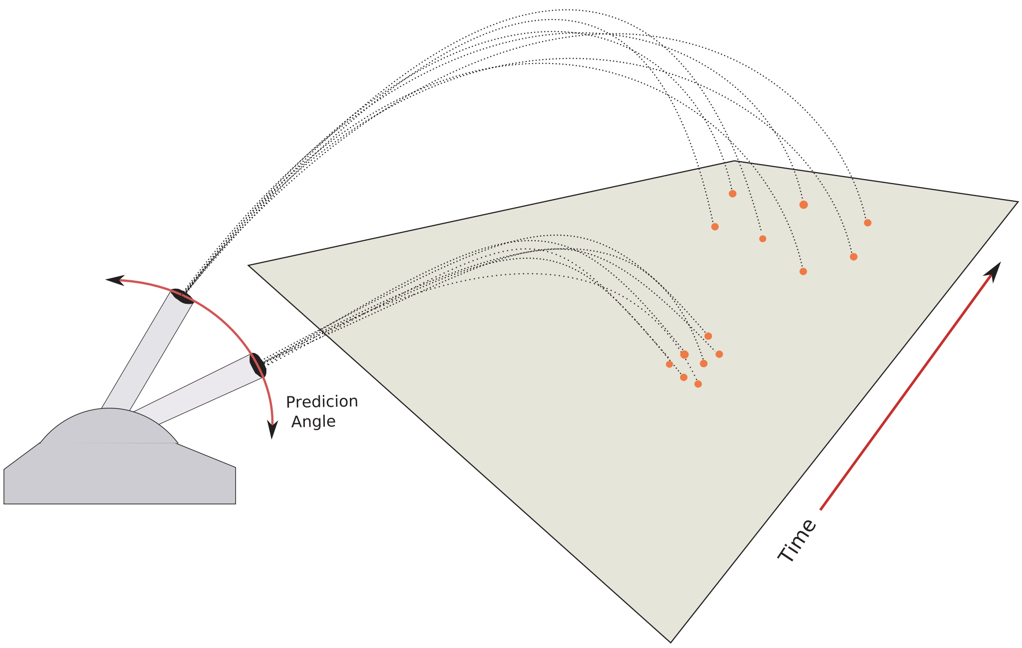


Figure 4.3: Artillery fire as an analogy to adjustable time horizon Monte Carlo prediction horizons. Higher angles have greater range (in time) but increased diffusion.

particles further into the space, with correspondingly greater uncertainty (depending on the model). A tilt back looks into the future and a tilt forward brings us back to the present. There has been debate in recent years on the particular mappings to use for this particular task. Should a tilt forward not let us look into the future and a tilt back bring us back to the present? We chose this particular configuration as the analogy to artillery fire was easy for users to grasp quickly. Our system has a 20m look ahead so at maximum extension (maximum tilt is 60 degrees in this version) the user will be hearing feedback from 30m ahead. At minimum extension they will be hearing feedback from 2m ahead. So we effectively have a function which allows users to look ahead in time and receive the feedback from that point in time in order to inform them if their current heading will require to be changed in the near future. The intention is that this aspect of our system will support users as they traverse the trajectory.

4.4 gpsTunes: Navigation By Audio

To our knowledge, the *gpsTunes* (Strachan *et al.* 2005) system was the first of its kind implemented in a truly hand-held, real-world situation. *gpsTunes* is a novel application with a location-aware element combined with a classic mobile application, the music player, and allows us to navigate in unknown environments via audio and vibrotactile feedback. The system has two modes of operation.

4.4.1 Non Probabilistic Approach

The first mode takes a simple non-probabilistic approach. This version of the system was designed to guide a user to a desired target by varying the volume and 'bearing' or direction of the currently playing song. So, for example, if a user enters an area with which they are not familiar and they wish to locate their desired building, they may inform the system of where they wish to go with a click of a map, which will then alter the volume level and bearing of the music being played. They then attempt to move towards the sound source keeping the music in front. As they move closer to the target, the volume of the music will increase, reaching the maximum

Figure 4.4: User holding the PocketPC and MESH in hand.

(user preferred) volume at the point where the target has been reached. At this point they will be notified of their arrival by an additional pulsing sound played over the current track. When building a system such as this the two most important pieces of information to convey to the user are the *distance from their desired target* and the *current direction, relative to targets* (Holland *et al.* 2002). In this mode the distance is conveyed by a change in volume. A Gaussian density is placed around the chosen target, and this is mapped to volume of the sound source. The music switches to the lowest audible volume on the edge (a threshold value) of this distribution. As the distance to the target is decreased the volume increases back towards the users preferred level. The direction of the current target is conveyed to the user by panning the sound 'source' around their head using a stereo shift in the audio. When the user clicks their desired target, the bearing to the target is calculated using the current GPS estimate of latitude and longitude. Using the heading calculated from the calibrated magnetometers in MESH allows the system to pan the sound source to the correct position, from the user's perspective. The user can rotate on the spot and hear the sound source effectively moving round their head.

4.4.2 Probabilistic Approach

The second mode of operation for this system uses the probabilistic approach described in section 4.3, in order to guide the user to where they wish to go. Using this approach a user may locate or *acquire* their target in the local area by probing and examining the locality. Panning to ascertain the correct direction and varying the Monte Carlo time horizon in order to gain a feel for the distance. Using this approach the user may guide themselves to their desired location or 'target' listening for impact sounds which represent Monte Carlo particles impacting with a target as illustrated in figure 4.9. It is also possible in this configuration to guide a user to their desired location along a set trajectory or path, using a density based approach, with the music they are listening to being adapted in a positive or negative way depending on whether the they are on or off the correct path.

4.5 Target Acquisition

Two different trials were conducted in the course of this work, both of them involving the acquisition of targets using our *gpsTunes* system. The first trial was an informal look at the effects of adding uncertainty to the display. The second trial was conducted in a more controlled environment indoors where participants were required to stand still and scan for targets placed around them by varying their bearing.

4.5.1 Granular Synthesis

In the target acquisition trials, a granular synthesis technique is used to display the output samples. Granular synthesis for probabilistic display is described in more detail in (Williamson and Murray-Smith 2005*b*). Each particle is displayed as a short audio impact sound drawn from a selection of waveforms (each goal has one set of distinct source waves). These sounds are drawn from samples of a number of real, physical impacts (e.g. wood, glass, water, etc.) and vary in timbre. In the Monte Carlo case described here, each grain is associated with a sample, and the likelihood of activation with a particular waveform is given by the proximity of the sample to the goal in the location space. More precisely, we define a distribution f_i around each goal i. This set of distributions is used to transform the physical space into the goal space, and the probability of activating a sample grain is given by this distribution. The goal densities are Gaussian in the target acquisition prototype. The particles can be thought of as impacting on the target densities; the value of the target map at which they impact modulates the volume of their presentation. This produces a continuously changing auditory texture which represents the total distribution of particles in the goal space. The sound has a flowing impression which varies from sharply defined audio at low uncertainty or low entropy to a vaguer mixture of sounds at increased entropy.

4.5.2 Outdoor Field Trial

The aim of this initial field trial was to test the hypothesis that a truthfully uncertain display can improve navigation in an environment with high sensor noise. In these trials, five participants were asked to find four different targets (in physical space) using only the audio and vibrotactile information presented to them, in an outdoor navigation task. In one case they were presented with an uncertain, dispersed audio display as illustrated in figure 4.9 and in the other they were presented with display of the mean only (i.e. without any uncertainty) as depicted in figure 4.10. The audio in both cases was augmented with a simple vibrotactile display, in which a short pulse was produced every time an audio grain was rendered.

As the GPS signal in this area was strong, noise typical of that in an occluded environment was artificially introduced to the GPS sensed position. This noise consisted of a random positional offset (of the order of a few metres), updated once every five seconds. The time horizon was fixed at this point to reduce the complexity of the task.

Method

A within-subjects experiment was used; each participant performed both versions of the trial and the experiments were performed outdoors on the university campus. All participants started in the same position, facing the same direction and were given a three minute introduction and demonstration of the system. The first target was selected and the user determined the direction of the target by panning the device until they heard the audio cues and felt the vibrotactile feedback indicating that they should move in that direction. When the participant came sufficiently close to the target the next target was presented and the same procedure repeated. The area traversed is illustrated in figure 4.5 along with the targets through which the users were required to pass.

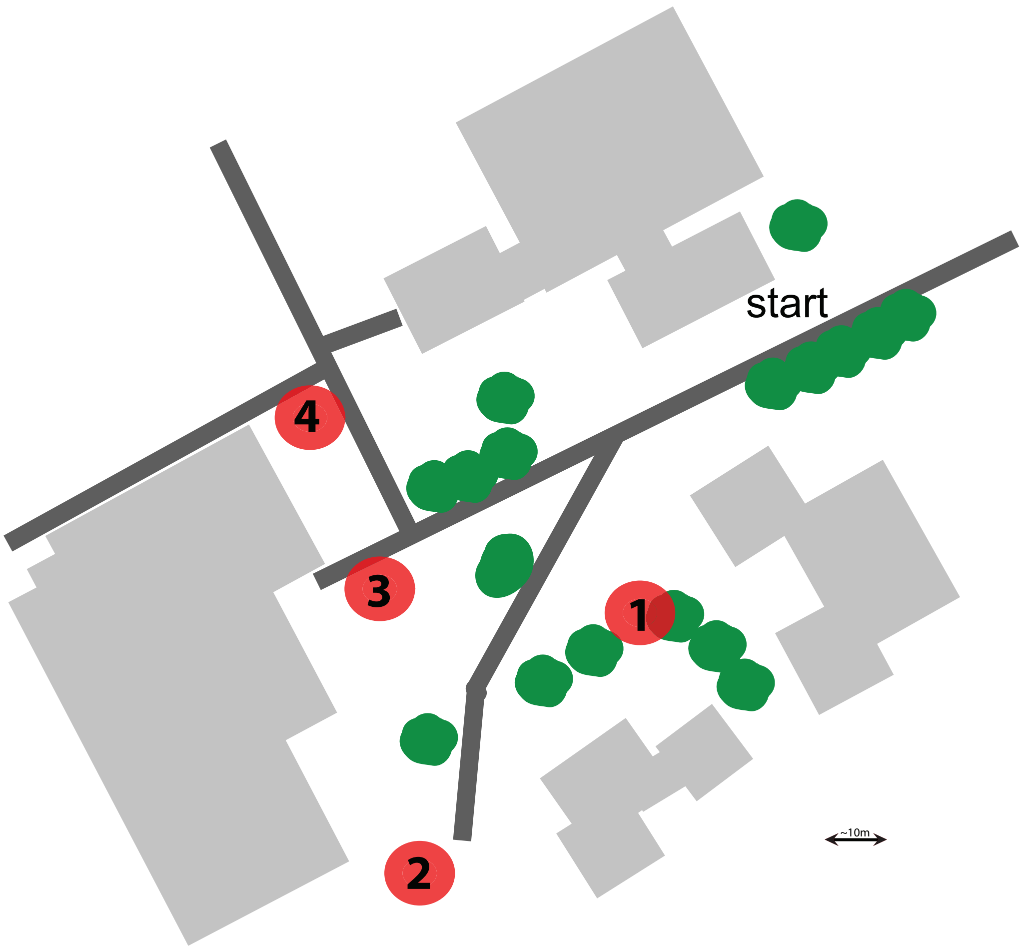


Figure 4.5: Schematic map of the area used for the outdoor field trial with four different targets indicated by numbered red circles to be found. The buildings are indicated by the grey blocks and the green circles are trees.

Results

As a general measure of performance, Figure 4.6 shows the time taken to complete the task successfully for each user. Time to complete the task is generally reduced when the display with accurate representation of uncertainty is employed. One reasonable hypothesis is that less effort should

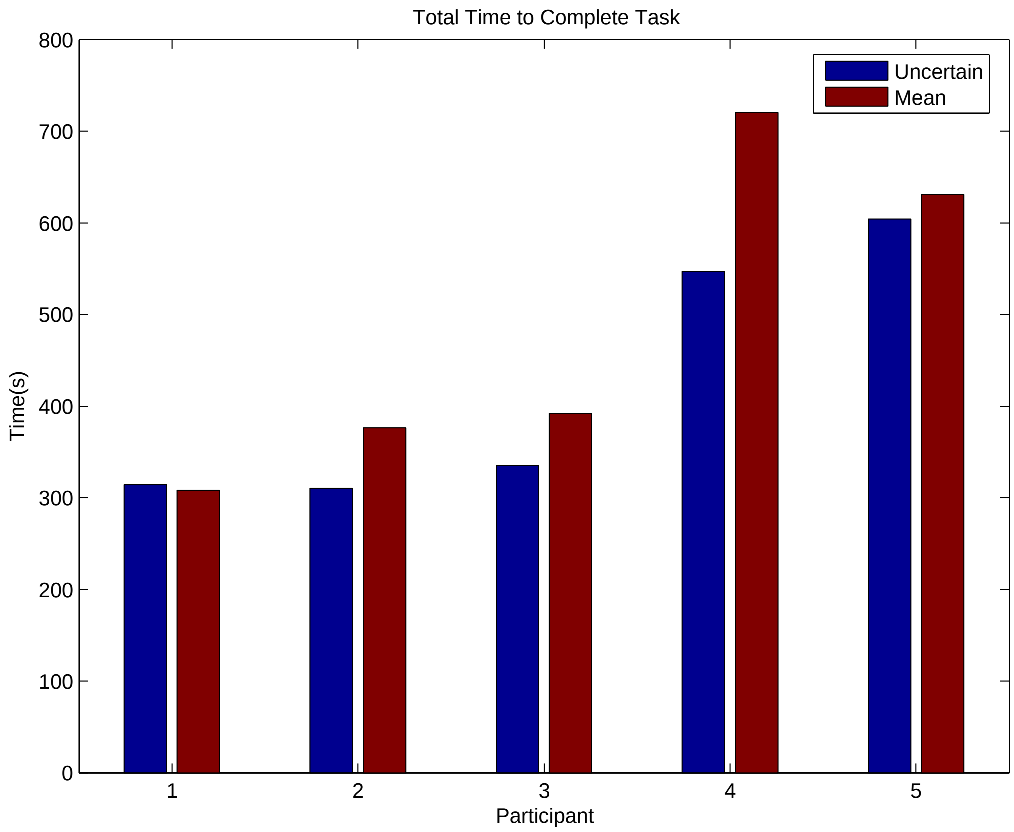


Figure 4.6: Completion times for the task for each participant. The task is completed in a shorter time for the uncertain case, except for the first participant, where the times are very similar.

be expended by a participant in searching for the target when the uncertain display is employed. The mean squared derivative of the bearing signal (i.e. energy normalised by time taken) gives an indication of the effort expended by the user; Figure 4.7 illustrates the values of the metric for each participant and condition. These results indicate that there was a large reduction in the scanning effort required by all participants in the uncertain-display case. Similarly, the mean squared derivative of the bearing signal (i.e. energy normalised by time taken) gives an indication of the effort expended by

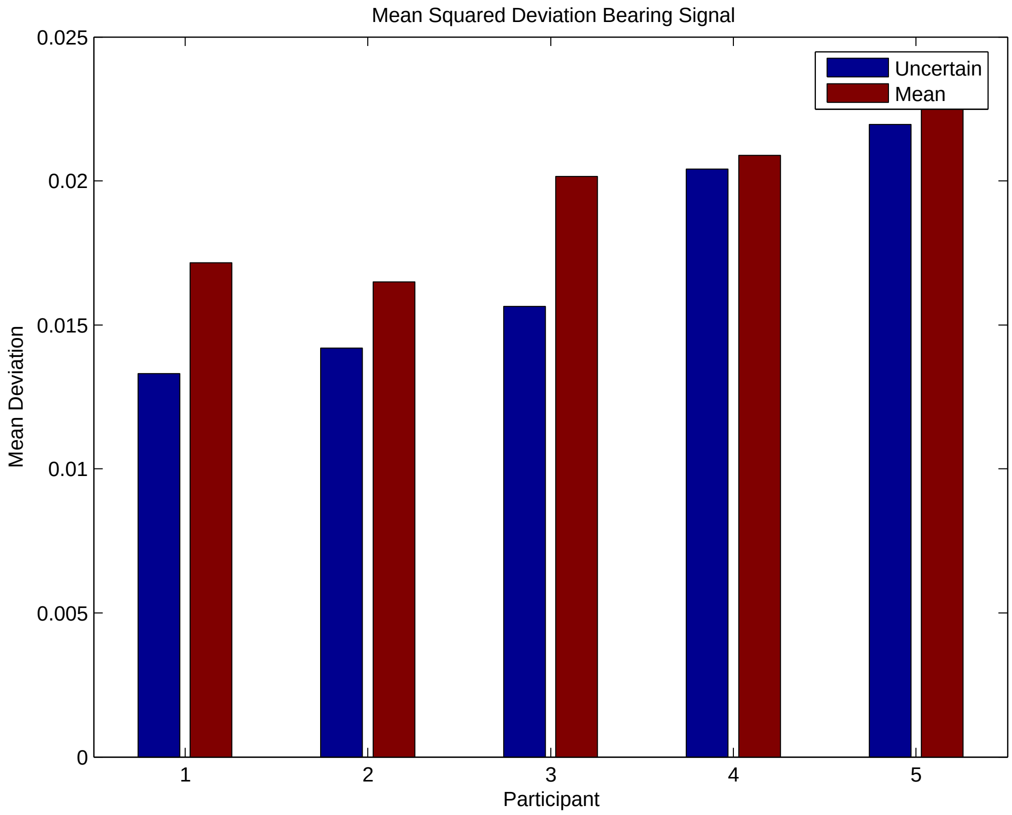


Figure 4.7: Mean deviations of the bearing signal $\frac{1}{T}\sum_0^T \left(\frac{d\theta}{dt}\right)^2$. Smaller changes are made by participants in the uncertain case.

the user; Figure 4.8 illustrates the values of the metric for each participant and condition. As in the total energy case, effort appears to be reduced.

Comments from Participants

Some informal comments were elicited from participants at the end of the experiment. Many of these concerned the apparent latency of the display, which seemed confusing until participants got a feel for it. One participant commented that they had difficulty gaining a feel for the dynamics of the system because of this delay. Another commented that the vibrotactile feedback sometimes seemed "more responsive" than the audio feedback.

Observations

From observing the participants it was clear that on first use, there was a lot of confusion in both the mean case and the uncertain case. However, after a short period of time, the participants rapidly acquired skill in using

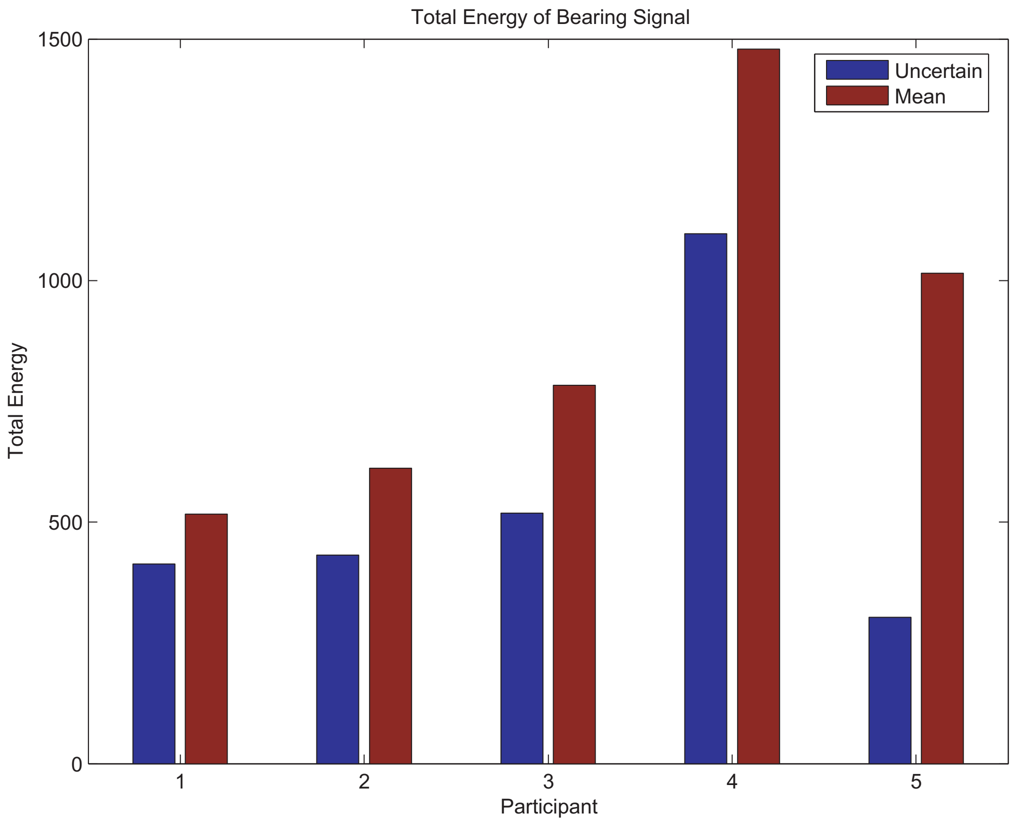


Figure 4.8: The total energy of the (unwrapped) bearing signal. Less scanning energy is expended by users in the uncertain display case than in the mean-only display.

the system. It was also apparent that participants had significant problems with the mean-display case since this resulted in more jumping of the audio. Once they had determined the direction of the target, sudden jumps in the audio signal proved confusing (Figure 4.14 illustrates the effect of a jump in sensor position causing a participant to begin more vigorous scanning). One participant in particular stopped every time there was a discontinuity in the feedback and then readjusted his position before moving on.

4.5.3 In A Virtual Environment

Since these results from the informal outdoor field trial provided encouraging evidence for our hypothesis, an experiment was conducted to study the effect of the uncertain audio display in a more controlled environment. In this setup, participants stood still and had to acquire targets arranged around them by scanning the space around them with the device as illustrated in figures 4.9 and 4.10. The participants were not required to move at all during the experiment and target acquisition occurred when their measured heading remained sufficiently close to the heading of the targets for a certain period of time. No GPS signal was used; however positional noise was simulated to produce the effect of a poor quality GPS fix.

Four cases were examined in the experiment: mean display, without additional noise; mean display, with additional noise; uncertain display, without additional noise; and uncertain display with additional noise.

Experimental Details

Five targets were laid out in the space, as illustrated in figures 4.9 and 4.10, each of which had to be acquired three times for each condition (fifteen acquisitions per condition). Acquisition was considered to have occurred when participants maintained the heading measured by the device within a funnel of 14.03 degrees for 5.4 seconds. Leaving this zone caused the countdown timer to pause until the participant re-entered the capture zone. The targets were arranged in an arc from $-\pi/2$ to $\pi/2$, at a distance of approximately 71 metres and the target positions were fixed throughout the trial. Sporadic noise (Gaussian distributed with 9m standard deviation) was used to shift the position of targets in the noisy cases. Noise occurred as steps updated every three seconds, resulting in a square wave like pattern similar to that of true GPS noise. Heading data was filtered with a low-

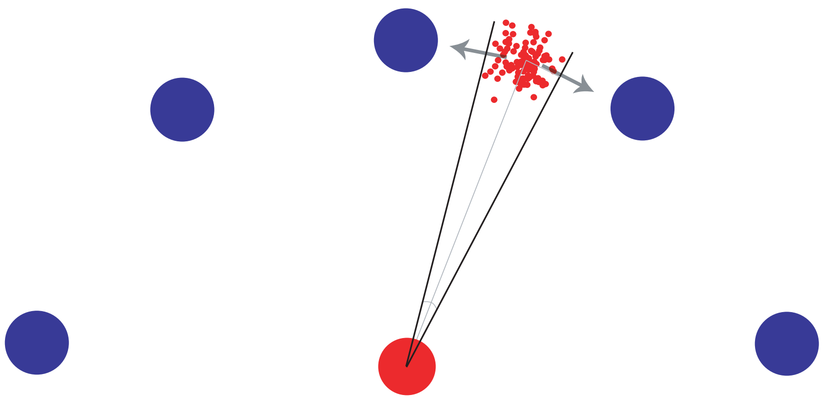


Figure 4.9: Targets are arranged in a semi-circle around the static user for the uncertain display case.

pass filter, with -3dB rolloff at 8Hz before being displayed and recorded.

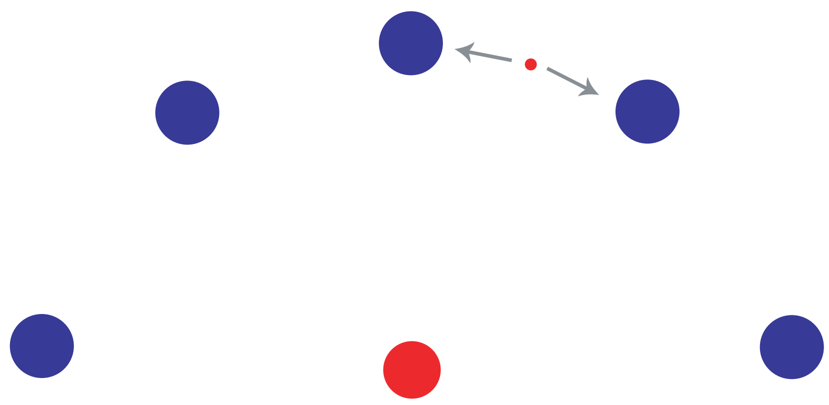


Figure 4.10: Targets are arranged in a semi-circle around the static user for the mean display case.

This eliminated most of the tremor signal (8-12Hz) from the sensed signals. The heading data and acquisition times were recorded. The experiment was within-subjects, with a counterbalanced presentation order and eight participants took part in the experiment.

Results

The mean case, for both the noise and no-noise cases, generally requires more time for acquisition than the uncertain display. Figure 4.12 gives a boxplot of the mean time for each acquisition (per participant), illustrating the distribution of timing in each of the cases. Figure 4.11 shows the energy (in the low frequency 0.1–2 Hertz band) for each condition and participant. There appears to be some reduction in scanning activity in this band for some participants, although the acquisition criterion may have lead to successful capture even without significant feedback, leading to anomalous cases where less energy had to be expended. Figure 4.13 shows a boxplot of the variance of the error between target heading and device heading. There is a significant reduction in the uncertain case compared to the noisy case. Large deviations from the target are less likely when the uncertain display is employed.

Figure 4.14 shows a typical time series from one participant for the mean and uncertain (with noise) conditions. There is noticeably more searching activity in the mean case, where the participant overshoots the target and has to search back. Figure 4.15 shows the histogram of error (for the same

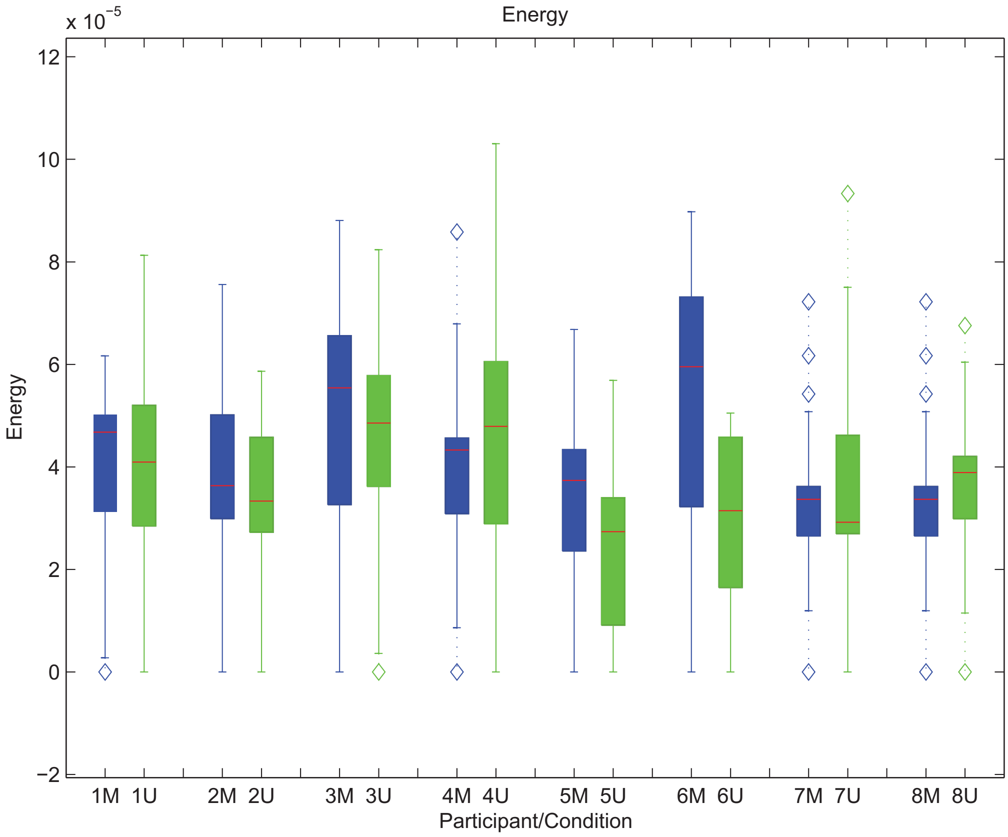


Figure 4.11: Low frequency energy (0.1-2Hz) in the heading signal for each condition. Boxplot shows distribution of energy for each acquisition (blue=mean, green=uncertain). Energy required is reduced for six of the eight subjects with the uncertain display. Energy is computed as $\frac{1}{T}\sqrt{\sum_{t=0}^{T}\left(\frac{dx}{dt}\right)^2}$.

participant) in the region after the error has been reduced by 63%, for both the mean noise and uncertain noise cases. The mean noise case leads to a distribution of error with heavier tails (more variation during the final stages of acquisition). This is compatible with the variance of error plots in Figure 4.13.

4.5.4 Discussion

The results support the hypothesis that the uncertain display requires less effort and results in more stable behaviour. However, the results would have almost certainly been stronger had the selection mechanism been less susceptible to "random" selections. The capture zone for acquisition was over-generous in this experiment, under-penalising the mean case. Sub-

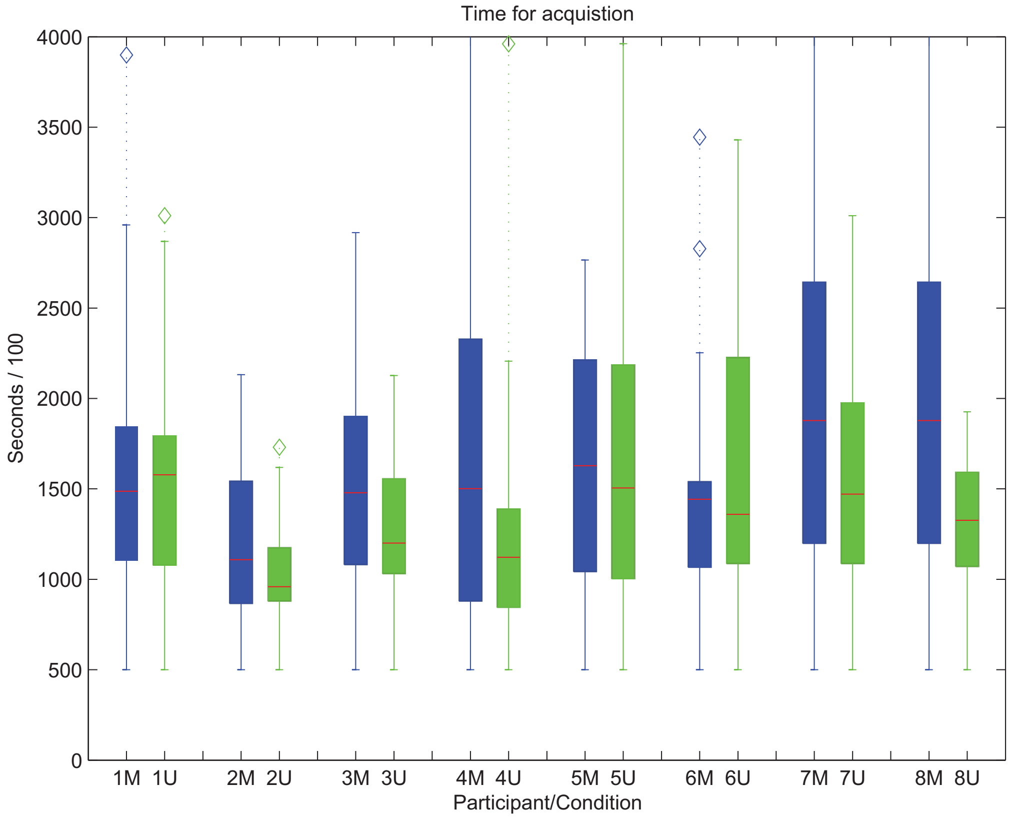


Figure 4.12: Boxplot showing the target acquisition times in each case. Mean time to complete is reduced in seven out of the eight cases with the uncertain display.

jective comments from participants suggest that they felt the targets were larger in the uncertain case than the mean case. They also apparently felt less "in control" in the uncertain case, despite performing better under these conditions. Some users (participants 7 and 8 in particular), when asked after the task, felt that they had not performed well and were confused but, in reality, looking at their results they appeared to have coped well. This is unsurprising given the unfamiliarity with ambiguous displays but does suggest a need for a careful choice of metaphor for uncertain interaction. Participants also noted no change in difficulty between the uncertain case where noise was applied and where no noise was applied; however they noted that the mean case was significantly harder when the artificial noise was applied.

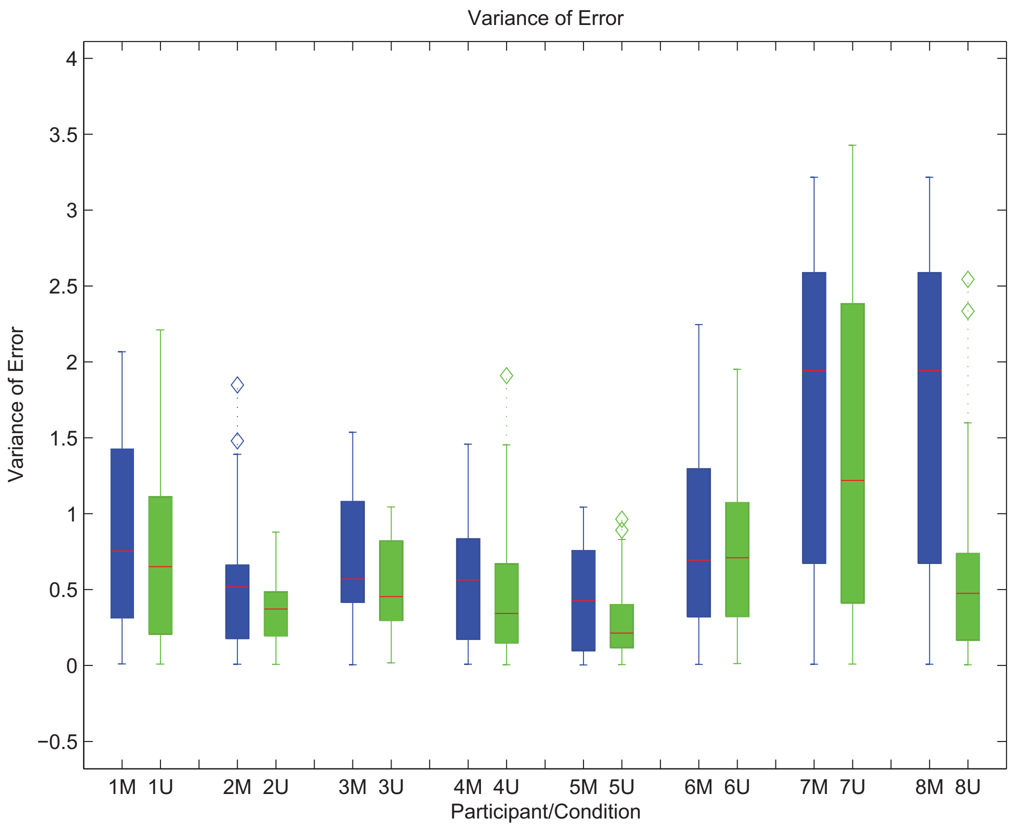


Figure 4.13: Boxplot showing the variance of the error during acquisition, for each condition. There is a visible reduction in the variability of the error with the uncertain display; large deviations are less common.

4.6 Trajectory Following

The notion of being guided to your destination is something intuitive for human beings. In this section we discuss a part of our system, within the *gpsTunes* framework, which utilises a combination of GPS, inertial sensing and Monte Carlo sampling and modulates a listener's music in an unobtrusive manner in order to guide or *persuade* them to a desired physical location through a density. In this case along a set trajectory or path.

Trajectory following is something we usually associate with robotics and autonomous controlled vehicles. In a very basic way, robots utilise their sensors in order to update their current state via complex control systems and trace out a desired trajectory with varying degrees of success. What if we wish to control a human and guide them from a starting position, along a trajectory to their desired location? How can we achieve this and what kind of behaviour should we expect from them?

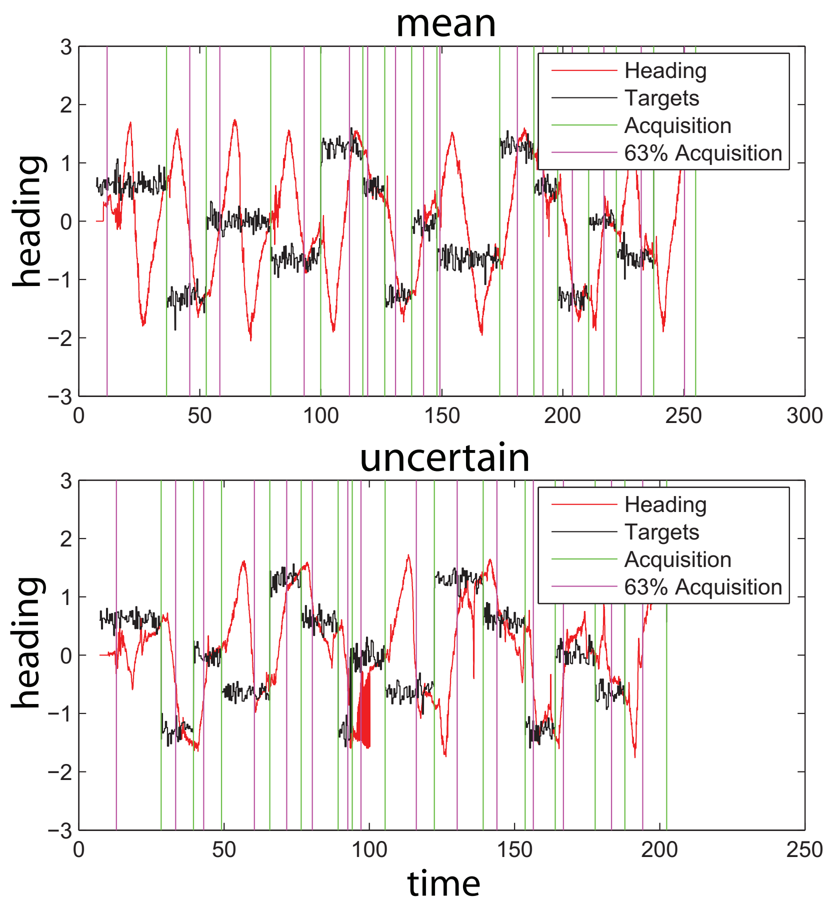


Figure 4.14: Heading time series in the mean and uncertain noisy cases for one participant (3). More scanning behaviour is visible in the mean case.

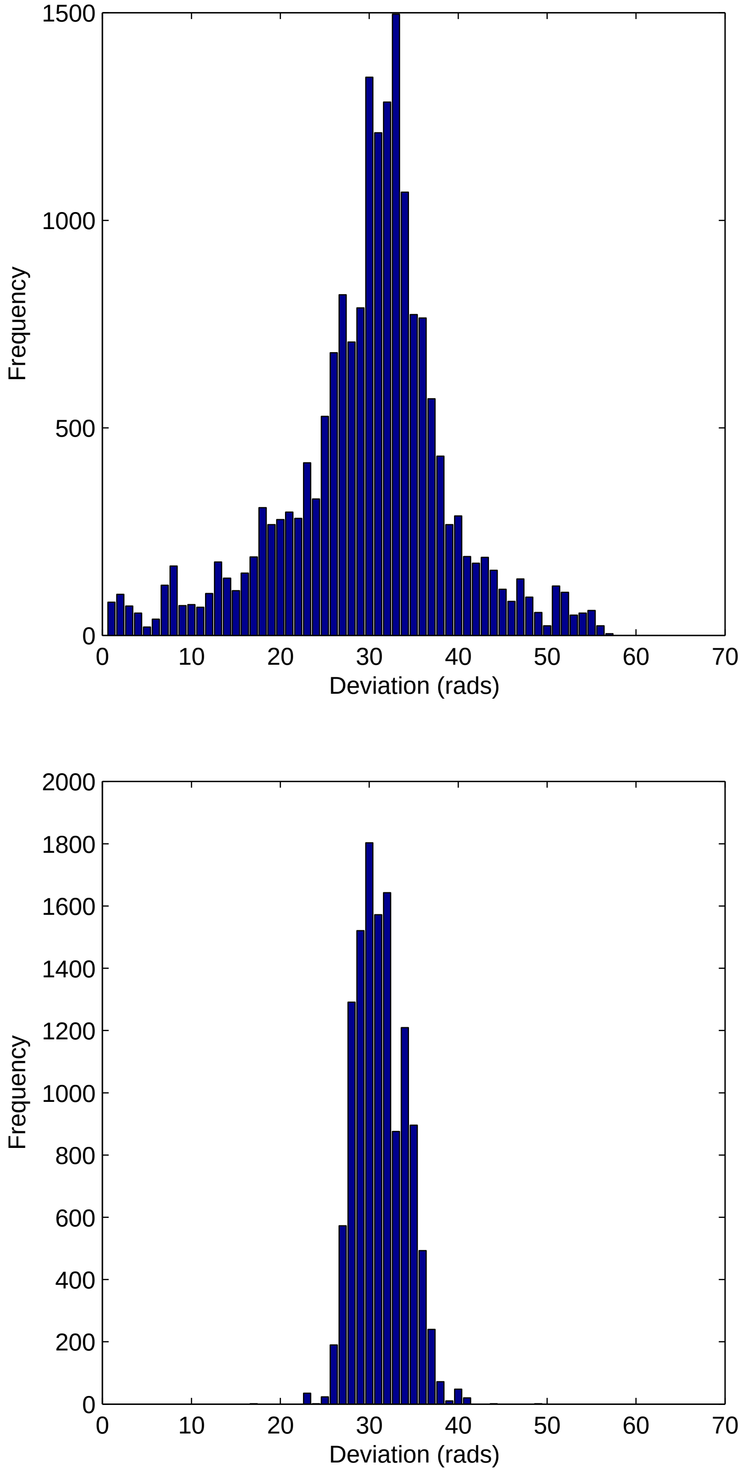


Figure 4.15: Histogram of error in the mean (top) and uncertain (bottom) noise cases for one participant (3). Larger deviations in error are more common in the mean case.

This kind of system has a number of different applications. One of the main applications is to mobile guides. It is possible to guide a user from the beginning of a tour to the end along the exact path which takes them through all locations of interest. Sports training is another obvious application. Using this system it is possible to record a runner's path both spatially and temporally. On subsequent runs it is then possible to convey this information to a user so that they have an idea of how they are doing compared to their previous personal best run. If they were behind their personal best run they would hear footsteps ahead of them from their *ghost runner* which they would need to increase their speed to catch, with the variation between front and back being displayed in an appropriate way.

4.6.1 Our Application

Our trajectory following application is a part of the overall gpsTunes system described previously. It is designed to guide a user along a desired trajectory using audio and vibrotactile feedback via their mobile device, be that a PDA, music player, mobile phone or Ultra Mobile Personal Computer (UMPC). If a user is traversing from one point to another in an area with which they are not familiar, there may be an optimal trajectory to that point or a trajectory which avoids any potential hazards. In this situation it is up to our system to guide the user through this preferred path.

The desired trajectory is represented by a density or uncertainty map (described later) layered on top of a map of the local area, as in figure 4.16. Monte Carlo propagation is then used for browsing this density map, which allows us to actively probe the locality by projecting possible paths into the future from some location along the current heading, enabling us to predict likely positions of the user at future time points. If the user, at a postulated future position, has strayed from the correct trajectory, this information may be fed-back to the user so that they may update their plan. Monte Carlo sampling is, exactly as in the target acquisition task described previously, used to predict likely positions of the user at future time points.

Figure 4.16: Example trajectory overlaid on a map. Although there are at least two different routes the user could take from the start point to the end point they will only receive crisp clear musical output if they stay on this black trajectory from beginning to end.

4.6.2 Feedback

Feedback in this system consists of both audio and vibrotactile. In this case the audio feedback consists of a distortion of the musical waveform. The distortion takes the form of a reverb effect which is modulated by the likelihood of the user being on the path at the time horizon. This is computed by summing the values of the likelihood map at the Monte Carlo sample points (the red dots in figures 4.18 to 4.25) to estimate the overall probability of being on the path at the horizon, $v = \sum_0^S \tau(x_t^s)$, where τ is trajectory probability density function. This value is used to modulate the reverb parameters such that a low probability of remaining on the trajectory results in increased reverberation and this is also mapped into vibrotactile feedback so that a low probability of remaining on the trajectory results in a 'stronger' vibrotactile feedback. This gives the user some sense of the changing probabilities without completely destroying their musical experience. Moving off the path produces echoing and muddy sounding output; sticking closely to the path produces clean, crisp sound.

4.7 Forming Trajectories

To gather some objective evidence from our empirical studies on the effects our system has on user behaviour it is important that we make the correct choice of trajectory. So what factors do we need to consider when designing for such an experiment?

- **complexity**: The complexity of a trajectory is important since we all have our limits. Will an overly complex trajectory lead to a complete loss of control and a highly frustrated user? And will an overly simple trajectory really tell us anything at all? Also, what are the effects of the plausibility of the perceived complexity of a trajectory to the user? Is there a threshold to what a user will tolerate?

- **width**: We can imagine that the trajectories in our everyday life vary in a lot in width. When we are walking through a wide open playing field our trajectory may be open and wide but when we come to the small foot bridge which crosses over the river our imagined trajectory reduces in width significantly. It is easy for us to perceive this change and it makes no difference to us in real life but what if our system guides a user through an unknown trajectory from an expansive, wide area into a narrow tight area? Will the user display a tightened behaviour in the thinner part of the trajectory?

- **location**: The location of the trajectory for our empirical studies is important. If we are confined to a cluttered environment there may be a significant number of visual distracters, which may tend to concentrate the users attention in that area. There may also be a number of natural distracters, which we tend to be drawn towards, such as roads, pavements and footpaths. Users would naturally want to keep to these paths even if our system is attempting to persuade them otherwise.

So ideally our initial experimental trajectories should be in an open uncluttered environment, perhaps in a playing field with very little in the way of visual distraction and will not be overly complex. Smooth curves should be preferred to straight lines as there is very little needed in the way of control whilst walking along a straight line although both may be included. One interesting feature which should be included in a trajectory is a sharp bend or even a right angle in order to examine user behaviour as they approach this point. Will they utilise the ‘look ahead’ function more at this point? Will they overshoot at the corner and be forced back into the trajectory?

4.8 Trajectory Following Trials

An experiment was conducted to demonstrate that the system may actually be used to guide users to a desired location by a number of different users and also to examine the effects of varying trajectory width and the presence of visual distracters.

4.8.1 Method

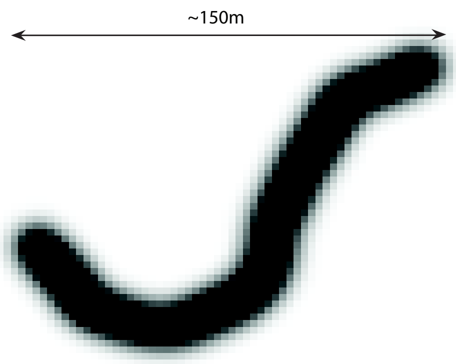
In total 6 participants took part in the experiments all aged between 20 and 29. All participants had used a mobile phone or PDA before but only 3 had any experience with GPS navigation.

Five trajectories were used in total with four of them taking the same form. The trajectory used represented a well known path from 1 place to another on the university campus as shown in figure 4.17(e). This trajectory was then translated over to a wide-open, featureless playing field and given three different widths. Trajectories one to three were identical but given varying widths and participants were required to traverse these three trajectories on the playing field. The presentation of these trajectories were given in a counter-balanced order, in order to reduce learning effects. Trajectory 2 was approximately 9m wide, trajectory one was approximately 18m wide and trajectory three was approximately 36m wide as shown in figures 4.17(a)-4.17(c). The fourth trajectory presented to the participants was a simple N-Shape which was also placed over the open playing field and was approximately 18m wide as shown in figure 4.17(d). The final trajectory presented to the participants was again the same shape as the first three with a 18m width but this time it was placed back over the campus, over paths and under trees. Before the experiment began participants were first given a 5 minute description of the system before being given a practice run to gain a feel for using the system over a relatively simple trajectory.

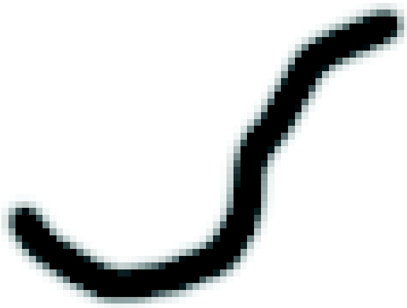
Our heading data was filtered with a low-pass filter, with -3dB rolloff at 8Hz before being displayed and recorded, eliminating most of the tremor signal (10-12Hz) from the sensed signals. The heading data was recorded along with the time, latitude, longitude, ground speed, pitch angle of the device and total uncertainty.

4.8.2 Results

The principal result from this experiment is that it is possible for this system to guide users to a set location with no user failing to reach the end point of any trajectory as illustrated in figures 4.28(a) to 4.28(e). A number of different strategies were employed by the users. Some users

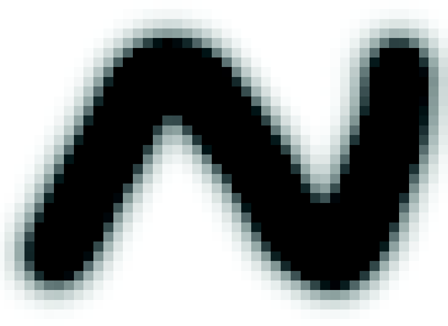


(a) This trajectory is approximately 18m wide.



(b) This trajectory is approximately 9m wide and is the narrowest.

(c) This trajectory is approximately 36m wide.



(d) This trajectory is approximately 18m wide.

(e) This trajectory is located in a campus setting and is approximately 18m wide. All other trajectories were located on the playing field in the top left corner.

Figure 4.17: All five trajectories used in the field trials.

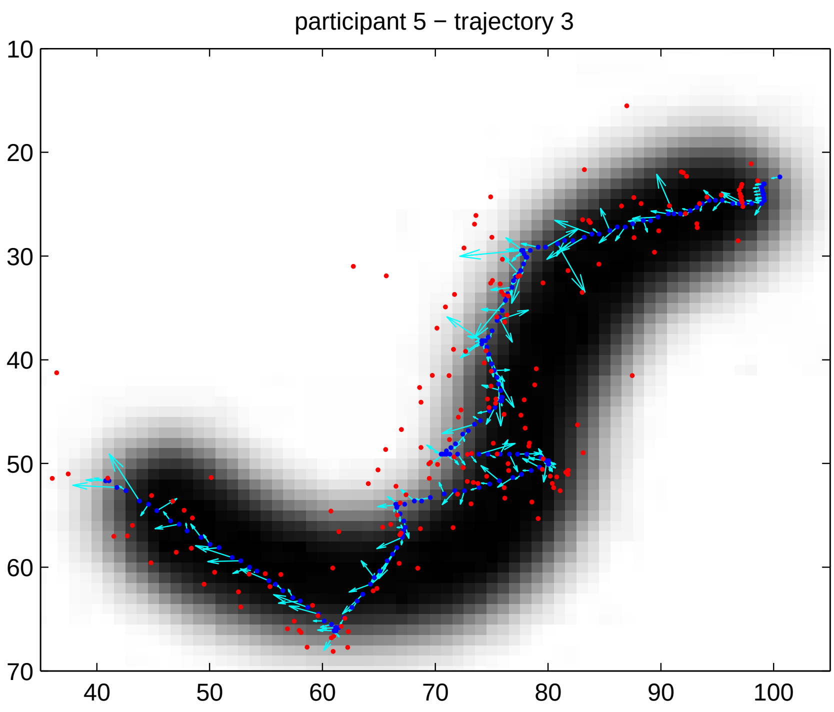


Figure 4.18: A ‘cautious’ approach to trajectory 3 by participant 5.

were highly active in probing the locality, taking a cautious and careful approach as in figure 4.18. This figure shows a quiver plot where the blue dots represent the user’s current position, the direction of the cyan arrows represents the heading direction, the length of the cyan arrows represent the tilt of the device and the red dots represent the current Monte Carlo prediction location. If these predictions are located on the white area, negative feedback is produced, if they are located on the black area there is no feedback. Other users were relatively inactive in scanning for the most part but became very active when it was required, employing a ‘straight-ahead’ approach while receiving no feedback and only scanning when they began to move off of the correct path to find another good direction leading to a zig-zagging or bouncing behaviour as shown in figures 4.19 and 4.20. Figure 4.21 gives an extreme example of this ‘zig-zag’ behaviour. One other interesting behaviour observed is when the user ‘clings’ to the edge of the trajectory, as in figure 4.24. They move along the path keeping touch with the edge, using it as a guide, reassuring themselves every so often that they are on the correct path although they are receiving poorer quality sound.

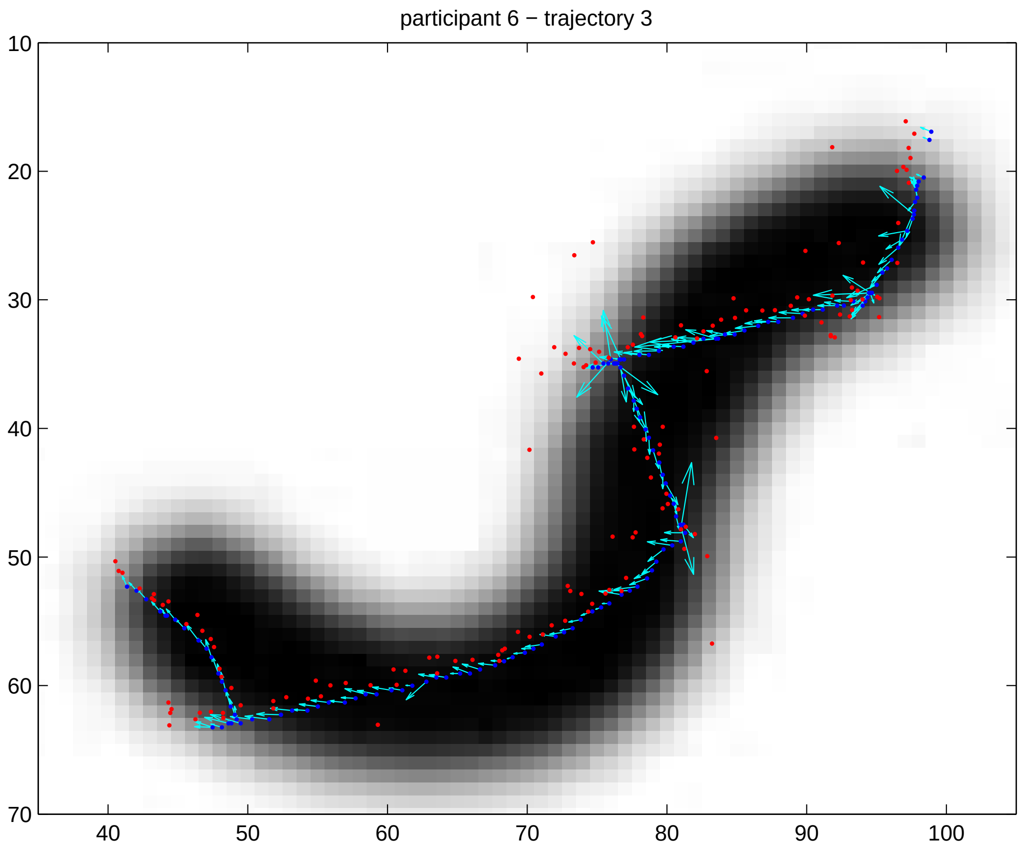


Figure 4.19: A 'bouncing' approach to trajectory 3 by participant 6.

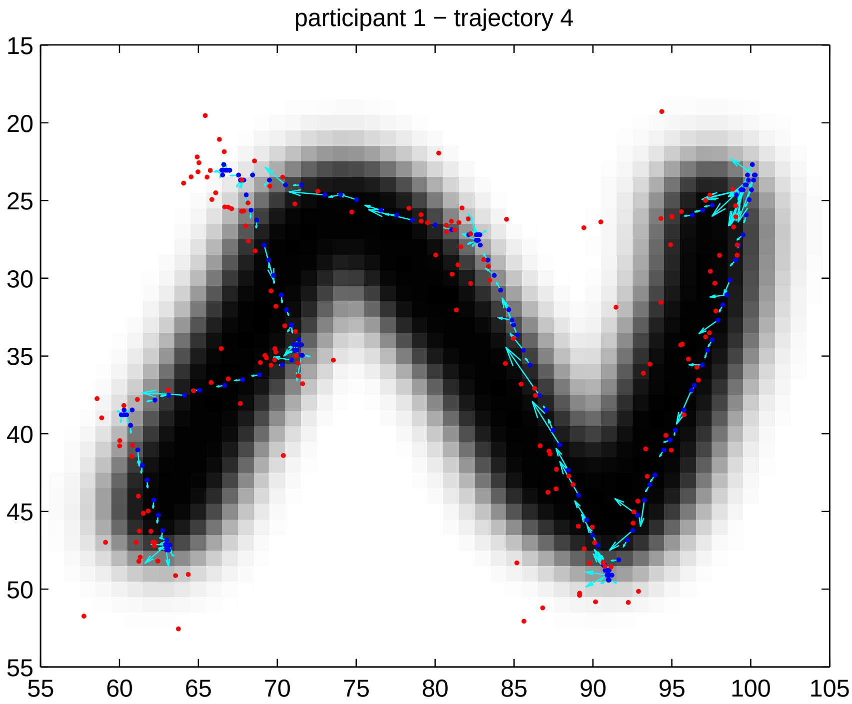


Figure 4.20: A 'bouncing' behaviour in the traversal of trajectory 4 by participant 1

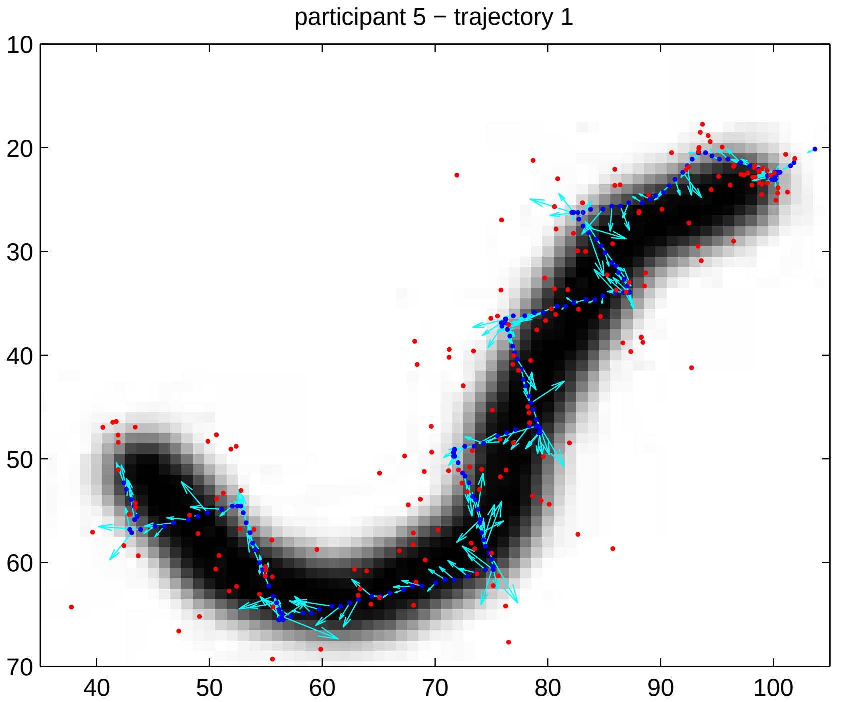


Figure 4.21: A 'straight-ahead' approach to trajectory 1 by participant 5 leading to a zig-zagging behaviour.

The inclusion of trajectory 4 allowed us to examine the effects of adding tight corners where a real understanding of the trajectory space is required in order to successfully complete the course and from figure 4.22 for completion time and from observation it was clear that users had most trouble with this trajectory. Figure 4.20 shows the path recorded for participant 1 on trajectory 4 in our field trials. This behaviour is typical and shows again a tendency to 'bounce' from the edges of the trajectory. When the user reaches the corners of the trajectory a lot more probing activity is observed in the quiver plot, since at this point the user is required to fully exploit their degrees of freedom, in order to recover the trajectory. Figure 4.27 shows the tilt and walking activity for the same example. We observe from the z-axis accelerometer data, that at the corner points in the latitude plot the user stops, then there is a burst of activity in the pitch angle, where the user is attempting to look-ahead, and a shift in the heading to the correct direction.

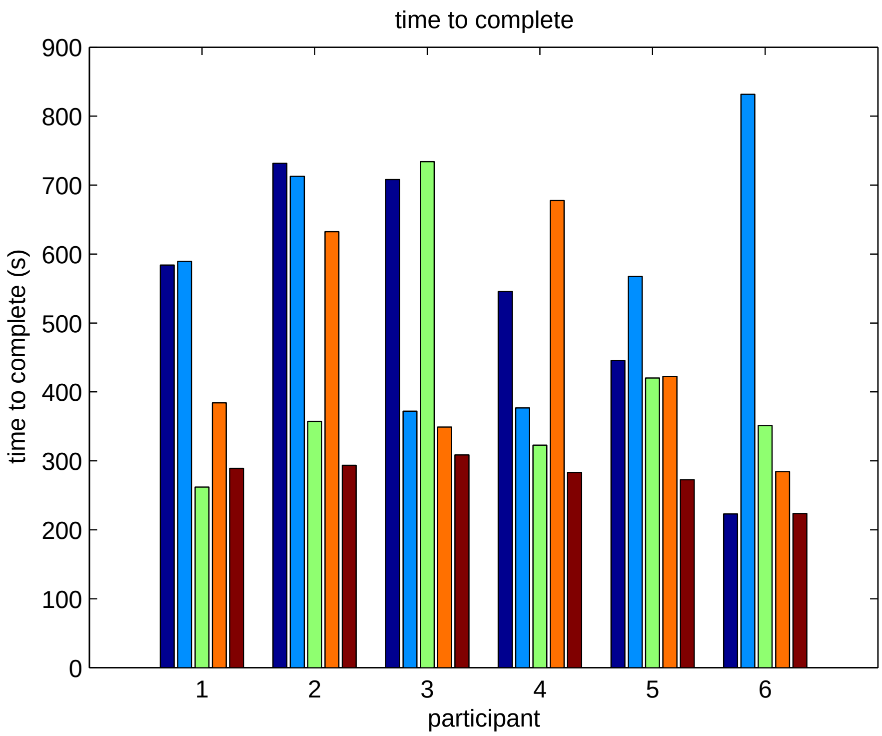


Figure 4.22: Completion times for all six participants over all five trajectories.

Looking at figure 4.22 showing the completion times for each participant,

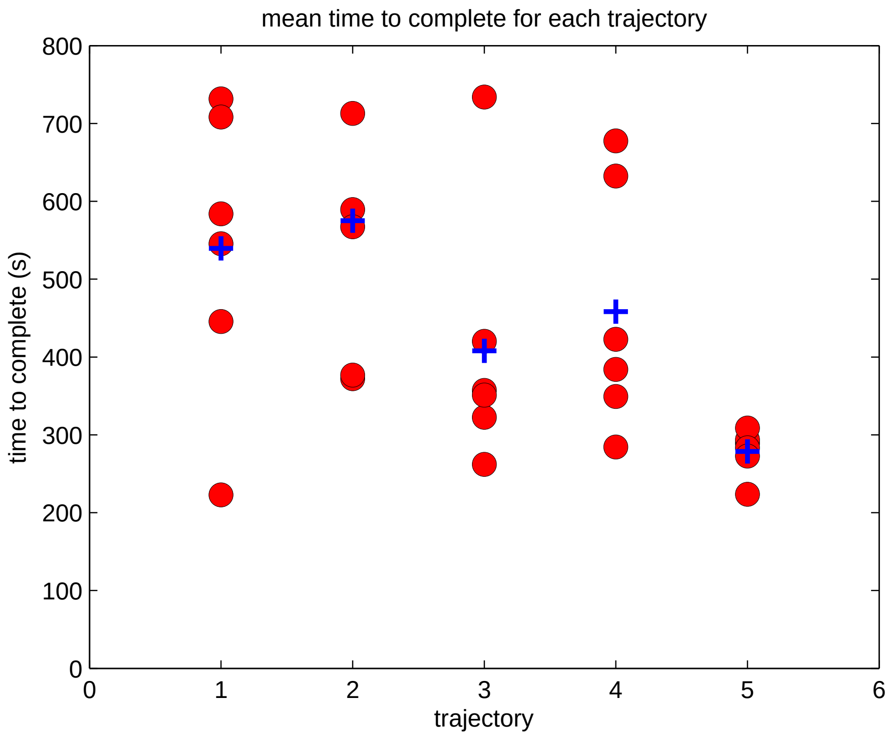


Figure 4.23: Plot of completion times for each trajectory for all 6 participants. The mean time for each trajectory is shown as a blue cross.

we see that, if we only consider the first three trajectories traversed on the open playing field at this point, the participants generally finished more quickly on the widest trajectory, trajectory 3, although comments from the user's suggested that some of them found the lack of feedback and relative freedom in this case slightly disconcerting. Figure 4.26 shows the plot for scanning energy, defined as $\sqrt{\sum_{t=0}^{T}(\frac{dx}{dt})^2}$, where x is the heading signal. This shows that users tended to scan less for the widest trajectory number 3 and most for the narrowest trajectory number 2. This is intuitive as we would expect users to react to and increase scanning immediately after feedback and in the case of trajectory 2 they are generally receiving more changes in feedback than in the wider trajectory number 3.

Interestingly, we see that the completion time for trajectory 5, from one point to another through the campus, is significantly lower than for all other trajectories, including its equivalent trajectory 1, on the open playing field. So, while we have shown that in a featureless environment like a playing field, people were able to follow the path, their performance improves signif-

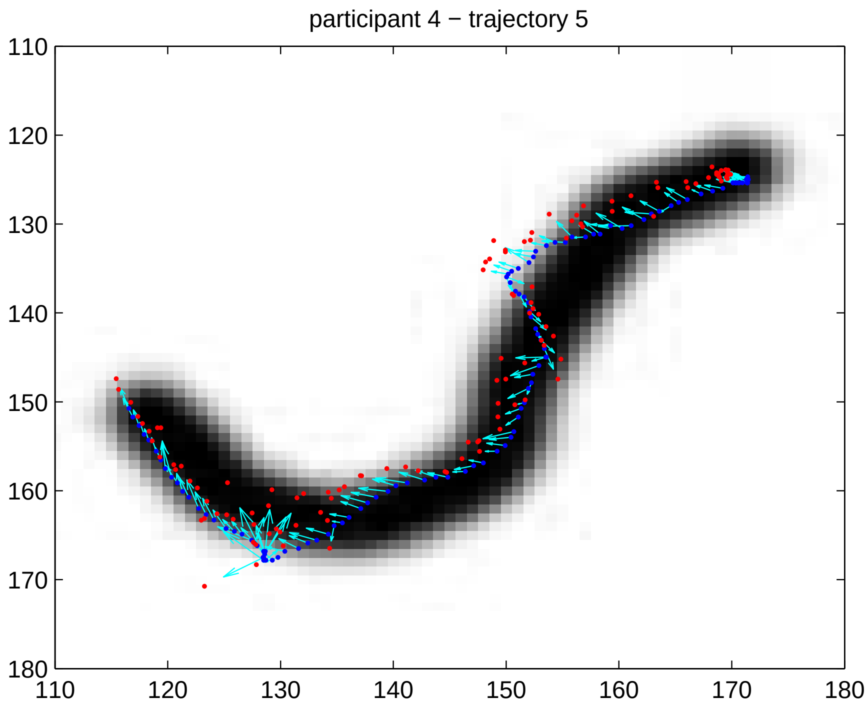


Figure 4.24: A 'clinging' approach to trajectory 1.

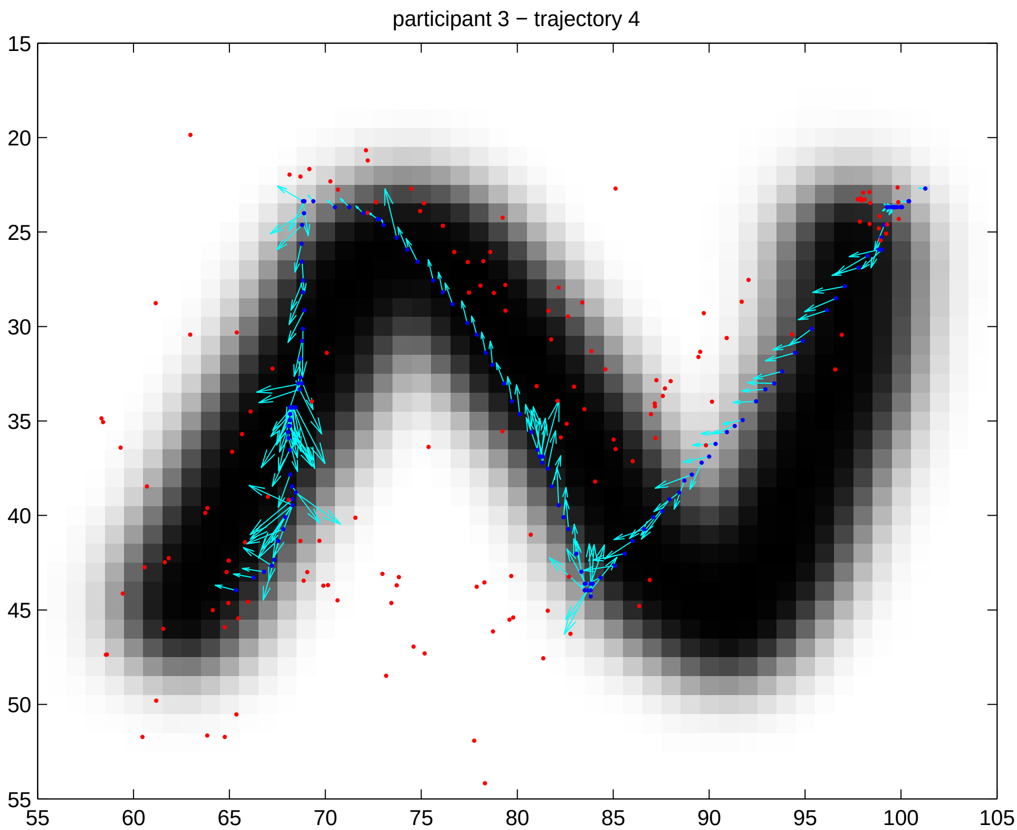


Figure 4.25: A user traversing the 'N'-shaped trajectory number 4.

icantly when the audio and vibrotactile constraints from the PocketPC are coupled with the natural constraints of the environment (paths, buildings etc). This is encouraging since most realistic use of such applications will be in settings with physically evident constraints in the environment.

Some users also commented that they found the vibrotactile feedback more useful than the audio feedback, although there was no difference in the way the feedbacks were triggered. This could be due to the on/off nature of the vibrotactile feedback (on if they were straying off of the path and off if they were ok) whereas the audio feedback was part of the music the they were listening to. It may have been difficult then to perceive small reverberations in the sound compared to small vibrational pulses.

The routes traversed for all of the participants over all five trajectories are shown in figures 4.28(a) to 4.28(e)

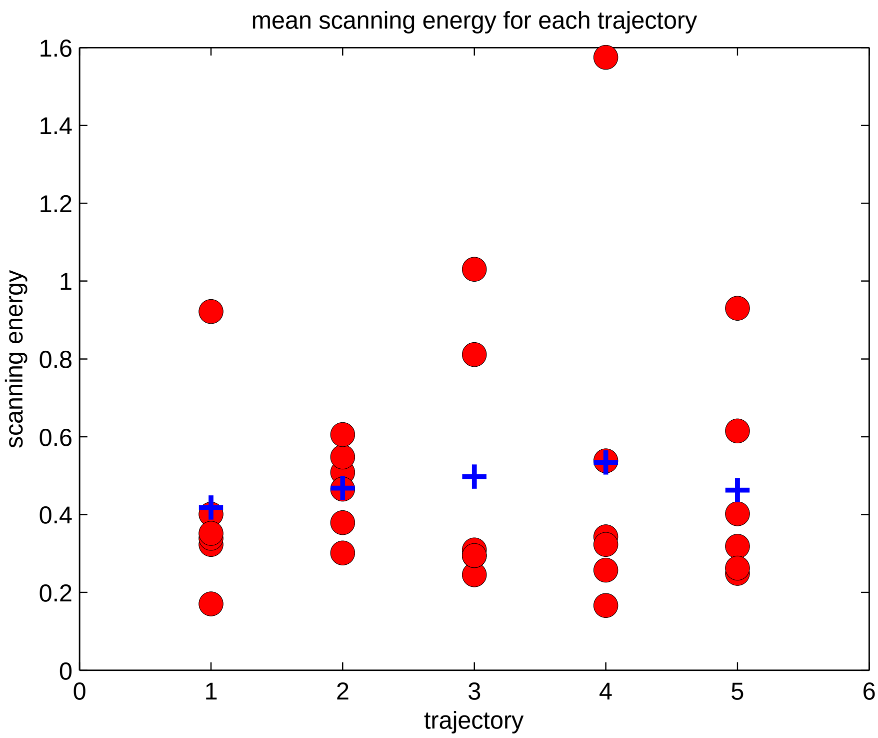


Figure 4.26: Plot of scanning energy for each trajectory for all 6 participants. Mean energy for each trajectory is shown as a blue cross.

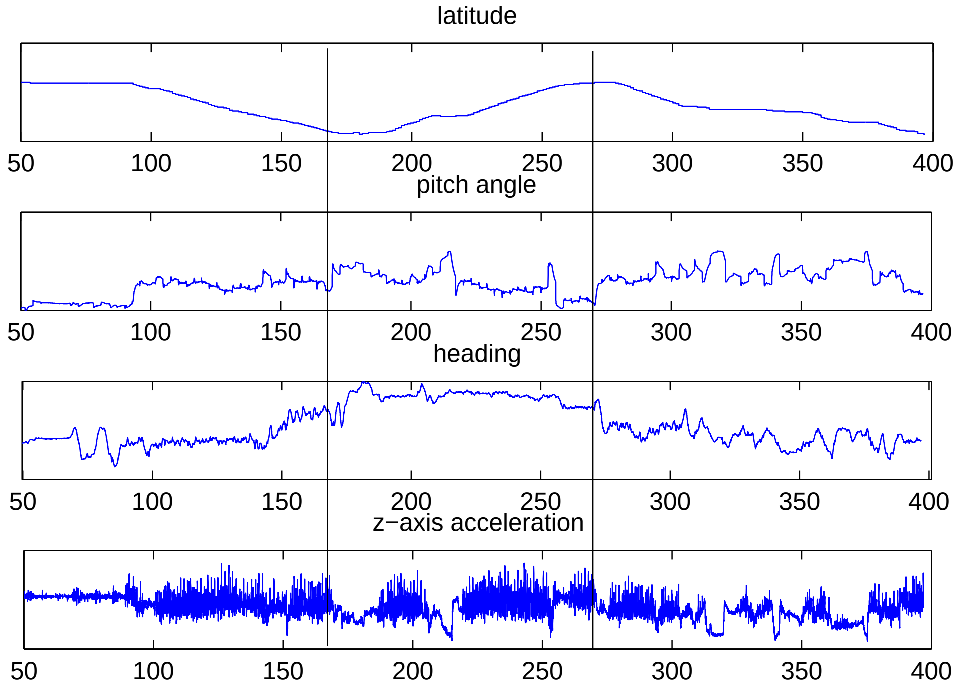
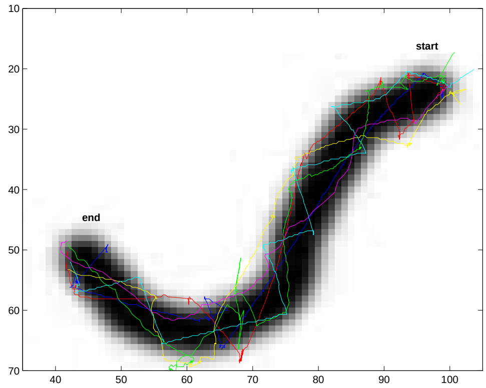


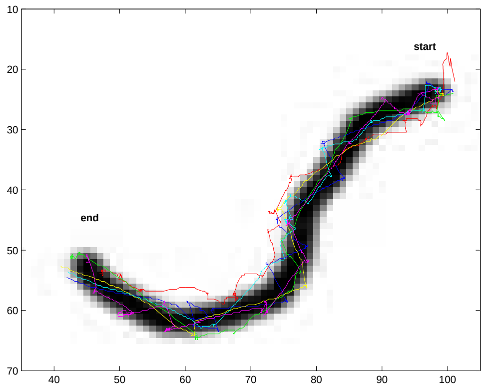
Figure 4.27: A comparison of the effect of a corner on trajectory 4. Vertical lines indicate the location of the corner.

4.9 Human Behaviour Modelling

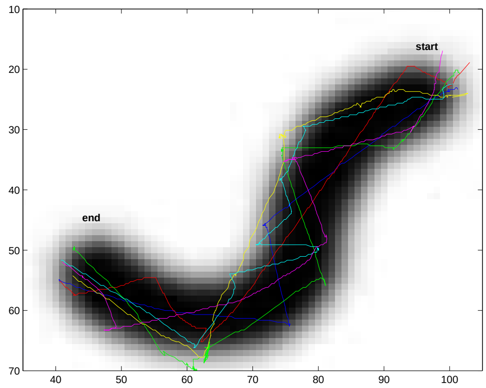
As was described chapter 2, manual control is the study of humans as operators of dynamic systems. It was realised very early by researchers in vehicle control that the human operator was a very important part of the system control loop. In order to predict the stability of the full system, they had to take into account mathematical descriptions of the human operators along with descriptions of the vehicle dynamics and this is also true for interaction design if we imagine the user's device as their vehicle, with specific dynamics. The Observation of human behaviour previously in this chapter prompts the question, can we model this specific kind of behaviour using the classical tools of control theory? There are two main reasons why this is an interesting question. First, quantitative models of human behaviour such as that described above may provide insights into basic properties of human performance (Jagacinski and Flach 2003). And second, constructing a model of this behaviour previous to conducting any experiment may give us an insight as to what kind of behaviour we might



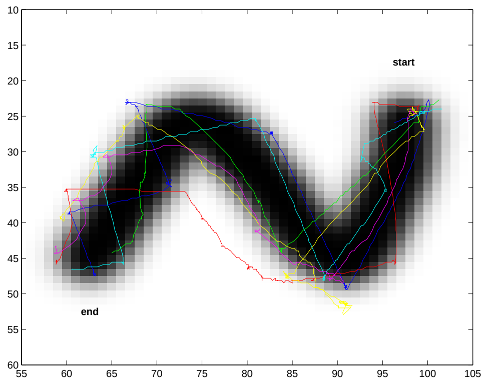
(a) The path recorded for all participants over trajectory 1.



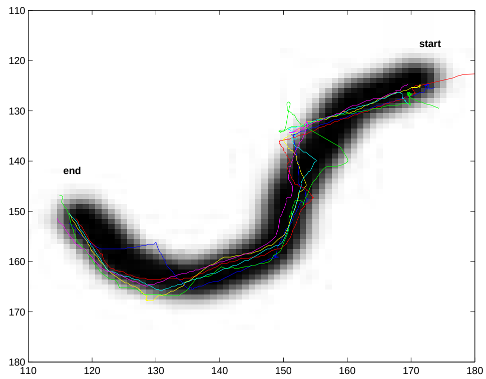
(b) The path recorded for all participants over trajectory 2.



(c) The path recorded for all participants over trajectory 3.



(d) The path recorded for all participants over trajectory 4.



(e) The path recorded for all participants over trajectory 5.

Figure 4.28: The paths recorded for all 6 participants over all 5 trajectories.

expect with our particular set up. For example, how could varying the users' look-ahead functionality in our trajectory following trial have potentially affected the their behaviour? And what kinds of behaviour might we have expected from varying the shape of the trajectory?

4.9.1 Our Model

We attempt now to construct a model which represents the behaviour of our participants. Obviously it is difficult to construct a model which perfectly describes human behaviour but it should be possible using basic assumptions to recreate the most basic behaviour.

When examining performance in a tracking task such as this, it is important that we have some measure for how well a particular participant is doing relative to another. Modern control theory provides the tools for developing models of "optimal control" that provide a basis against which to measure human tracking performance (Jagacinski and Flach 2003). When approaching an optimal control problem such as this, there are three sets of constraints that must be addressed. The *dynamic constraints*, which in this case are the dynamics of the human controlling the system. The *physical constraints*, which in this case are the constraints of our trajectory, which has a set beginning, end and width. And the *value constraints* or the 'performance criteria', which provide a value system for ranking an *optimal* or *best* path. Generally, this is defined in terms of minimising or maximising some 'cost function', which we are required to construct for our particular tracking task.

First we need to consider what the user is controlling or perceives themselves to be controlling when using the system. In a basic way the participants are attempting to traverse to the end of the set trajectory, they are controlling their position on a playing field. They are not attempting to traverse to a specific position on the field because they do not know where the end of the path is, although they are attempting to keep moving forward as instructed. The only information the user has at his disposal comes from the interface with the scan and look-ahead functionalities. What the user is attempting to do with this tool is maintain their path within the

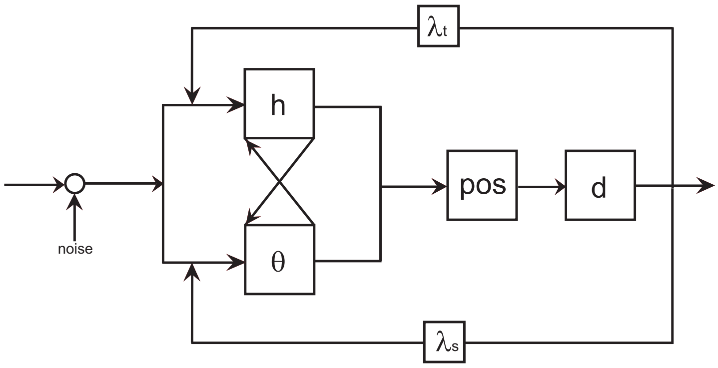


Figure 4.29: Block diagram for our model of human behaviour for the trajectory following task. θ is our heading direction, h is our level of tilt and d is the density of the trajectory at some position, indicated by *pos*

trajectory by keeping any audio or vibrotactile feedback to a minimum, i.e. minimising some cost. Users may scan around themselves checking for the best direction, with least feedback and they may look-ahead to check what will happen in the distance, again aiming to keep any feedback to a minimum.

Our cost function consists of three main parts as illustrated in equation 4.9.1. The first part is the density of the trajectory at the current level of look-ahead in the current direction, which we label d. The second part of the cost function represents the amount of scanning activity displayed by the user. We represent this current heading direction by θ. The third part of the cost function represents the amount of look-ahead activity displayed by the user and is represented as h. Some users exhibit a lot of activity and some users exhibit little activity and these characteristic differences can affect how effectively they traverse the trajectory and what kind of behaviour they display. We represent this difference in user characteristics by the parameters λ_{tilt} and λ_{scan}. Our cost function is constructed such that higher values of λ_{scan} penalise higher scanning activity and higher values of λ_{tilt} penalise higher tilting or look-ahead activity.

$$cost_t = d + \lambda_{scan}(\theta - \theta_{t-1})^2 + \lambda_{tilt}(h - h_{t-1})^2 \tag{4.1}$$

$$cost = \sum_{t=0}^{t}(cost_t/T) \tag{4.2}$$

So for example, high values for λ_{scan} and λ_{tilt} will represent the behaviour of an inactive user who does not utilise their scanning and look-ahead functionality to its full potential and low values for λ_{scan} and λ_{tilt} will represent a very active user. Within our model, illustrated in figure 4.29 we also include noise, which in this case we consider to be natural noise from walking and general arm movements. We could also include other kinds of noise such as physiological noise from tremor in our muscles, for example.

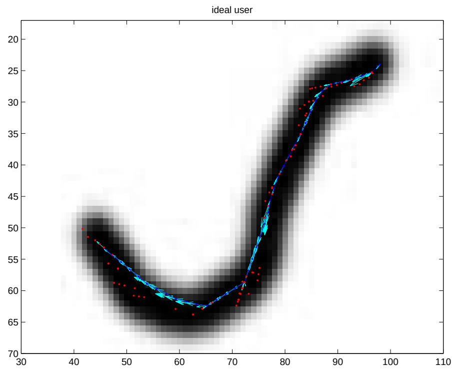
4.9.2 Model Testing

Figures 4.30(a) to 4.30(c) allow us to gain an insight as to what we might expect if one part of the interface functionality was penalised. Figure 4.30(a) shows the behaviour of an 'ideal' user who displays both high scanning and tilting behaviour and traverses the trajectory smoothly, only scanning where it was required. Figure 4.30(b) shows what we might expect then if the amount of scanning was penalised. We see from this figure that, given good initial conditions, the model tends to move in a straight line until it approaches the edge of the trajectory and is forced to change direction. Figure 4.30(c) shows what we might expect if instead the scanning was active and the tilting look-ahead was penalised. Although the model stays on the trajectory, we see that the model displays more bouncing behaviour since where the combination of a lack of ability to see ahead and a more unbounded scanning behaviour causes the model to perform less smoothly on the curve.

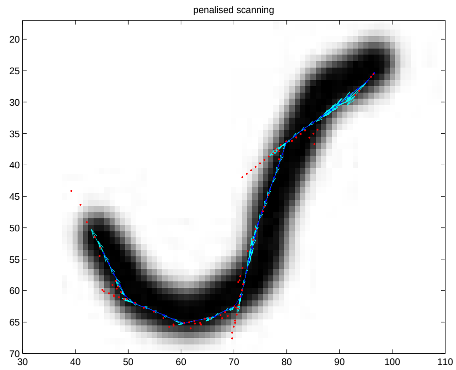
Recreating Behaviour

We may also show that it is possible to mimic, to a certain extent, some of the behaviour we observed in our trajectory following field trials using this simple model and gain an insight as to why this behaviour may have been observed.

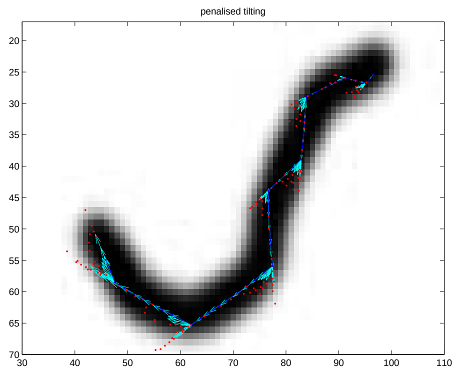
Figures 4.31(a) to 4.31(d) show modelled behaviours for the three different kinds of behaviour described above, i.e. *cautious*, *bouncing* and *clinging*. Although the model we have constructed is too simple to recreate fully the behaviour observed in the real trials, they can still provide us with an in-



(a) Modelled ideal user who exhibits high scanning and tilting behaviour.

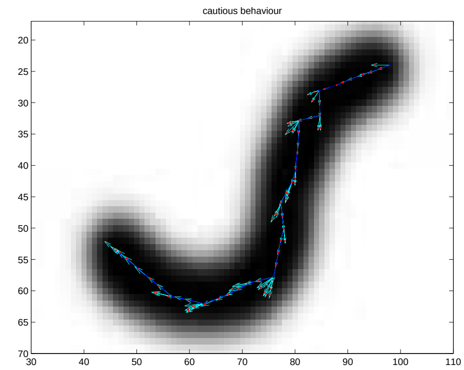


(b) A modelled user who displays active tilting behaviour but little scanning behaviour.

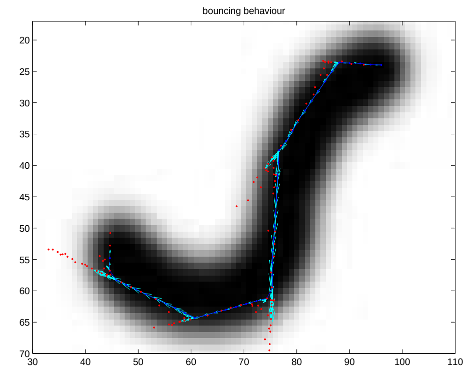


(c) A modelled user who displays active scanning behaviour but little tilting behaviour.

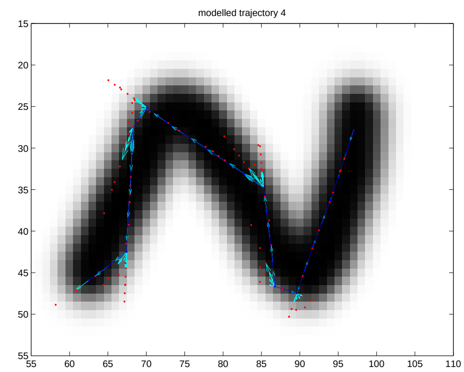
Figure 4.30: The effects of penalising the different parts of our cost function.



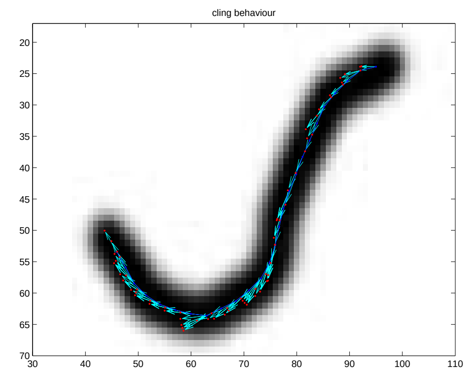
(a) A modeled cautious behaviour.



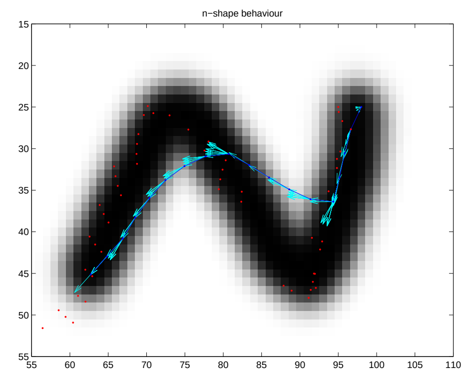
(b) A modeled bouncing behaviour.



(c) A modeled bouncing behaviour for the n-shaped trajectory number 4.



(d) A modeled clinging behaviour.



(e) A modeled n-shape behaviour.

Figure 4.31: Recreation of the main behaviours observed in the trajectory following field trial.

sight as to why this behaviour was observed. For example, the bouncing behaviour in figure 4.31(b) was recreated by penalising use of the scanning functionality whilst allowing the look-ahead behaviour to be more free, similarly for figure 4.31(c). This is intuitive, since we would expect a user, if they were not utilising their scanning functionality, would tend to move along in straight lines until it was impossible to go any further as they hit the edge of the trajectory and were forced to readjust. The clinging behaviour in figure 4.31(d) was created by penalising the scanning component and to a lesser extent the look-ahead functionality. For this behaviour we also increased the minimum look-ahead, indicating that this user was not utilising their look-ahead functionality as effectively as they could have been. The behaviour observed in figure 4.31(e), where the user strays out of the trajectory for a short period, was created by first penalising the scanning behaviour more than the look-ahead behaviour but more significantly the look-ahead was constrained to a high value (10-15 pixels ahead). This meant that the model was looking too far ahead and was actually receiving feedback from another part of the trajectory, meaning that it became possible for the model to stray into the white area of the density, exactly as is observed in figure 4.25 for a real participant, indicating to us then that this user was constantly looking too far ahead and needed to learn how to use the functionality more effectively.

4.9.3 Improvements

Our model is not perfect. It is a very simple representation of a human operator. Some improvements, which could be made include the incorporation of real dynamics. It is likely that if we included a dynamic representation of the user's motion, instead of the model utilised in this instance with an assumed constant velocity, we would see a lot more realistic motion at the edges of the trajectory, for example, where a high velocity may lead to the overshooting of the trajectory. This becomes much more important in the situation where the user may achieve higher velocities, if they were riding a bike, for example. Another factor that should be considered is learning effects. Most users at the beginning of the trajectory display slightly different behaviour than at the end, since by the end of the trajectory they have learned exactly what is happening and have adjusted their behaviour appropriately. It is also important that we consider such factors as the computational delay on the display of feedback to the user, as this is also likely to affect the user's behaviour somewhat.

4.10 Discussion and Conclusions

In this chapter we have demonstrated that probabilistic, multimodal, handheld interaction techniques can be applied effectively to allow users to explore density functions in space, with an example of pedestrian GPS navigation. The Monte Carlo sampling method provides an effective way of integrating probabilistic models into practical interfaces, and displaying the results in a multimodal fashion.

Results from our initial outdoor field trial and our more controlled indoor field trial support the hypothesis that the uncertain displays require less effort and result in more stable behaviour. The trajectory following field trials have shown that it is possible to guide users to a desired location over a set trajectory or path and a number of interesting behaviours have been observed. Interactive sonification of the exploration process produced a navigation system which may be used eyes-free, where the user brings their sensorimotor systems into the interaction with an augmented environment. It is clear from this initial data that it is possible for users to navigate to the final location in a featureless environment like a playing field, using audio and vibrotactile feedback alone. Their performance and confidence improves significantly when the audio and vibrotactile constraints from the system are coupled with the natural constraints of the environment, suggesting that the system is promising for a range of use cases.

We have shown the potential of using a simple model of human behaviour to recreate the kind of behaviour observed in our field trials. A more fully developed model has the potential to provide an insight as to how a user may perform in an experiment, prior to the experiment and has the potential to explain, in a quantitative manner, some of the behaviour observed in a field trial where the user was required to control some system.

The system presented here though has applications well beyond simple trajectory following. This system offers a new general mechanism for providing highly interactive context-aware applications. The densities here could represent $P(C_i|x)$ - the probability of context state C_i given the current state vector x. By treating our system as a separate density layer in any application it is possible to provide different functionalities. For example, densities could be used to represent differing contexts such as local socioeconomic levels or crime rates, areas of interest to tourists or various Geographic Information Systems data. In the following chapter we will introduce and demonstrate the use of one other application using this system and discuss the potential further applications.

Chapter 5

Messages in the Air

5.1 Summary

This chapter introduces an application, which combines the work conducted in the previous two chapters to provide a mechanism for producing highly interactive context aware applications based on the probing of a local density. We present an example application, *airMessages*, which enables the locating and viewing of messages left in the virtual environment. We demonstrate the utility of a system such as this with a small field study. Finally, we describe a number of potential applications for this system.

5.2 Introduction

The work developed and lessons learned in the previous two chapters open up a wealth of opportunities for the creation of highly interactive location-aware mobile computing environments. We have thus far developed an egocentric location-aware interface for interaction around the body in chapter 3 and exocentric interface for interaction in the real-world in chapter 4, so it is a natural extension now to combine these two forms of interaction to produce a real world application enabling an embodied, active and gestural interaction with the 'real' world. The work we describe here is an attempt to bring augmented reality and virtual environments to everyday handheld devices without the use of explicit visual information. What we have achieved with this work is to build an audio based eyes-free augmented reality style system fully contained in a hand-held device, enabling the construction of augmented reality style applications.

The application developed here differs from previous augmented reality systems in that we do not require the use of external markers or sensors. This kind of system can be considered desirable because it opens the door for the active augmentation of our real-life surroundings as well as our own bodies. For example, it is still possible in this situation to carry our most used tools, most listened to music or most important documents in egocentric 'virtual pockets' around our body, but also it is possible using the exocentric interface introduced in chapter 4 to leave these objects in certain locations in the real world to be picked up at a later date. This embodied interaction with the virtual environment opens up the opportunity for the development of techniques enabling audio shape perception or moulding of virtual objects. It becomes possible for the user to shape and augment this environment over time in his own personal way, creating a highly a personalised virtual skin.

5.3 Augmented Reality and Virtual Environments

By definition, a virtual environment involves the replacing of the real world with a virtual world. In augmented reality a virtual world augments or *supplements* the real world with additional information. Previous work in this area has focussed principally on the use of visual augmentation and addressed a wide range of application areas including aircraft cockpit control (Furness 1986), the aiding of manufacturing processes (Caudell and Mizell 1992), assistance in medical applications (Lorensen *et al.* 1993) or personal tour guides (Feiner *et al.* 1997).

The first augmented reality system was developed by Sutherland (1968) who constructed an elaborate system designed to present a user with a perspective, wire-frame image, which changed as the user moved. One of the more important applications of augmented reality is to the medical field and one example of its use is for Ultrasound imaging. Using an optical see-through display, an ultrasound technician can view a rendered image of a fetus overlaid on the abdomen of a pregnant woman. Another example includes that of Lorensen *et al.* (1993) who describe a procedure for surgical planning and surgical support that combines live video of a patient with a computer-generated 3D anatomy of the patient. This permits surgeons to plan access to the pathology that exists within the patient and provides them with a live view of the patients internal anatomy during the operation. Another significant application of augmented reality is to manufacturing or maintenance processes. It is easy to imagine a machinery technician, instead of flicking through his repair manual or searching through an online guide, simply taking his Head Mounted Display (HMD) and visualising any problems the machinery or computer equipment may possess. Feiner *et al.* (1993) describe a system for printer maintenance, *KARMA*, which explains simple end-user laser printer maintenance tasks using a head mounted display, overlaying images indicating the location of the ink cartridge or paper tray, for example.

Early augmented reality applications were confined principally to indoor settings. One of the first outdoor systems to be implemented was the Touring Machine (Feiner *et al.* 1997). This self-contained backpack-based system includes magnetometers and accelerometers for head orientation tracking and a differential GPS for location information. This system also contains a mobile computer with a 3D graphics board and a see-through HMD. The system presents the user with information about their urban environment, in this case the campus at Columbia. Although these "backpack" systems have been successful proof-of-concept prototypes, they lack the convenience of a fully hand-held system.

The development of smaller and more powerful devices in recent times has led to the development of an increasing number of applications on handheld devices for truly mobile augmented reality. Some completely handheld AR applications make use of the screen and cameras available on these devices. Wagner and Schmalstieg (2005) describe a system designed for use as a museum guide. Using external markers, which are recognised by the device, the system may overlay extra information or animations on the museum exhibits. There exists a number of hand held AR systems which make use of these external markers (Wagner and Schmalstieg 2005, Henrysson *et al.* 2005, Mohring *et al.* 2004, Mitchell 2006). Baillie *et al.* (2005) describe a fully contained handheld system which combines GPS and attitude information to visualise a virtual image of a building in the present or past on screen by simply pointing their device at that building.

Due to the limited screen space and resolutions on these mobile devices it is beneficial to concentrate more on the audio and haptic senses and less on the visual sense. There are augmented reality systems which focus completely on the audio sense, leaving a user's visual attention free, which is important, especially when a user is mobile. Bederson (1995) describes a prototype automated tour guide which superimposes audio on the world based on a user's location. Users in a museum may hear information about exhibits in their local vicinity using a hand held device and sensors located in the ceiling of the museum. Lyons *et al.* (2000) describe another audio augmented reality system that uses a wearable computer and an RF based location system to play sounds corresponding to the user's location and current state. They describe a fantasy style game implemented with this system. Audio Aura (Mynatt *et al.* 1997) is a system which augments the physical world with auditory cues allowing passive interaction by the user. By combining active badges (Want *et al.* 1992) and wireless headphones, the movements of users through their workplace can trigger the transmission of auditory cues and convey information to the user such as the current activity of their colleagues or the arrival of new emails.

5.4 Contextual Interaction

As discussed, the reliable detection of user intention is one important area of research for future mobile applications. Context detection is another. But the two are not mutually exclusive since the correct classification of a user's current context may be extremely important when attempting to infer a user's intention.

Context-aware computing is defined as *"an application's ability to detect and react to environment variables"* (Barkhuus and Dey 2003). The most common use of context in human-computer interaction is to tailor the behaviour of a system to patterns of use. Brummit *et al.* (2000) describe a system, *Easy Living*, which enables the dynamic aggregation of a number of I/O devices into a single coherent user experience in an intelligent environment, in this case a living room. Cheverst *et al.* (2000) describe an intelligent electronic tourist guide, *GUIDE*, which was built to overcome many of the limitations of the traditional information and navigation tools available to city visitors. The system combines mobile computing technology with a locationing system to present city visitors with information tailored to both their personal and environmental contexts.

Since the notion of context-aware computing was introduced by Schilit *et al.* (1994) there have been a number of different definitions, often related to the level of interactivity. Chen and Kotz (2000) define the notions of *active* and *passive* context awareness. They define active context as that which influences the behavior of an application and passive context as that which is relevant but not critical to an application. As an example we may think of the clock update on a mobile phone as being active if it updates automatically and passive if it prompts the user first before updating. Cheverst *et al.* (2001) introduce the notion of information push versus information pull. A 'pull' system is one in which the emphasis is on the user to decide when context-aware information is presented and they may pull this information to themselves and a 'push' system is based on information being presented automatically to the user, which is triggered by contextual events. The system we describe is a 'pull' system and is also 'active' but in a slightly different sense. Our system is highly interactive, meaning that users can probe a density in their immediate environment and actively check for information in their surroundings rather than relying on the system to make decisions for them.

5.5 Density Exploration

The application developed here acts as a general mechanism for providing highly interactive context aware applications. By representing 'context' in this situation as a density overlaid on the real world we can search this density and probe for information stored there using the functionality of our interface, described in chapter 4.

This density or context can take many forms. The density may contain messages, which can be left for our friends, the density may contain information about a particular location, information about when your next train leaves, information about the local socioeconomic levels in an area or information about what kind of offers your favourite record shop has at the moment, all of which can be accessed by simply interacting with the information placed in that density. Other possibilities for this kind of system include the cooperation between two or more systems so that you may keep track of the location of your friends or loved ones, akin to social networking but in the 'real world'. It is easy to imagine having the location of all of the people in your friends network displayed on the screen and leaving messages, videos, pictures or games in specific locations for people in your network to pick up. Essentially, it becomes possible with this kind of system to overlay a 'virtual skin' on the real environment, which we can alter depending on what we are interested in at that point in time. Some of the possibilities for this system are described later in the chapter.

5.5.1 airMessages

AirMessages is an example application of our density exploration mechanism, which combines the functionality of the applications described in the previous two chapters. Combining again the use of a global positioning system, a model of the user's local environment and Monte Carlo propagation, users are able to 'drop' and retrieve messages in their own personal virtual environment. Users can leave messages, represented as local densities, anywhere in the environment, which is *overlaid* on the real world.

Espinoza *et al.* (2001) describe a similar system, *GeoNotes*, arguing that location-based systems must allow users to participate as content providers in order to achieve a social and dynamic information space. Their system attempts to blur the boundary between physical and digital space while at the same time striving to socially enhance digital space by letting users leave virtual messages, which are linked to specific geographical positions. Jung *et al.* (2005) present the design of an enhanced mobile phone messaging system, *DeDe*, that allows a user to define in what context a message will be delivered. 'Context' in this particular situation is either the time of day, the location, whether a specific bluetooth device is currently in range or whether a certain number is calling. Similarly Ludford *et al.* (2006) develop a location-based reminder system, *PlaceMail*, and demonstrate with a field study that their system supports useful location-based reminders and functional place-based lists. E-graffiti (Burrell and Gay 2002) is a context-aware application, which senses a user's location and displays notes dependent on that location and allows users to create notes that they can associate with a specific location. They conduct a field study with 57 participants, finding the idea of location-specific notes was something that appealed to users.

Our system differs from those described above in that a user may actively probe his environment in an embodied manner using the tilt of the device to control the variable Monte Carlo propagation time-horizon and effectively 'look-ahead' or 'project' themselves into the distance (higher tilt gives a further look ahead). The user can sense if they are interacting with an object anywhere in the local area by hearing, via audio and feeling, via vibrotactile feedback 'impacts' with the message, represented by the Monte Carlo predictions interacting or impacting with this overlaid density as illustrated in figure 5.3. If the user senses that there may be something in a specific part of the local area they may then move towards that area to examine what has been left there with the message being displayed visually when they are in close enough proximity.

The mechanism for dropping messages is gestural and uses the same approach as that described in chapter 3, eliminating the need to use any buttons at all with this application. To drop a message the user simply gestures to their hip, as illustrated in figure 5.1, where they are notified via vibrotactile feedback that a message has been successfully left in that particular location. Future extensions of this functionality might include more creative ways to drop a message. For example, a more realistic dropping gesture could be used where the message is flicked from the end of the device to the ground or a ball, representing the message is rolled from the device into a virtual pocket.

Figure 5.1: A gesture to the hip, used as the mechanism for dropping a message into the virtual environment.

Location Projection

The functionality of our interface combined with the kind of activity performed in this application allows us to describe a new way of thinking about our 'look ahead' functionality. In this situation we may think of a user 'projecting' their location as they interact with the environment and the local density. The user in figure 5.2 is passively walking through the local density, sensing anything which they happen to pass through. But in figure 5.3 the user is probing the local environment whilst projecting their current position into the distance, the user is effectively saying "*what* *would I feel if I was over there...*". This way of thinking allows users to

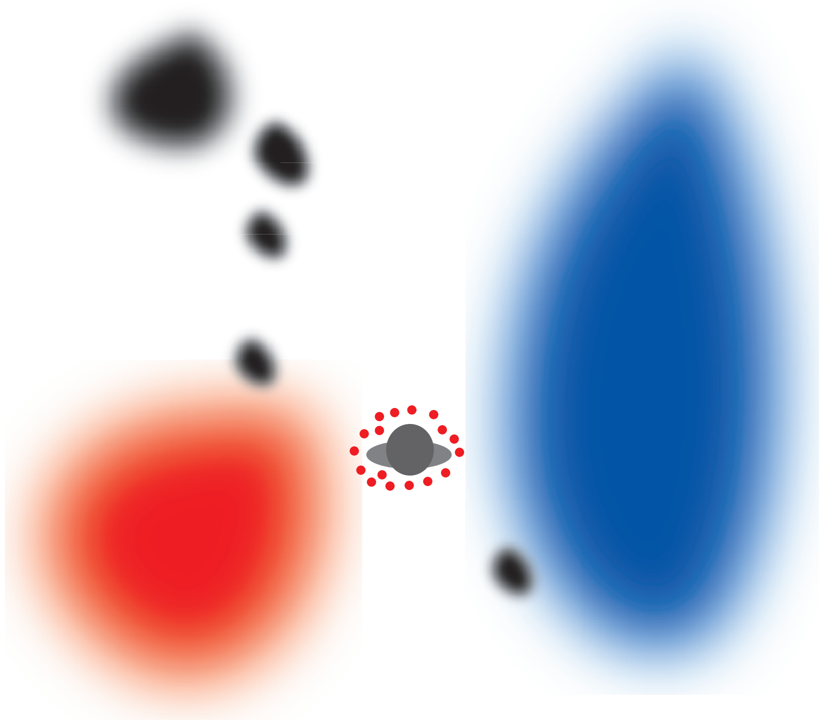


Figure 5.2: The user takes a passive approach to interacting with the local density (coloured and black areas) with particles spread around the body, which can potentially interact with any part of this density causing some audio and vibrotactile feedback to be displayed to the user. It is possible to alter the spread of the particles by varying a simple parameter in the Monte Carlo simulation.

take a much more active and embodied approach to retrieving information from their current context, listening and feeling for objects located in their virtual environment. They can scan the horizon and project themselves forwards in time to build a mental model of their virtual environment and any objects which it may contain. This kind of interactive system promotes again the concept of spatial organisation aiding a user's memory, as a user may wish to leave specific things in specific locations in their exocentric real world interface, actively 'grabbing' this information as they go.

5.5.2 System Testing

A field study was conducted to test if this system could be used by a number of different users and to examine how they interacted with this system.

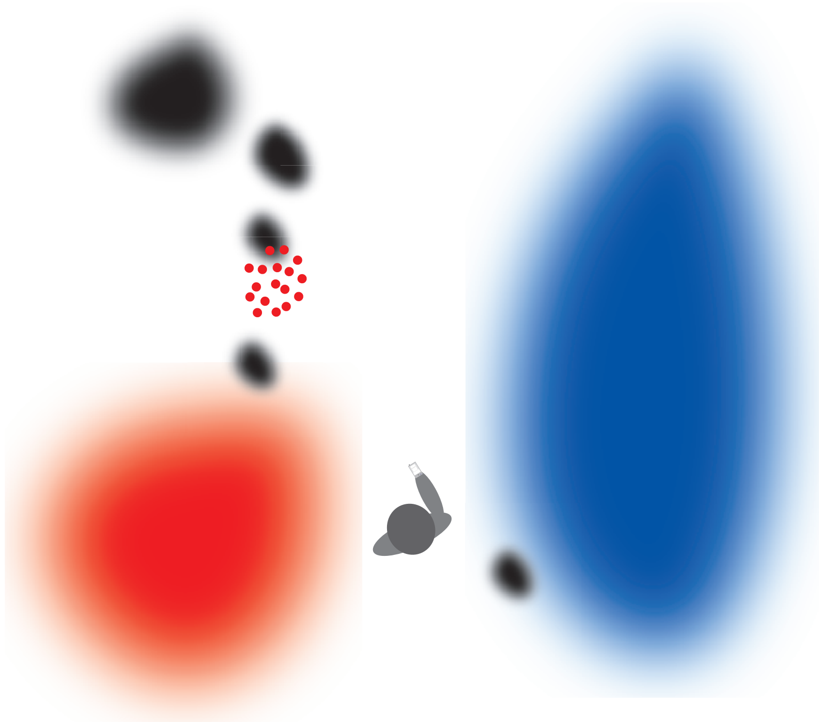


Figure 5.3: The user takes an active approach to interacting with the density. They consciously probe the locality to locate object situated there in the density without the need to physically walk to those areas.

Method

In total 6 users completed the trials, which involved following a set scenario around a small area of the university campus. They were first given an introduction to the system and a brief explanation of the whole concept. They were also instructed about how to use the gestural interface and allowed to practice before beginning the walk through. Participants all started in the same position (indicated in figure 5.4) and were asked to manually locate the first message, which was placed close by (location 1 in figure 5.4), using the functionality of the interface. When they had located this message they then headed towards that location with the message they had found being displayed as soon as they were in close enough proximity. This message read:

"go to the Hume building door and drop a message"

The participant then heads to the correct location (location 2 in figure 5.4) and gestures to their hip in order to drop a message. This dropped message then appeared on the screen and read:

Figure 5.4: The are used for the trial. Participants start at target 4 and try to locate target 1. They then move to target one, then to 2, then to 3, back to 4 and back to 2 again where the trial is complete.

> *"you dropped this message...go and pick up a message in the car park"*

The participant then heads to the car park (location 3 in figure 5.4) where they are aware a message exists but do not know exactly where in the car park. As they get closer to the car park they again begin to probe the local area using the panning and 'look-ahead' functionality of the interface, attempting to locate this message with the aid of audio and vibrotactile feedback. When located, this message read:

> *"drop a message outside the Hamilton Institute"*

The user then headed back to the Hamilton Institute (location 4 in figure 5.4) and dropped a message there which read:

"go back to the message you dropped at the Hume building"

When the user returns to the message they originally dropped at the Hume building (location 2 in figure 5.4) it now read:

"you're finished!"

and the trial is completed.

What this simple scenario enables us to do is examine how users interact with the system generally and allows us to observe any interesting behaviour or any problems that people may encounter. It also allows us to understand how easy people find the whole concept to grasp. For example, are they confident that these virtual messages will be where they are told? Especially when they are told to return to a message that they dropped in the first place? From a more technical point of view it allows us to examine how users interact with the functionality of the system and how the use of a Monte Carlo simulation and a model of the local environment really aids the user. All data from each of our sensors was logged for each user.

5.5.3 Results and Observations

All users successfully completed the tasks required of them. The locating of the first target provided users with the most problems as they attempted to gain a feel for the system. All participants display a very active behaviour when trying to locate their targets, indicated by the increased arrow length at the beginning. This is particularly prominent in figure 5.6(a) for the acquisition of target 1 by participant 1 and for the acquisition of target 3 also by participant 1 in figure 5.7. This quiver plot shows the current position of the user (blue dots) as measured by the GPS, the direction they are looking in at that point (direction of the arrow), the level of look-ahead at that point (length of the arrow) and the Monte Carlo predictions that provide the feedback (red dots).

Some users (figures 5.6(a) and 5.8(a)) seemed to be aware that the target they were trying to locate was somewhere in the distance and were using the look-ahead function effectively, to draw themselves towards the target and it is observed that with participant 1 in figure 5.6(a) and participant

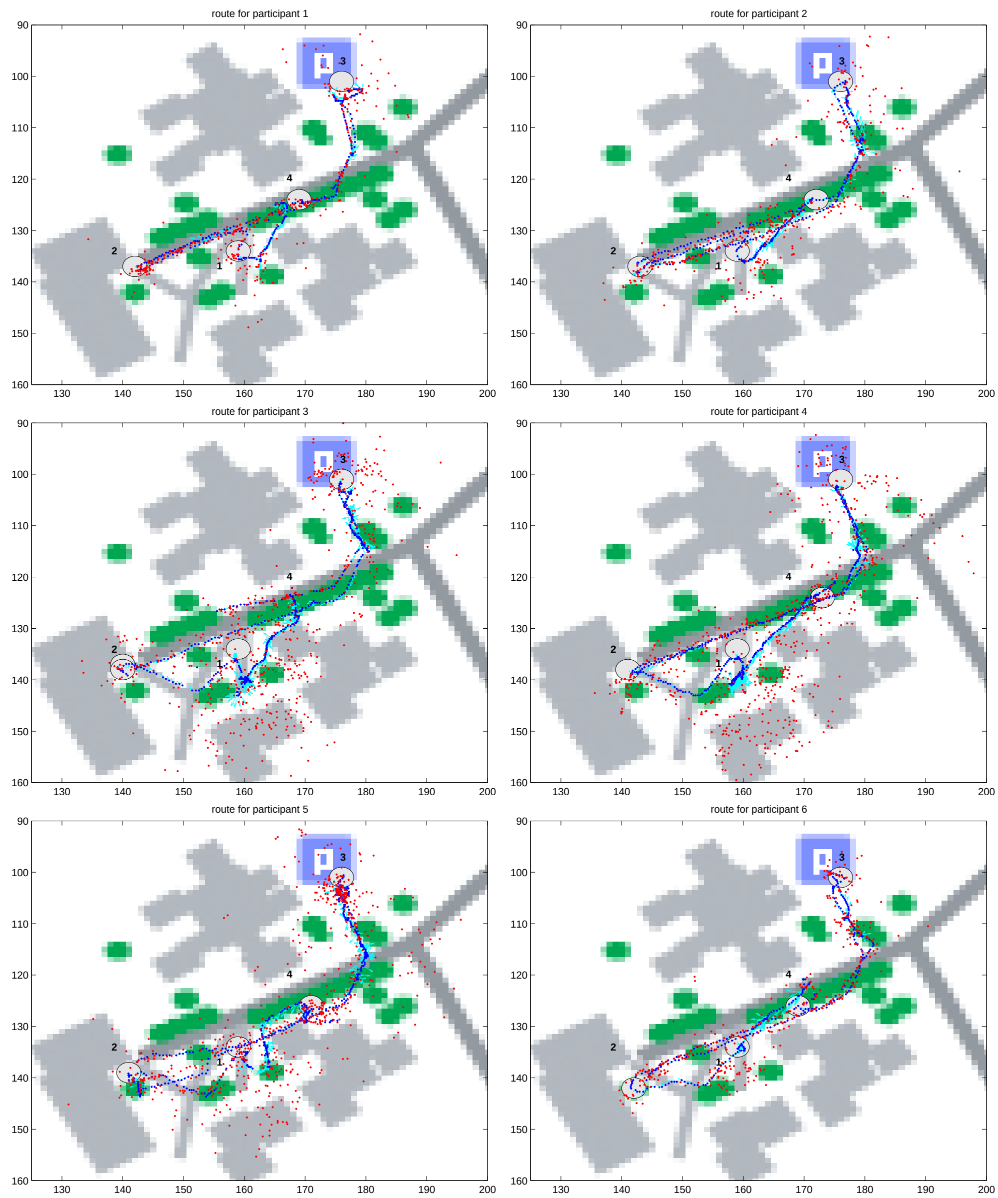
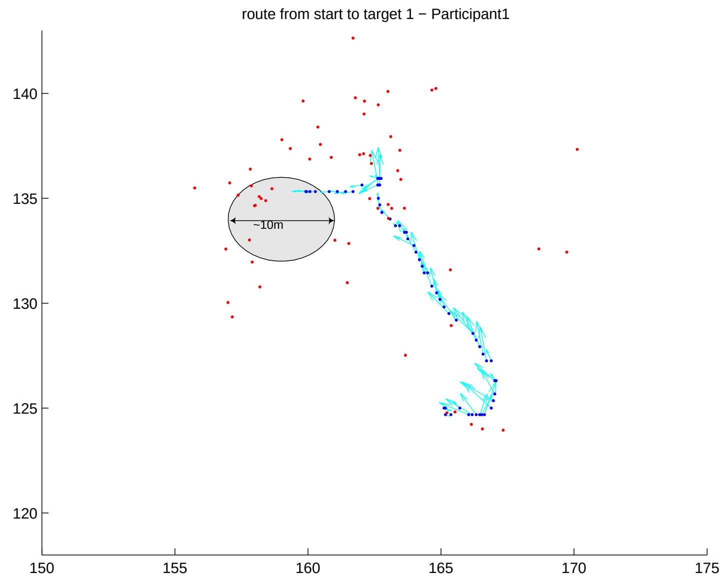


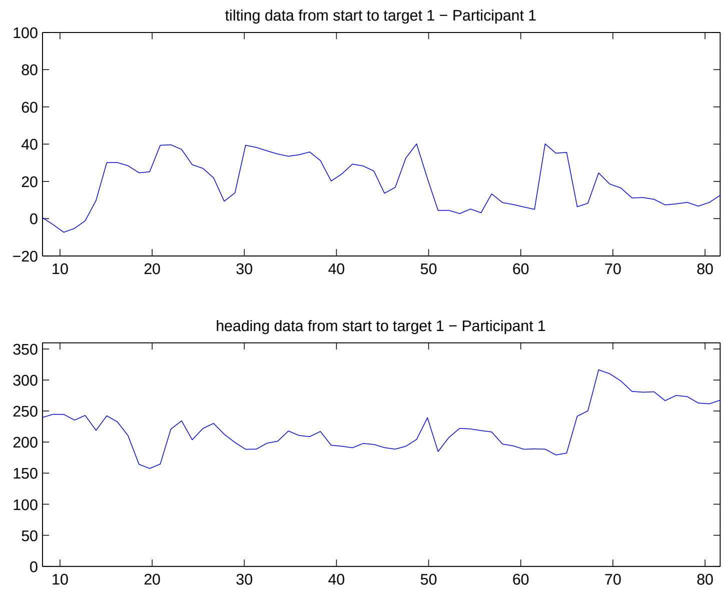
Figure 5.5: The route traversed by all 6 participants.

2 in figure 5.8(a) the look-ahead arrows decrease in size as the user moves to the target, indicating that they have some kind of 'fix' on the target as they are moving towards it. This is confirmed if we examine the device tilt data in figures 5.6(b) and 5.8(b), which shows a gradual decrease as the participant moves towards the target, acquiring it at the end.

A more common strategy was to move 'straight ahead' without utilising the interface much at all after a roughly correct direction had been determined using the full functionality of the interface. This lack of use of the interface is confirmed if we look at the tilting and scanning energies in figures 5.10 and 5.11, which show that participants 3 and 4 used considerably



(a) Participant 1 locating target 1. This participant displays a very active look-ahead at the beginning of the route in order to locate the correct direction of the target and then draws themselves toward the target as the look-ahead decreases. There is a slight overshoot just as the target is acquired.



(b) Tilt and heading data for the acquisition of target 1 by participant 1. We observe a gradual decrease in the tilt of the device, indicating a decrease in the level of look-ahead, as the participant moves towards the target.

Figure 5.6: Acquisition of target 1 by participant 1.

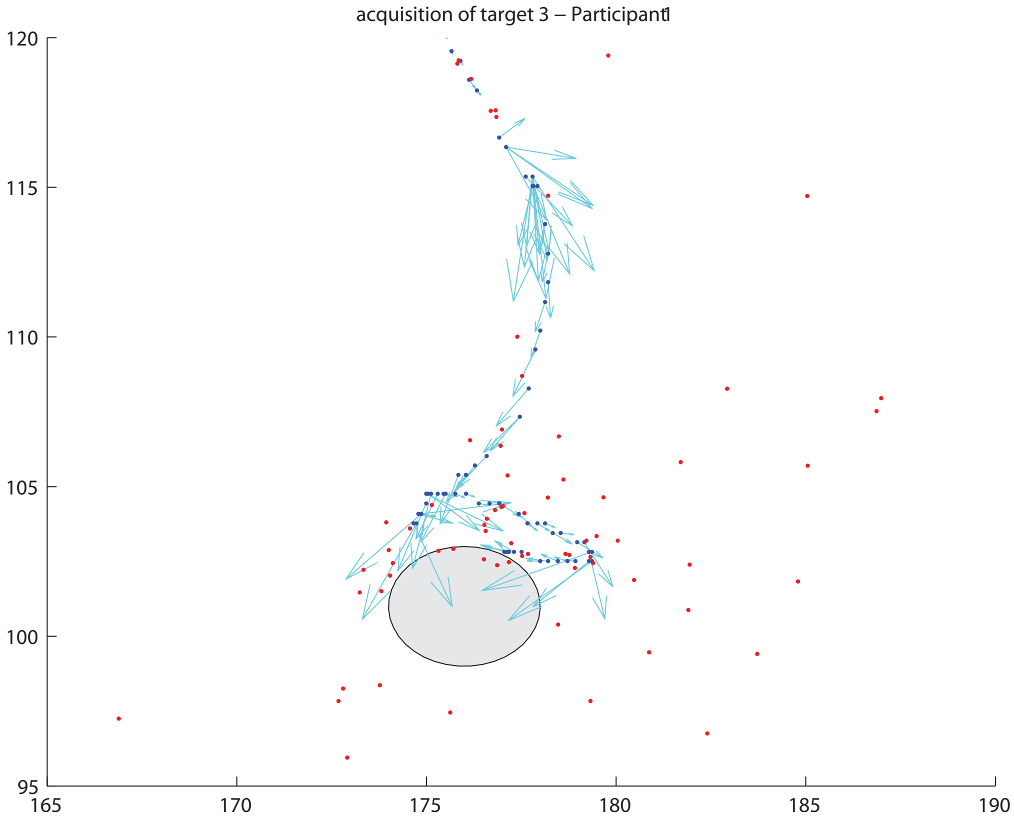


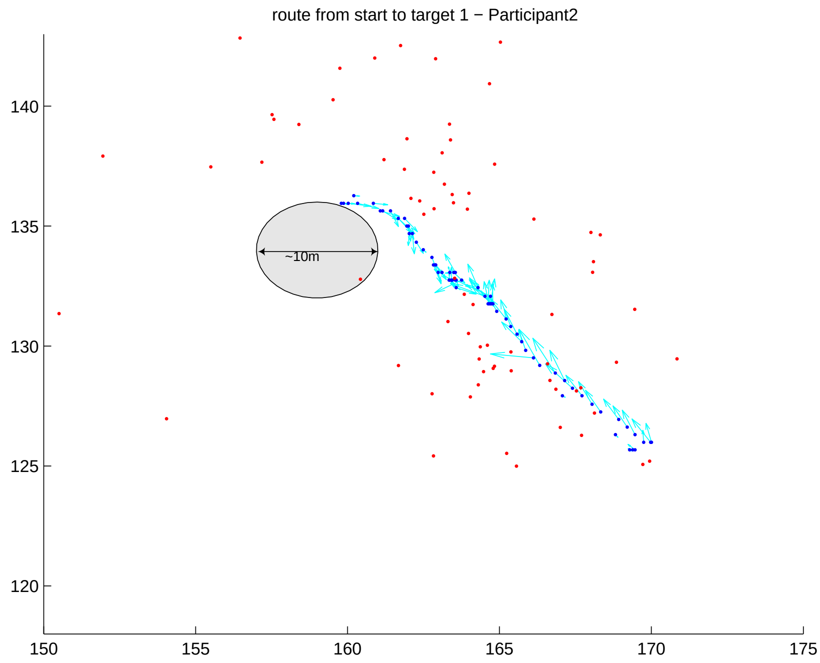
Figure 5.7: Participant 1 locating target 3.

less tilting and scanning for target 1 than the rest of the participants. This is also confirmed if we examine the tilt data from the device for these target acquisitions.

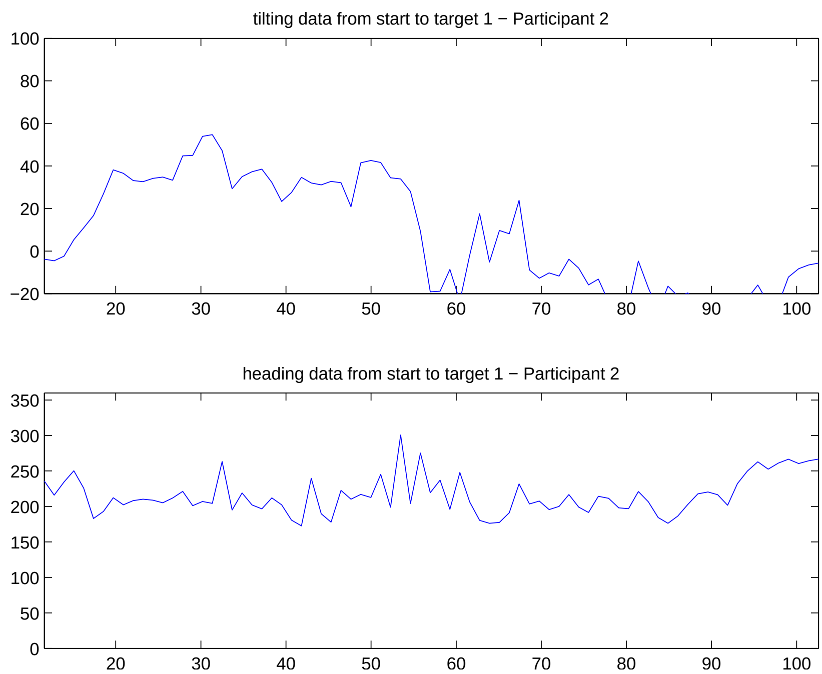
Figure 5.13(b) shows a long relatively inactive period as the user walks in the correct direction towards the target and figure 5.12(b) shows the same but to a lesser extent.

Also, from figure 5.9 we observe that participants 3 and 4 took considerably longer to locate this target than the other participants, indicating that the 'straight ahead' strategy was not a good one.

This strategy also caused an overshoot as observed figures 5.12(a) and 5.13(a) meaning that the participants became slightly confused and were forced to readjust their strategy and locate the target again. This was because participants 3 and 4 were not utilising their look ahead function sufficiently and the Monte Carlo predictions were always slightly ahead of the users at the horizon they chose to hold the predictions at. This meant



(a) Acquisition of target 1 by participant 2



(b) Tilt and heading data for the acquisition of target 1 by participant 2. There is a gradual decrease in the tilt of the device, indicating a decrease in the level of look-ahead, as the participant moves towards the target. This indicates that the participant has a 'fix' on the target as they move towards it.

Figure 5.8: Acquisition of target 1 by participant 2.

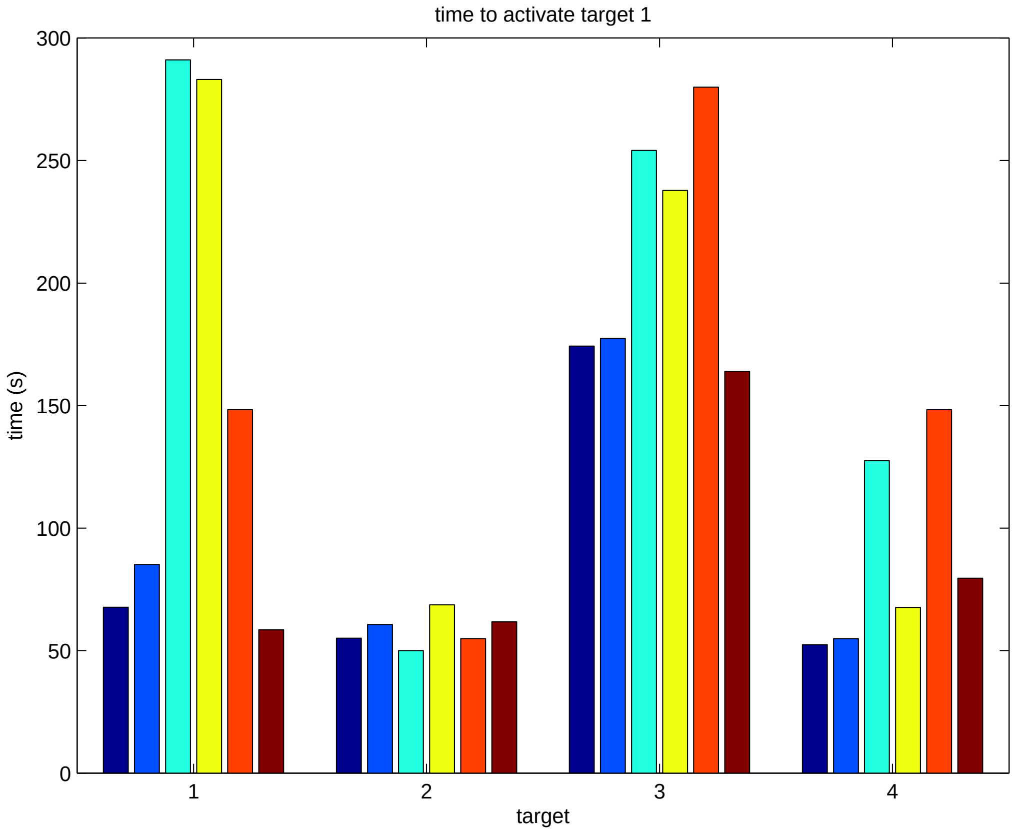


Figure 5.9: Time to acquire each message for each participant.

that when the participants received good feedback they were not seeing the message, indicating that they were in fact not in the correct area. They then move on again in that direction, missing the actual target area, and suddenly the feedback was lost.

One strategy used by the participant in figure 5.14 was successful and we observe from figure 5.9 that this participant took least time to reach target 1. This participant also displays high scanning and tilting energy for target 1 relative to the other participants meaning that they have employed a 'sweeping' strategy, covering a large area with Monte Carlo predictions, which although effective in this case, is a sign that the user did not fully grasp how to use the system. This participant employs the same sweeping strategy for the acquisition of target 3 in figure 5.15 but to a lesser extent.

Observing figures 5.10 and 5.11 allows us to observe how one participant compared with another for the acquisition of a particular target or message but it is unfortunately not valid to compare the two acquisitions at targets

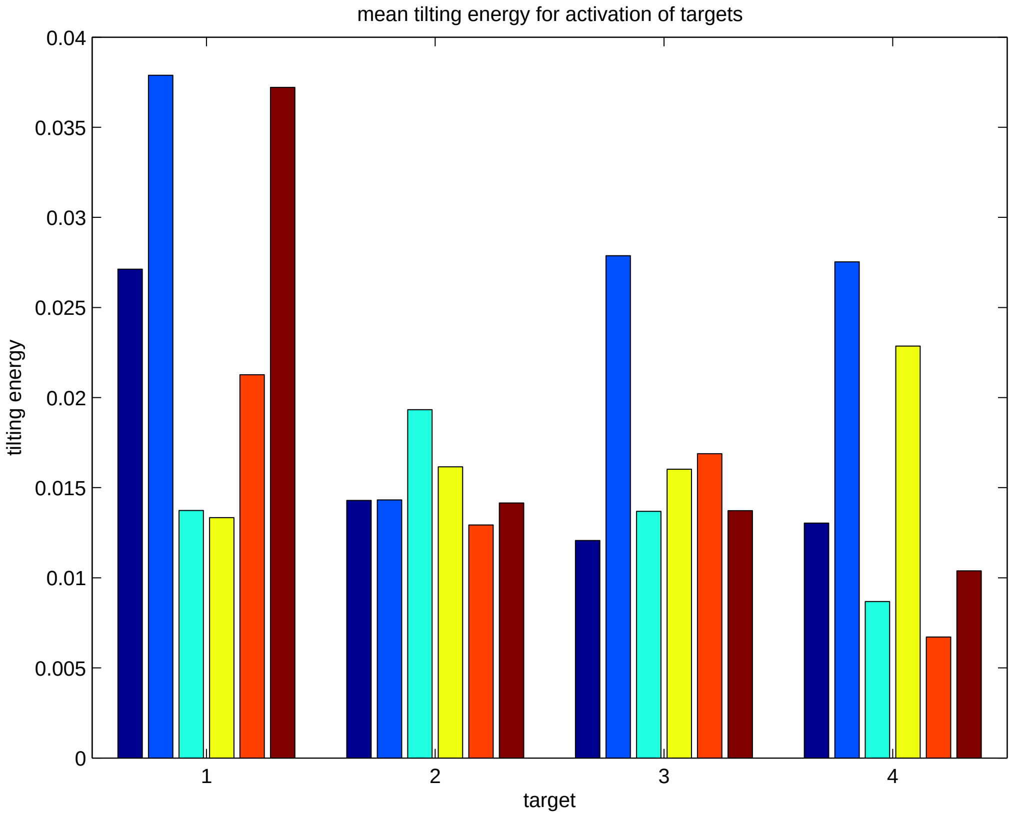


Figure 5.10: Mean tilting energy for each participant for the acquisition of each target.

1 and 3, where the actual functionality of the interface was being used, since there was a considerable walking distance between target 2 and target 3. We may though observe from these figures, the energy in the walks between two targets. From the point where target 1 is acquired to target 2 and from the point where target 3 is acquired to target 4 (targets 2 and 4 in figures 5.10 and 5.11), users are simply walking and not utilising the functionality of the interface at all.

It is also interesting to note that the participants who generally display the largest scanning and tilting energies relative to the other participants, i.e. participants 1 and 6, also show the lowest acquisition times relative to the other participants. This, although subjective at this point, is promising evidence that the functionality and interactivity provided by this interface does help the user in this kind of task.

After the users had reached target 1 they were requested to move to target 2 and drop a message there. We can see here the message drop in

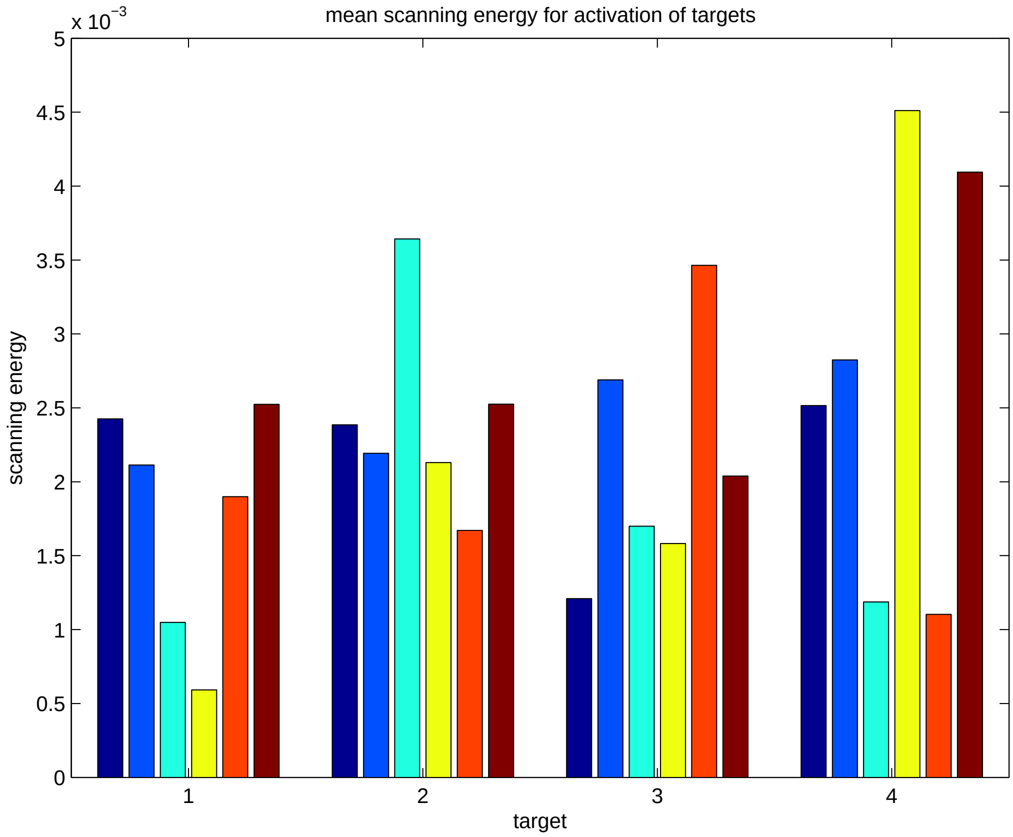
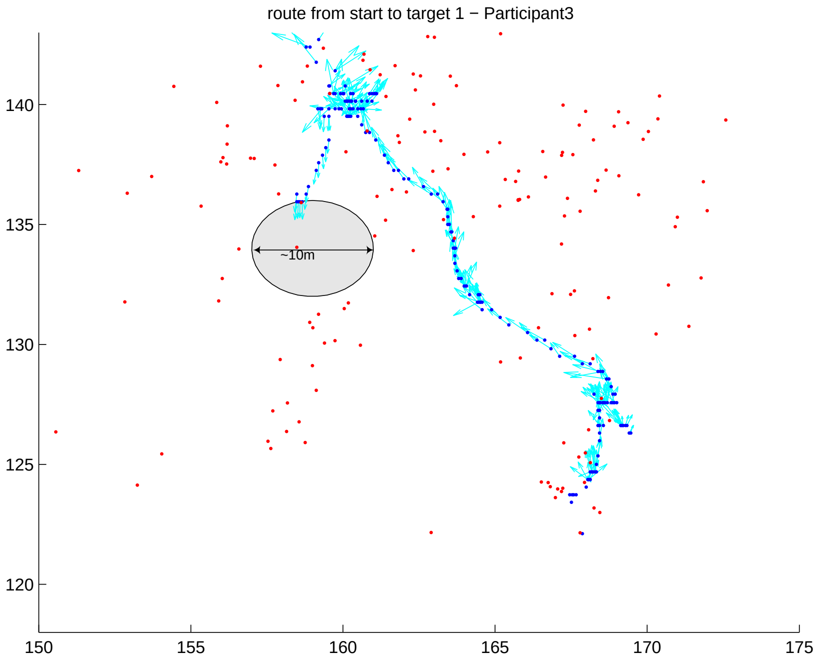


Figure 5.11: Mean scanning energy for each participant for the acquisition of each target.

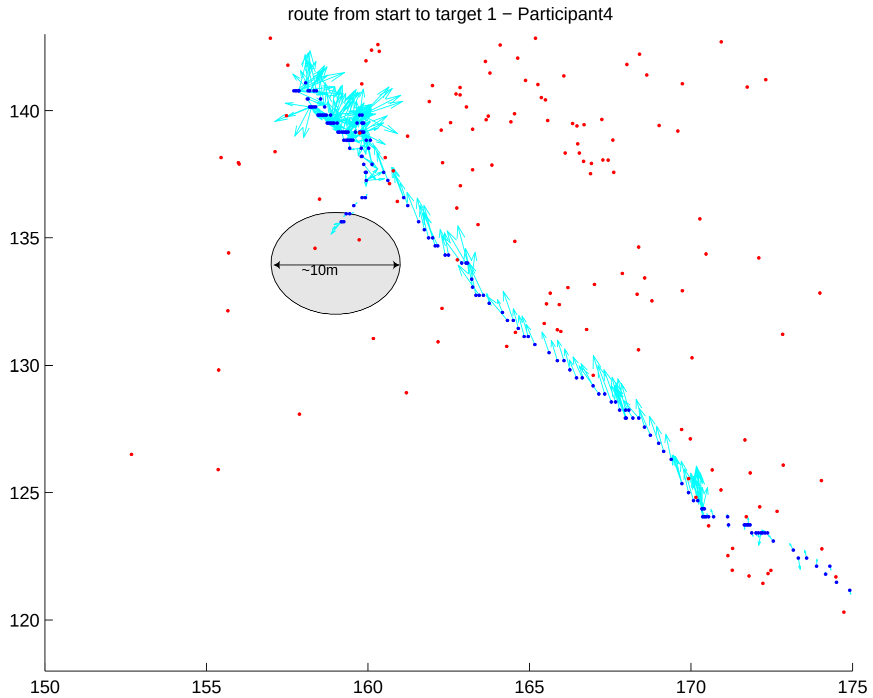
the accelerometer data in figures 5.16, 5.17 and 5.18. We see the data first as it is coming from a device held normally in front of the chest, then we observe the abrupt change as the device is rotated and moved to the hip, then we can see the steady relatively invariant data of a device held at the hip.



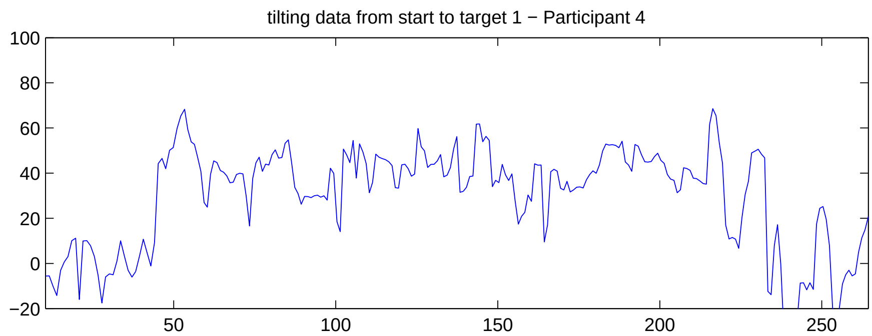
(a) Participant 3 locating the first target. This participant displays significant activity at the beginning to obtain the correct direction but then stopped using the full functionality, suffering a large overshoot. The participant was then forced to scan again to obtain the correct direction before acquiring the target.

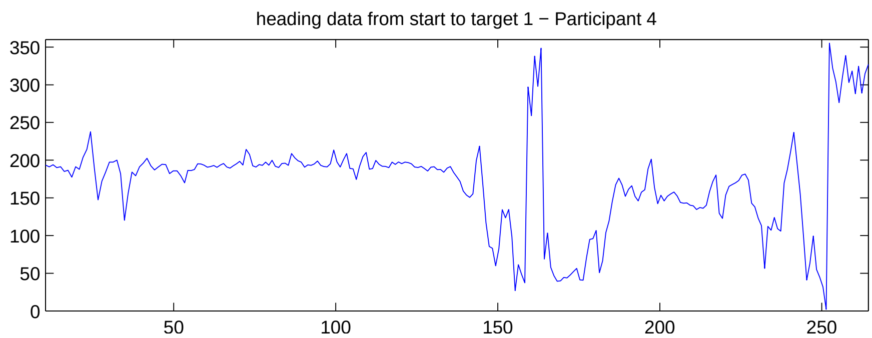
(b) Tilt and heading data for the acquisition of target 1 by participant 3.

Figure 5.12: Acquisition of target 1 by participant 3.



(a) Participant 4 locating the first target. This user obtained the correct direction but then utilised the interface little after, which led to an overshoot and a new search for the correct direction.





(b) Tilt and heading data for the acquisition of target 1 by participant 4.

Figure 5.13: Acquisition of target 1 by participant 4.

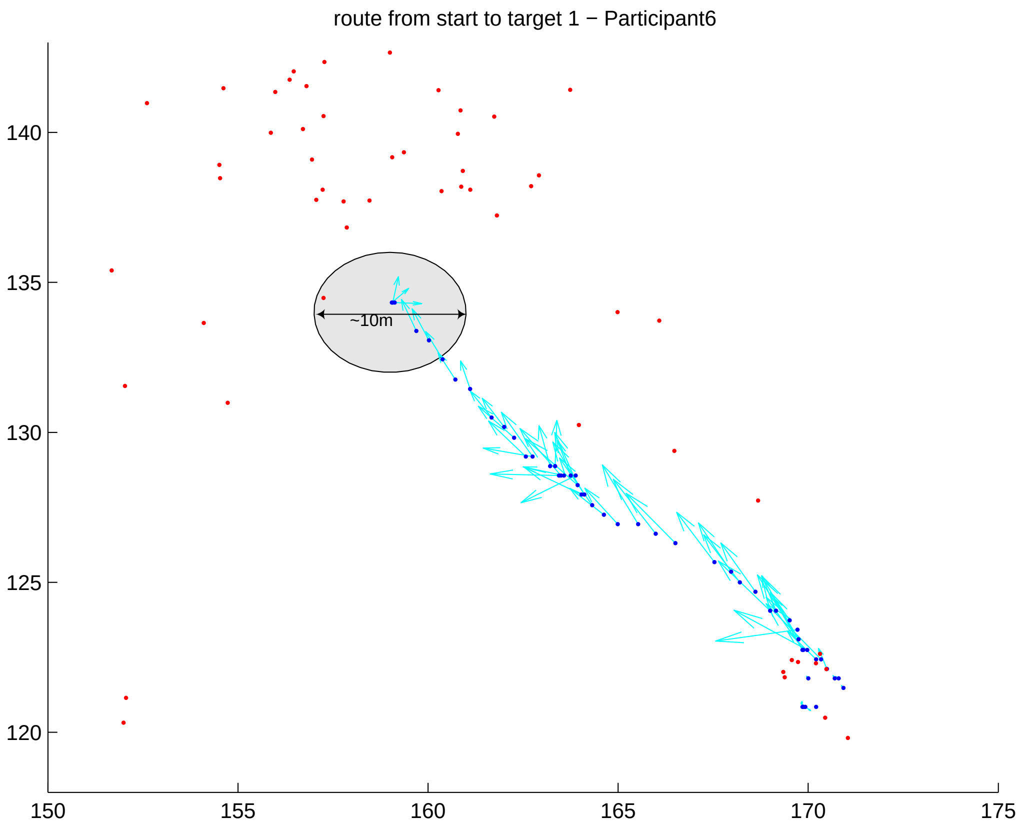


Figure 5.14: Participant 6 locating the first target. This participant obtains the correct direction first time and heads straight to the target.

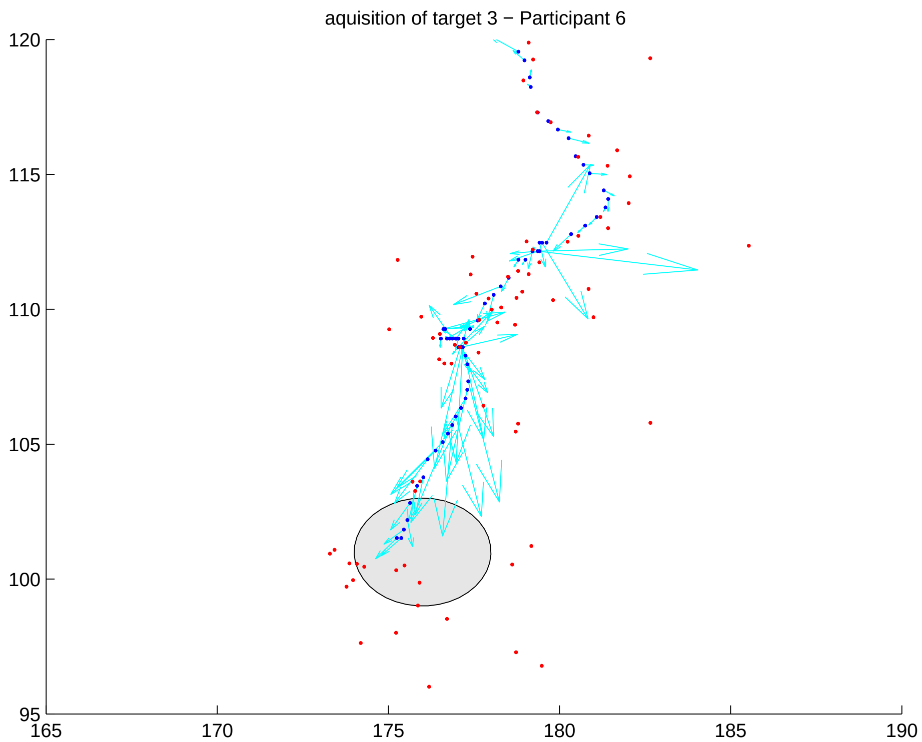


Figure 5.15: Participant 6 locating target 3.

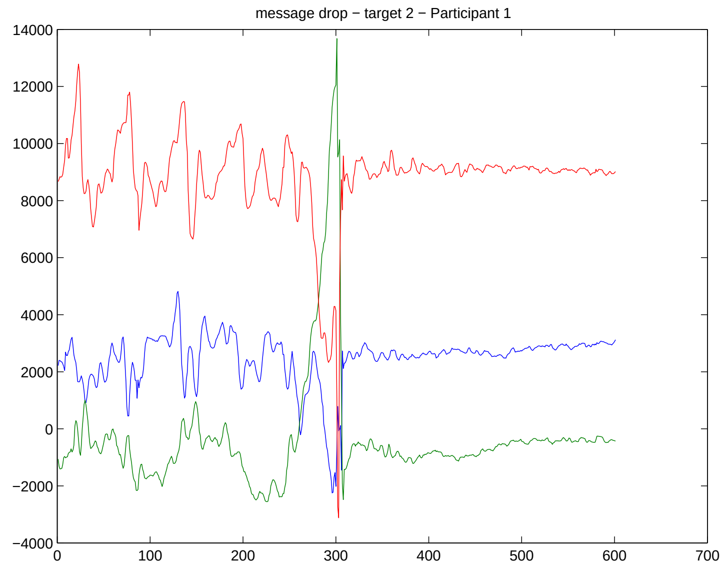


Figure 5.16: Acceleration output for participant 1 dropping a target at location 2.

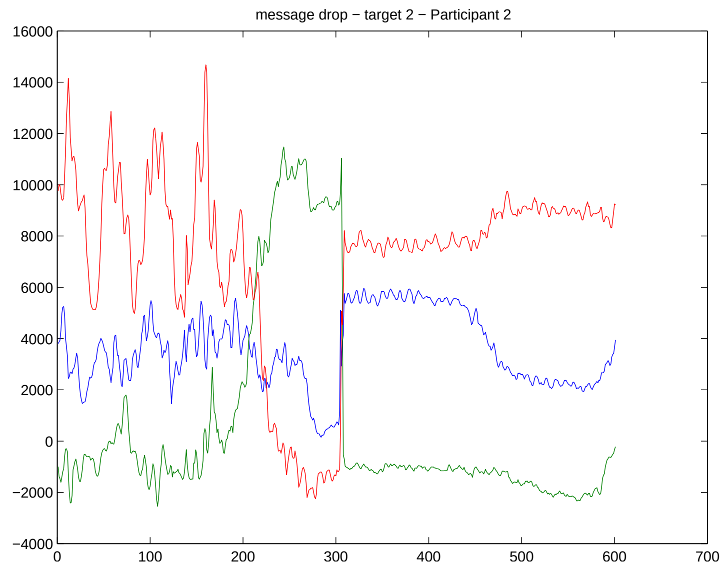


Figure 5.17: Acceleration output for participant 2 dropping a target at location 2.

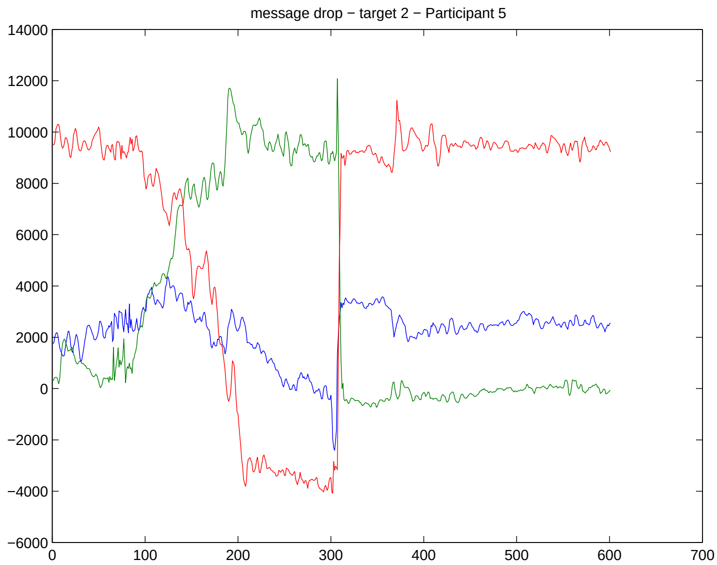


Figure 5.18: Acceleration output for participant 5 dropping a target at location 2.

5.6 Potential Applications

By including a separate density layer in any location-aware application it is possible to provide rich context-aware functionalities since this density has the potential to represent anything of interest to the user. Below we describe a number of potential applications for this system.

Figure 5.19: A user interacting with information left at each side of their garden path. Work objects are left on one side and leisure objects are left on the other side. The user can feel for and take any objects he is interested in or leave objects in the appropriate context for use later.

5.6.1 Social Networking

Social networking is one natural application for this kind of interactive location-aware system. There is already a multitude of social networking websites and even sites which inform users by text message, when they are in the vicinity of a friend, or even a friend of a friend as long as the user keeps the website updated with their current location (Google Inc. 2007). Using this density based approach it would be possible to keep track of friends and have their locations represented by a densities in the local vicinity, which it would be possible to interact with. It is easy to imagine

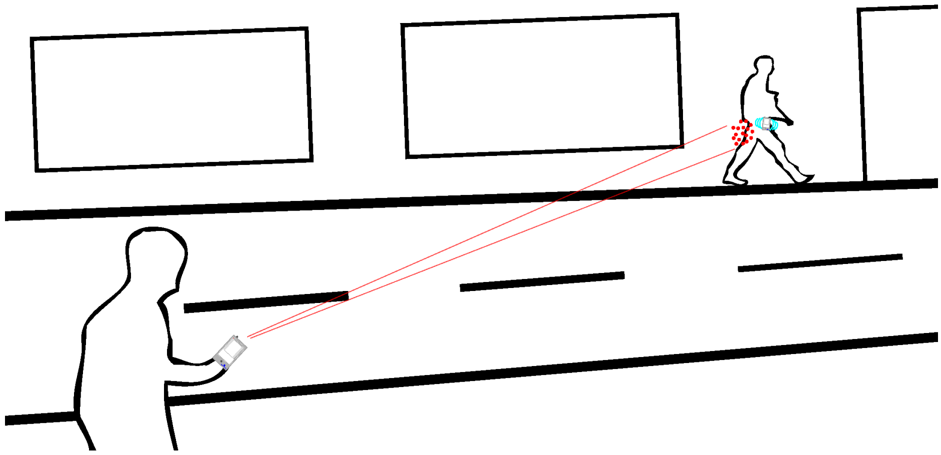


Figure 5.20: A user is actively interacting with another person's device at the other side of the street

interacting with a friend who was near by, by probing the local density representing that friend and have the effects of this probing made apparent on the other persons device as illustrated in figure 5.20. A context aware device could build a detailed picture of its owner's personal state, represented by a state vector, which could indicate, for example, the owner's current mood and whether they are open to contact with other people at that point in time. This creates the potential to build rich state vectors, which could be interacted with in this density, providing users with particular feedback depending on the particular structure the vector. Users can also have full control over their own density, which could become a parameter in their state vector. If a user didn't want to be contacted or disturbed by anybody they could increase the size of their density to cover a wide geographical area, it would then become difficult to locate them. If the user was available to be contacted they could decrease the size of their density down to a very local level.

5.6.2 Geographical Blogging

Recently we have seen the emergence of a number of so called 'geographical blogs'. One such example is a global public art project known as *Yellow Arrow* (Counts Media 2007). Yellow arrow stickers can be obtained from a website and placed anywhere, pointing to something of interest. The Yellow Arrow can signify that there is more to see in this particular location, such as a funny story, a memory or an interesting experience. Each arrow links digital content to a specific location using the mobile phone. When a sticker is found, a unique code printed on the sticker can be sent as a text message to a particular phone number. Moments later a text message is received with a message left by the sticker's original owner. By using the functionality of a system such as ours it becomes possible to eliminate the need for physical arrows and the sending of text messages. Arrows could be represented by densities placed over a map and users could sense the arrow in their local vicinity, automatically retrieving the message left by the previous person.

Another possibility for the use of a system such as this is for Geocaching (Peters 2004). Geocaching is a new and increasingly popular sport that involves using a hand-held GPS unit to travel to a specific longitude and latitude, often involving walking or hiking to natural locations, in search of objects hidden by other Geocachers. Using our system it is possible to build a Virtual Geocaching network, where Geocachers are instead rewarded with virtual objects or prizes. They would still be required to travel to specific latitude and longitude coordinates to receive their prize but once there, they are required to interact with and probe the local environment to find the virtual object represented by a local density *object*.

5.6.3 Demographics

Using a density layer to represent such demographics as socioeconomic levels or crime rates could be very useful. It is unlikely if your device sensed a high crime rate in a particular area, for example, that you would want to leave your car there. This kind of information can also be useful in situations where people may require quick information about the population structure of a particular area including such demographics as Age, Sex, Religion etc. By representing this information as a density and probing it with such an interface this information, and any fine structure that exists within this information, can be readily available. Chicago Crime (Holovaty A. 2007) is a website, which combines crime data for the Chicago area with Google Maps to provide local geographic information about crime in a particular area. It is easy to imagine creating a density map based on this crime data, which could be probed and explored using our interface at a local level.

5.6.4 Tourism

There have been a number of GPS based tourist guides developed in recent times. By and large these guides rely marking general areas of interest for users to pass by or locate on a map and move towards. By overlaying a density representing particular areas of interest it would be possible to guide tourists along a set route towards particular areas. This could be considered beneficial as in current systems, users are required to guide themselves, through any route they choose, into local 'hotspots' but in a system like this could be taken down specific paths or specific routes to particular areas of interest. This approach would also be useful for indoor guide systems if there was an indoor positioning system available.

5.6.5 File Sharing

With the release of the Microsoft Zune, we have seen the first mass market mobile device with active file sharing capabilities. It is possible using a Zune to share music with other people in the immediate vicinity and in the near future we will see the emergence of so-called 'Zune filling stations'. These are particular places, perhaps in a local McDonalds or Starbucks, where it is possible to download music for your player. This is follows on from the use of 'proximity servers', which are used to push information to people in close proximity with Bluetooth and infrared-enabled phones, usually at conferences. It is possible in this situation to dispense with the need for a proximity server and instead use a density layer to represent local areas of shared files. Users can navigate around their locality, through this density hearing and feeling close-by density objects and probing around with their device to locate their exact positions and receiving feedback depending on the contents of the object (Eslambolchilar and Murray-Smith 2006). Density objects could contain photos, videos or music, for example. It would then be possible, if in a particular area, a user found a music file that they enjoyed listening to, to take this file, effectively removing the file from the density. They could then carry this file with them until a point in time where they no longer wanted it and drop this song again wherever they may be at that point in time. This would then make the file available again for others to take from this new location.

5.6.6 Sculpting the Virtual Environment

Because various variables in the user's local environment are represented by a density it is possible to actively update that layer in real time. It becomes possible then to potentially sculpt and mould your own personal environment in an embodied manner by carving into 'blocks' of local density using the stream of Monte Carlo particles to effectively carve into the block. Work has been conducted into the audio perception of visual information (Wang and Jezekiel 1996, Rath and Rocchesso 2005). Hollander (1994) examined the ability of subjects to recognise geometric shapes and alphanumeric characters presented by sequential excitation of elements in a "virtual speaker array" finding that subjects were able to identify the patterns with "significantly better than chance" accuracy. The creative sonification of Monte Carlo particles impacting with a local density is a potential new way of conveying shape information via audio and vibrotactile feedback.

5.7 Discussion and Conclusions

In this chapter we have demonstrated the construction of an application which draws together the work conducted in chapters 2 and 3 to allow the dropping and retrieving of messages in the user's personal virtual environment. The system here is an extension beyond the simple target acquisition and trajectory-following applications presented in chapter 4. This is a system which has the potential to provide a general mechanism for providing highly interactive context-aware applications. By treating our system as a separate density layer in any application it is possible to provide different functionalities. We described a number of potential applications using this kind of density based context aware system.

As an example application we constructed the *airMessages* system, which is a combination of the *bodySpace* and *gpsTunes* systems described in chapters 3 and 4, enabling an embodied and gestural interaction for the searching and retrieving of messages left in the real world. Results from the field trial show that users are able to probe the local density effectively, using the full functionality of the designed interface to complete the task. We also found that users who used their interface functionality more fully were also the users who completed the task more effectively. This, although slightly subjective at this point, is promising evidence that this kind of embodied interface aids this kind of interaction.

Chapter 6

Conclusions

6.1 Theoretical Framework

The work in this thesis explores the design space of mobile devices equipped with inertial and location sensing and audio and vibrotactile feedback. We have demonstrated a new theoretical style of interaction design starting with the basic notion of treating this new kind of continuous interaction as a loop of control and building our application around this principle. We have demonstrated the necessity to think carefully about the inputs to this control system, the processing of those inputs and the feedback provided. We have demonstrated two distinct application interfaces built on a solid theoretical foundation and created interfaces using these principles, which use the egocentric body and exocentric real-world environment as interfaces for the interaction.

We have also demonstrated the utility and generality of a model-based approach to the interaction with mobile devices with the aim of allowing other HCI researchers to extract this approach and adapt it to their own interfaces. We have demonstrated the utility of incorporating uncertainty and constraints into the interaction design, which it is hoped can be adopted for the general improvement of interaction with location-aware applications.

6.2 BodySpace

For the egocentric bodySpace interface we developed an example application, which utilised inertial sensing and basic pattern recognition to enable the gestural control of a music player by placing the device at different parts of the body and gesturing to control the functionality, rather than having to press buttons or wear instrumented clothing.

We described a new general 'end-point' or goal based approach to the detection and segmentation of gestures and planar motion, showing that this approach could be used by a number of users in a small trial. We also demonstrated the use of a model-based approach to the design of this kind of interaction, demonstrating that interaction based on the simulation of a physical model of a ball in a bowl was both intuitive and easy for users to understand and may be easily applied to other interface designs. Although users displayed initial problems while using the system, as we would expect from the first use of any system, this initial testing provided us with some interesting usability and motor-control insights as to how our model based approach to this kind of interaction actually coped with real people. For example, we found that each user tended to have their own comfortable posture, which emerged after only a few minutes of practice, indicating that any system adopting this kind of approach would need some kind of personalisation, although this could be an iterative process. We also found that users were particularly susceptible to hand drift, which tended to cause a number of false positive recognitions. We also found that the participants were somewhat more limited with one 'flicking' direction than the other, with forward flicks of the device, when placed at the ear, being more successful than backward flicks.

The work in this chapter also allowed us to demonstrate the use of real-world constraints in the inference of user intention and how this notion of using real-world constraints can be considered as part of the interaction design process. One level of constraint that may be utilised in our particular example interface comes from the fact that the gesture is performed by the human arm, which is restricted in its potential movement. Another constraint comes from the fact that when the device is placed at a particular part of the body, it is restricted to a plane around a part of the body. We described the use of a dynamic representation of a bodySpace style gesture as a dynamic system, which could potentially enable the easier provision of formative feedback and how this kind of dynamic systems approach to representing gestures can be used in interaction design. We also demonstrated the use of tremor from our muscles as a potential new source of information in the inference process and as a potential proxy for pressure sensing in a mobile device equipped with accelerometers.

6.3 Whereable Computing

We developed an application, which utilised the 'real world' as an exocentric interface. We demonstrated that probabilistic, multimodal, handheld interaction techniques can be applied effectively to allow users to explore density functions in space, with the example of pedestrian GPS navigation. The Monte Carlo sampling method provides both an effective way of integrating probabilistic models into practical interfaces and of displaying the results in a multimodal fashion, which could be of great use to interface design in general. We described the extension of this system to one which provides a general mechanism for providing highly interactive context-aware applications. By treating our system as a separate density layer in any application it is possible to provide different functionalities. The densities here could represent $P(C_i|x)$ - the probability of context state C_i given the current state vector x.

We have shown that feeding back uncertainty can improve performance in location-based interaction and that the use of natural constraints in the environment, similar to what we achieved in the design of the bodySpace interface, can aid the interaction. We have also shown that the use of tilt to control the Monte Carlo sampling time horizon and the use of magnetometers to provide rapid bearing updates aided the design of a more embodied kind of interaction in this context.

We conducted two main field trials, the first of which supported our hypothesis that the uncertain displays required less effort and results in more stable behaviour. This field trial also allowed us to introduce new metrics for usability analysis, which provide objective measures of the way in which a system was used and removed the need to rely on the more traditional subjective measures. The second field trial showed that it is possible to guide users to a desired location over a set trajectory or path and a number of interesting behaviours were observed, which we attempted to classify into certain categories. Interactive sonification of the exploration process produced a navigation system which could be used in an eyes-free manner, where the user brings their sensorimotor systems into the interaction with the augmented environment for a more embodied interaction. This initial data from these trials has shown that it is possible for users to navigate through a set path over a featureless playing field using audio and vibrotactile feedback alone. Their performance and confidence improves significantly when the audio and vibrotactile constraints from the system are coupled with the natural constraints of the environment, suggesting that the system is promising for a range of realistic use cases.

We also demonstrated the potential possibilities for modelling this kind of behaviour using a simple control model. We have demonstrated initial steps towards creating a control model of human behaviour to help us understand some of the behaviour exhibited by participants in our field trials. This approach, when more fully developed, has the potential to aid the design of future interaction and interfaces in general as it has the potential to give prior knowledge of how users may perform with a particular interface design.

6.4 Outlook

This thesis has demonstrated the development of a new kind of location-aware computing and there is great scope for extension of the ideas presented here. In chapter 2 we introduced the idea of using a control theoretic approach to interaction design and in subsequent chapters we demonstrated the use of this approach. For example, in chapter 3 we introduced the concept of controlling a system from a simulation of a physical model and in chapter 4 we demonstrated the development of a simple model of human behaviour. This control-theoretic approach to the design of this kind of interface can improve both the use of this interface and enable the generalisation of this approach for other HCI researchers. The appropriate treatment of uncertainty is something critical for the successful use of location-aware systems and we have shown in chapter 4 that the explicit use of uncertainty in our interface design has proved to be beneficial to the interaction process. Likewise, our embracing of natural constraints in the environment has been proven to aid the interaction process. In chapter 3 we demonstrate the use of explicit constraints around the body to shape the gestural interaction for our gesture-based interface and in chapter 4 we use the natural constraints of the local environment to infer future user positions in our location-aware interface. By applying this kind of approach to all areas of interaction design, it not only has the potential to greatly improve interaction with this new kind of system but also increase the general acceptance of these novel approaches to the larger interaction design community.

The use of inertial sensing to create a more embodied and highly interactive style of interaction in this location-aware context has shown that it is possible for users to engage with a system and interact with objects placed in their own personal egocentric or exocentric virtual worlds. This kind of interaction with virtual objects opens the door for the development of an abundance of novel applications. Virtual objects can take a number of forms; local objects of information, text messages or even other people. There is great potential for the development of social networking applications in this context, which allow people to interact and negotiate directly with friends in their personal virtual worlds. This has the potential to change the way that people think about location and context-aware computing. Systems change from static, unresponsive on/off systems to dynamic, responsive, flowing, highly interactive systems and ultimately the work presented in this thesis has the potential to become a basis for this rapidly growing field.

Appendix A

The Global Positioning System

The Global Positioning System (GPS) consists of 32 satellites orbiting the Earth, transmitting radio signals which enable GPS receivers anywhere in the world to determine their location, altitude and speed. The first experimental satellites were launched in 1978 and GPS has since become indispensable for navigation around the globe.

A.1 Navigation

Navigation is the principle application of the Global Positioning System. But how does this work? A GPS receiver calculates its position by measuring the distance between itself and three or more GPS satellites using a technique known as trilateration (Bajaj *et al.* 2002), a method of determining the relative positions of objects using the geometry of triangles, not dissimilar to triangulation, as illustrated in figure A.1. Knowing the position and the distance of a satellite indicates that the receiver is located somewhere on the surface of an imaginary sphere centered on that satellite and whose radius is the distance to that satellite. When four satellites are measured simultaneously, the intersection of the four imaginary spheres reveals the location of the receiver, according to the World Geodetic System WGS84 (National Imagery and Mapping Agency (NIMA) 1991) coordinates system. Often, these spheres will overlap slightly instead of meeting at one point, so the receiver will provide a mathematically most-probable position and indicate the uncertainty in this estimate.

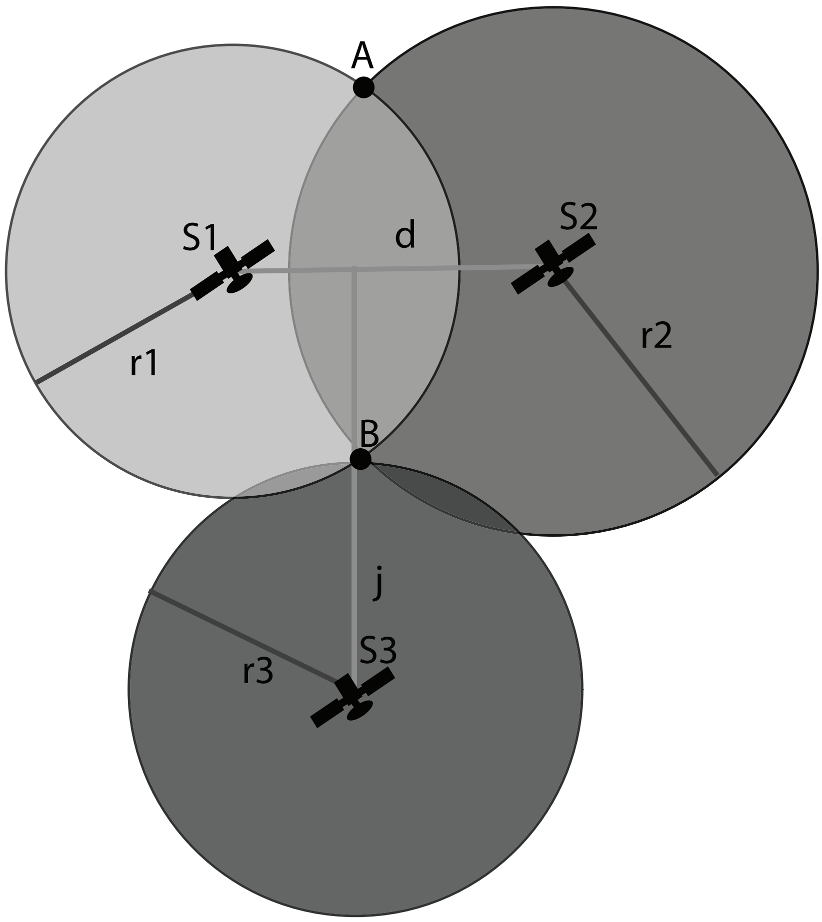


Figure A.1: Standing at B, you want to know your location relative to the reference satellites S1, S2, and S3 on a 2D plane. Measuring $r1$ narrows your position down to a circle. Next, measuring $r2$ narrows it down to two points, A and B. A third measurement, $r3$, gives your coordinates at B. A fourth measurement could also be made to reduce error. Figure adapted from (Bajaj 2002)

A.2 Accuracy

The position accuracy calculated by any GPS receiver is primarily dependent on the satellite geometry and signal delay but can be affected by a number of different sources. Figure A.2 shows a log of GPS data over an 8 minute period. The number of satellites visible in this time varied between 4 and 8 and we see that there is considerable variation in the receivers estimated position. So what factors were contributing to this variation?

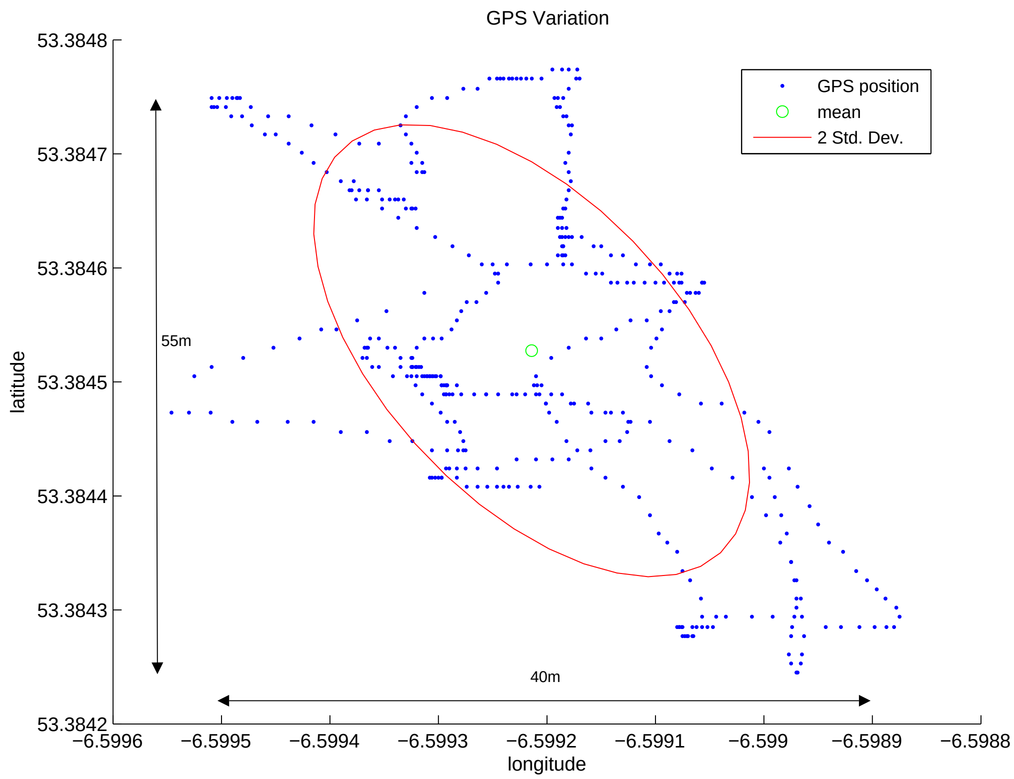


Figure A.2: GPS position variation while the unit is at standstill in an eight minute period. Units are degrees.

A.3 Sources Of Error

A.3.1 Satellite Geometry

The most significant factor affecting the accuracy of a GPS measurement is the "satellite geometry", which describes the position of the satellites to each other from the view of the receiver. We have a "good" geometry if all the satellites that our receiver can currently see are well distributed across the sky leading to the kind of geometry illustrated in figure A.3-A. In this case we can take position estimates with an error of as little as 2-3 m. A so called "bad" geometry arises if all currently locked satellites appear in the same part of the sky as illustrated in figure A.3-B. This kind of geometry can, in the worst case lead to no position estimate at all but generally this kind of bad geometry will cause an error of 100-150m. To indicate the quality of the satellite geometry, the DOP values (dilution of precision) are commonly used. There are five variants of DOP:

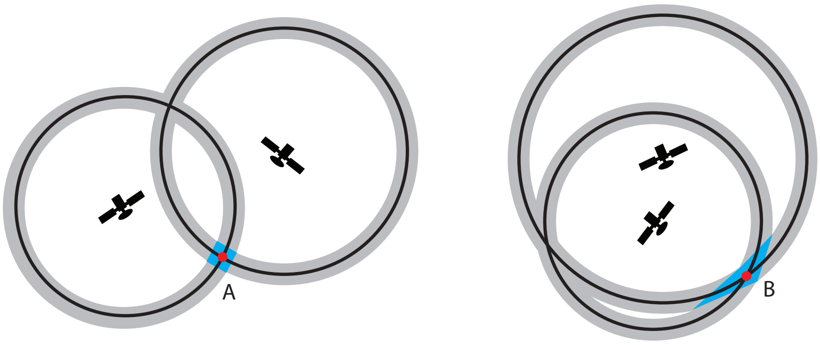


Figure A.3: The blue area at the point of intersection of the two circles indicates the possible positions of the receiver, given the uncertainty in the satellite position indicated by the grey circles. In the good case (case A) the blue area is small indicating good geometry. In the bad case (case B) the blue area is larger, indicating bad geometry. Figure adapted from (Köhne 2007)

- Geometric Dilution Of Precision (GDOP) - Overall-accuracy in 3D-coordinates and time
- Positional Dilution Of Precision (PDOP) - Position accuracy in 3D-coordinates
- Horizontal Dilution Of Precision (HDOP) - horizontal accuracy in 2D-coordinates
- Vertical Dilution Of Precision (VDOP) - vertical accuracy in height
- Time Dilution Of Precision (TDOP) - time accuracy

Generally speaking HDOP-values below 4 are good and above 8 are bad and for an accurate position determination, the GDOP value should not be smaller than 5. (El-Rabbany 2002).

A.3.2 Signal Shadowing

Signal shadowing describes the situation when the line of sight to a satellite is obscured by a large object or mountain. In urban environments this is a significant problem since the skyline generally has a higher elevation, restricting the amount of sky that can be seen by the receiver and decreasing the likelihood that the receiver will see the minimum 3 satellites required to make a positional fix. Figure A.5 illustrates the situation where a building is obstructing the path to the receiver. Because the satellites are in non-stationary orbits, even if a GPS unit is in a static position, the GPS availability will change over time making signal shadowing a significant problem. Steed (2004) describes a tool *satview*, which visualises the current likely availability of GPS coverage.

A.3.3 Atmospheric Effects

Changing atmospheric conditions can change the speed of GPS signals and can have a significant effect on the accuracy of GPS signals. These effects though are minimised when the satellite is directly overhead, and become greater for satellites nearer the horizon, which is why GPS accuracy is inherently lower at extreme latitudes, since the signal is affected for a longer time. The effects of the ionosphere are generally slow-moving, and can be averaged over time making it relatively easy to remove this effect. This effect is illustrated in figure A.4. Humidity can also be a source of error for GPS signals. This effect is much more localised, occurring in the troposphere, and changes more quickly than ionospheric effects, making precise compensation for humidity more difficult. Altitude also causes a variable delay, as the signal passes through less atmosphere at higher elevations. Since the GPS receiver measures altitude directly, this is much simpler correction to apply (A. and M. 2007).

A.3.4 Ephemeris and clock errors

The navigation message from a satellite is sent out only every 12.5 minutes but in reality, the data contained in these messages tend to be “out of date” by an even larger amount. Consider the case when a GPS satellite is boosted back into a proper orbit; for some time following the maneuver, the receivers calculation of the satellite’s position will be incorrect until it receives another ephemeris update. The onboard clocks are extremely accurate, but they do suffer from some clock drift. This problem tends to be very small, but may add up to 2 meters (6 ft) of inaccuracy.

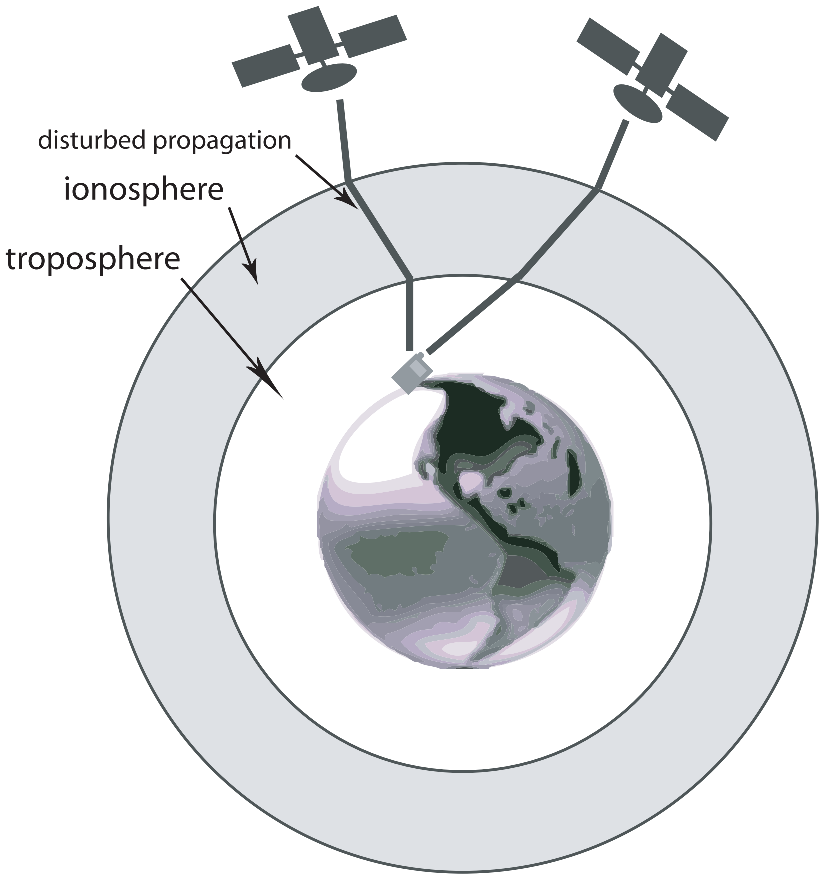


Figure A.4: the effect of the earths atmosphere on the radio signals from the GPS satellites

A.3.5 Multipath Effects

GPS signals can also be affected significantly by multipath effects, where the radio signals from the satellites are reflected off surrounding buildings, mountains, hard ground, etc. These delayed signals can cause inaccuracy. A variety of techniques have been developed to reduce multipath errors and for long delay multipath, the receiver itself can recognise the delayed signal and ignore it. Multipath effects though are much less severe in moving vehicles. When the GPS antenna is moving, the false solutions using reflected signals quickly fail to converge and only the direct signals result in stable solutions but this is a significant source of error for pedestrian GPS based applications.

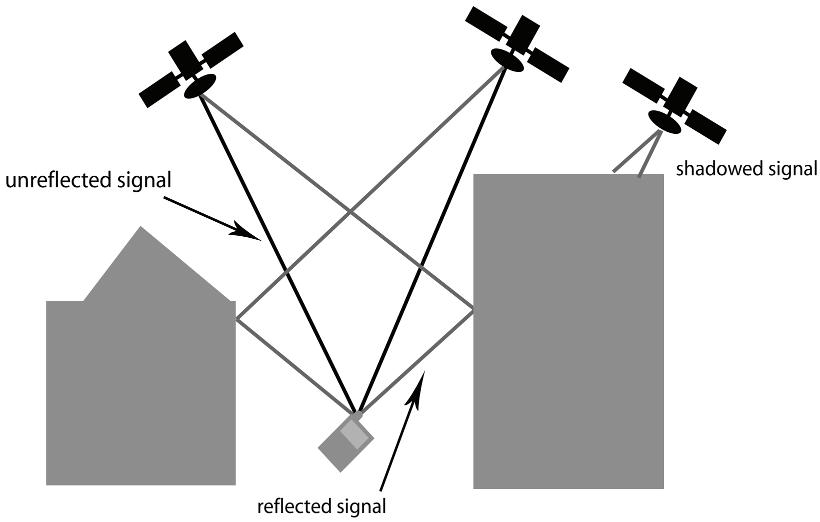


Figure A.5: interference from signal reflections and signal shadowing

A.4 Other Applications

The GPS is not just used for military or navigation applications. Surveying and mapping is one common use for the GPS. Survey-Grade GPS receivers can be used to position survey markers, buildings, and road construction. High precision measurements of crustal strain can be made with differential GPS. This works by finding the relative displacement between GPS sensors. Multiple stations situated around an actively deforming area, such as a volcano or fault zone, can be used to find strain and ground movement. These measurements can then be used to interpret the cause of the deformation. The availability of hand-held GPS receivers has also led to the development of games as mentioned in a previous chapter. One such game is Geocaching (Peters 2004), which is a new and popular sport that involves using a hand-held GPS unit to travel to a specific longitude and latitude to search for objects hidden by other geocachers. This popular activity often includes walking or hiking to natural locations. Combining GPS position data with photographs taken with a digital camera allows people lookup the locations where the photographs were taken on a website (Spinellis 2003) and automatically annotate the photographs with the name of the location they depict.

Appendix B

Inertial Sensing For Mobile Devices

B.1 Introduction

Although inertial sensing has been researched extensively over a long period of time and there exists a wealth of detailed literature on the application of the tools developed for this field, there is a lack of information about how to apply these techniques in a mobile or handheld domain. This is understandable as until recently there was no real demand for this kind of information. For this reason we describe here some of the main points relevant to the mobile domain and try to apply these in a relevant way. To describe these techniques in a detailed way would be futile, since it is unlikely that we could ever reproduce the kind of tight navigation that we see from robotics or military applications. Despite the obvious differences such as the small-scale, rapidly changing movements on a mobile device compared to the large scale, more constant movements on an aircraft we must also consider some more subtle details such as the physical sensors we use, which are cheaper and less accurate, and the kinds of angular rates and accelerations we would expect to see in a mobile application compared to that of an aircraft or missile application.

B.2 Mobile Movement

The types of movement we experience with an instrumented mobile phone are highly variable. The kinds of movement sensed range from discrete hand gestures up to large arm gestures. We may also detect motion from our general environment or 'context'. The device may pick up movement from the train your're riding in or movement from your pocket while you walk down the street so this variability provides us with much more complicated problem than the traditional aircraft or missile traveling in a straight line for long periods of time. Human physiology is another factor. The human body contains a complex system of oscillators, which make up the human body and transfer movement into our mobile phones. Tremor from our muscles adds another dimension to this problem which again adds to the complexity of this task and makes it distinct from the more traditional problems.

It's not all negative though. Human physiology, may actually work as a constraint on the potential range of movements. The same is true for movement of the human arm and hand/wrist, which has a finite range of movement so these constraints in the potential range of motion of a typical mobile device may act to simplify the problem slightly if we possess a detailed knowledge of this range of possible movements.

B.3 Coordinate Systems

While the emergence of location-aware and context-aware computing has opened paved the way for a wealth of new applications, it also poses a number of challenges. Traditional navigation theory is based around the prior definition of a number of reference frames. Traditional navigation around the Earth requires the definition of axis sets, which allow inertial measurements to be related to the cardinal directions of the Earth, that is, frames which have a physical significance when navigating in the vicinity of the Earth (Roth 1999). We will first consider the case of navigation round the Earth then attempt to apply some of these ideas to navigation with an instrumented mobile device.

Each frame is defined as an orthogonal, right-handed axis set as shown in figure B.1. The first frame we define for navigation around the earth is referred to as the *inertial frame (i-frame)*. This frame has its origin at the centre of the Earth and axes which do not rotate with respect to the fixed stars (Britting 1971). The orientation of the coordinate axes may be chosen arbitrarily but it is usual to picture the z_i axis running from south to north along the earth's rotation axis with the orthogonal x_i and y_i axis in the equatorial plane, i.e. the plane normal to the earths rotation axis. This frame may seem too 'large-scale' in the context of a mobile device but it is a necessary and useful basis for the definition of the rest of our reference frames. We then define the *earth frame (e-frame)*, which again has it's origin at the centre of the Earth and axes which are fixed with respect to the Earth. The z_e axis runs from south to north along the rotational axis of the earth and the x_e and y_e axes again lie in the equatorial plane rotating with an angular rate Ω with respect to the *inertial frame*. The *navigation frame* has its origin at the location of the navigation system and has axes aligned with north, east and the local gravity vector $\mathbf{g}$. The x_n and y_n axes lie in the local horizontal plane and have a turn rate ω_{en} with respect to the earth frame, often referred to as the 'transport rate' (Titterton and Weston 2004). The *accelerometer frame* is an orthogonal axis set whose origin is the point where motion is measured by the accelerometers. We may also define an analogous *gyroscope frame* whose origin measures the point of motion. The *body frame* is an orthogonal axis set, which is aligned with the roll, pitch and yaw axes of the vehicle or device to which the navigation system is 'strapped'. We use this frame to describe the orientation of our device. We will assume for simplicity that the origin of this frame is coincident with the origins of the accelerometer and gyroscope frames.

What defining these frames allows us to do is picture the device in varying situations, which may arise for different applications. For the BodySpace application described in chapter 3 we can imagine the principle movements of interest happening in the body frame. Whereas for the gpsTunes application in chapter 4 we are principally interested in a combination movements from the navigation frame for general movement around an area and from the body frame for the inertial interface control.

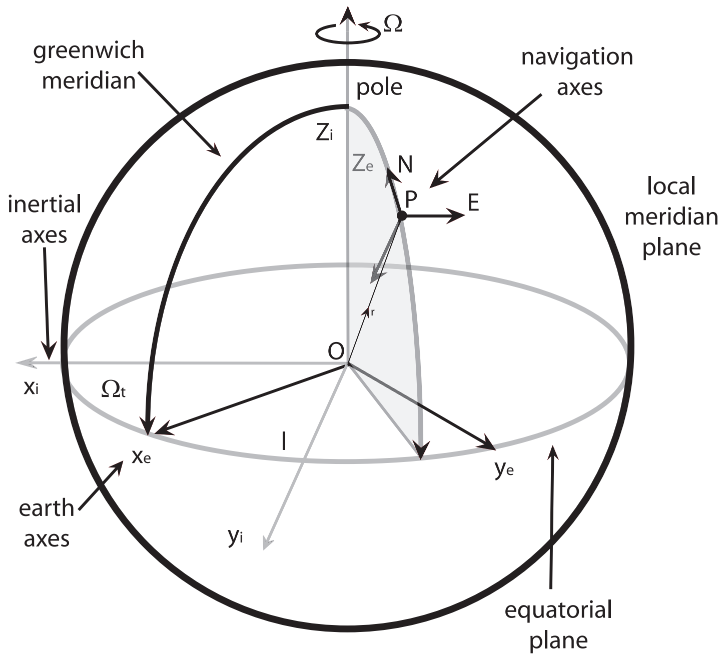


Figure B.1: The Earth, Inertial and Navigation frame axes. Figure adapted from (Roth 1999)

B.3.1 Navigation Equation

The navigation equation allows us to generate estimates of acceleration, velocity and position in our desired reference frame. We can assume that we will be required to navigate with respect to a fixed, or non-accelerating, non-rotating set of axes. The particular component of acceleration in the direction the movement, referred to as the 'specific force', and estimates of the gravitational field are summed to determine components of acceleration with respect to a space fixed reference frame (Titterton and Weston 2004).

Let **r** represent the position vector of a point P on the sphere in figure B.1 with respect to O, the origin of the reference frame. The acceleration of P with respect to a space-fixed axis set, i.e. the *i-frame*, is defined by:

$$\mathbf{a}_i = \frac{d^2\mathbf{r}}{dt^2}|i \tag{B.1}$$

From our accelerometers we can take a measure of the specific force, **f**, acting at a point **P** where

$$\mathbf{f} = \frac{d^2\mathbf{r}}{dt^2}|_i - \mathbf{g} \tag{B.2}$$

where **g** is the mass attraction gravitation vector. Rearranging we get

$$\frac{d^2\mathbf{r}}{dt^2}|_i = \mathbf{f} + \mathbf{g} \tag{B.3}$$

which is known as the navigation equation. To obtain the velocity in the *i-frame* we integrate one time

$$\mathbf{v}_i = \frac{d\mathbf{r}}{dt}|_i \tag{B.4}$$

with a second integration theoretically giving its position in that frame.

In practice we will often be required to resolve velocity and position with respect to a rotating reference frame, when navigating in the vicinity of the earth for example. In this situation we will need to revise the navigation equation slightly to take into account additional apparent forces acting, which are functions of reference frame motion.

To obtain the velocity in the *earth-frame* from the velocity in the *i-frame* we may use the theorem of coriolis, as follows,

$$\mathbf{v}_e = \mathbf{v}_i - \omega_{ie} \times \mathbf{r} \tag{B.5}$$

where $\omega_{ie} = \begin{bmatrix} 0 \\ 0 \\ \Omega_z \end{bmatrix}$ is the turn rate of the *e-frame* with respect to the *i-frame.*

Accelerometers usually provide measures of specific force in a body fixed axis set, denoted $\mathbf{f}^b$. In order to navigate it is necessary to resolve the components of the specific force in the chosen reference frame. If we choose the inertial frame, for example, we may resolve the components of specific force by multiplying the body fixed measurements, $\mathbf{f}^b$ by the direction cosine matrix, $\mathbf{C}_b^i$ using

$$\mathbf{f}^i = \dot{\mathbf{C}}_b^i \mathbf{f}^b \tag{B.6}$$

where $\mathbf{C}_b^i$ is a 3x3 matrix which defines the attitude of the body frame with respect to the i-frame. The direction cosine matrix $\mathbf{C}_b^i$ may be calculated from the angular rate measurements provided by our gyroscopes using the following equation:

$$\mathbf{C}_b^i = \mathbf{C}_b^i \Omega_{ib}^b \tag{B.7}$$

where Ω_{ib}^b is the skew symmetric matrix:

$$\Omega_{ib}^b = \begin{bmatrix} 0 & -r & q \\ r & 0 & -p \\ -q & p & 0 \end{bmatrix} \tag{B.8}$$

This matrix is formed from the elements of the vector $\omega_{ib}^b = \begin{bmatrix} p & q & r \end{bmatrix}^T$ which represents the turn rate of the body with respect to the *i-frame* as measured by the gyroscopes.

Determining the orientation of our device is one of our main aims. The orientation of the device is described by the relative difference between the axes of the *body-frame* and the *navigation-frame*. The orientation at a time t after the start of motion is a function of the initial motion at $t = 0$ and the angular motion of the device which followed. The angular motion is thus defined by the time history of the angular velocity of the *body-frame* relative to the *navigation-frame* frame (Roth 1999).

B.4 Sensors

In a typical Inertial Measurement Unit (IMU) the essential sensors are accelerometers, gyroscopes and magnetometers. Any other sensors added may aid the system in some way but are not essential. The construction of devices which are used to sense motion may be classified as either mechanical or solid-state. Mechanical accelerometers, for example, are well established and can provide highly accurate measurements of acceleration even down to a few micro-g in some cases. These sensors though are generally very large, larger than your average mobile device, and so we must find a suitable alternative. For this reason we focus on solid-state sensors which have made significant advances in recent years in terms of their size and accuracy.

B.4.1 MEMS inertial sensors

The sensors used in a typical IMU for a mobile device are Micro-machined Electromechanical System or 'MEMS' sensors. New applications for inertial sensing have, in recent times, demanded much smaller, less power consuming, less expensive sensors and MEMS technology has successfully fulfilled these demands. However, the introduction of MEMS technology will bring with it more limitations. In general, they bring a decrease in sensitivity/scale factor and an increase in noise. It may also make thermal sensitivity much more of a problem since silicon is very sensitive to thermal fluctuations. Despite these limitations though, MEMS sensors provide good enough performance for the acceptance of this trade-off for the reduction in size and price alone.

B.4.2 Accelerometers

In a basic way, accelerometers are essentially mimicking the human vestibular system. This system is essential for stable posture control and enables humans to move freely since it is not earthbound. This is also the system utilised by our brain to measure head movements without a frame of reference. Recently, micro-machined inertial sensors, i.e. accelerometers and gyroscopes, have become much more widely available. They are small in size, can be worn on the body and like the vestibular system, the working principle of these sensors is based on omnipresent inertia, enabling measurement anywhere without the need for a frame of reference (Luinge 2002).

To give a feel for how exactly an accelerometer works we may consider the 'mass in a box' analogy where we imagine a mass suspended inside a box by a spring, as in figure B.2. This mass is allowed to move in one direction which is the sensitive direction of the accelerometer. The displacement of the mass with respect to the casing is proportional to the acceleration along that axis. We can imagine 3 such accelerometers with orthogonal orientations giving us a measure of the 3D acceleration.

B.4.3 Gyroscopes

Gyroscopes are used to sense the angular rate of turn about an axis. Like accelerometers, gyroscopes can come in a number of different forms. Spinning gyros, laser gyros and vibrating mass gyros are the most common form in use today. The spinning and laser varieties of gyroscope are mainly used for large-scale navigation are not suitable for use in a mobile device, since they are both expensive and large (Söderkvist 1994). Vibrating mass gyroscopes on the other hand are ideal for incorporation into mobile devices because they are small, inexpensive and have a low power requirement. A

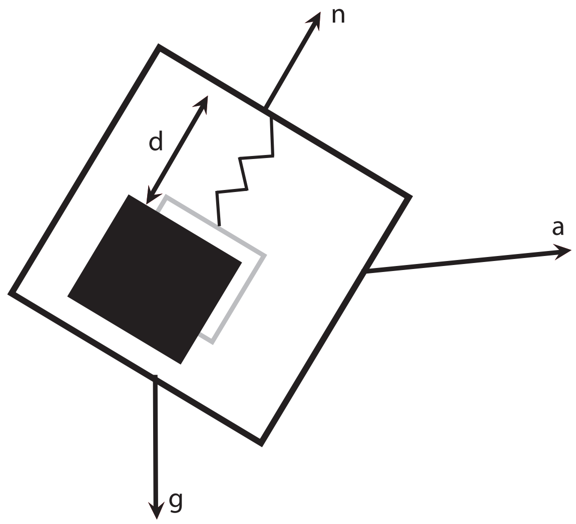


Figure B.2: This figure conveys the 'mass in a box' representation of an accelerometer whereby a mass is suspended by a spring. Any displacement of the mass with respect to the outer casing is mapped to a corresponding acceleration. Figure adapted from (Luinge 2002)

vibrating mass gyroscope, as would be used in most mobile applications, is based on the principle of a vibrating mass undergoing an additional vibration caused by the coriolis effect as in figure B.3. It consists of a mass, actuated in the direction given by r_{act}. The displacement of the mass is measured in the direction perpendicular to the actuation direction. If the box is rotated with an angular velocity perpendicular to the plane, the mass will experience an apparent force in the direction perpendicular to the angular velocity and momentary mass speed. The displacement of the mass in the direction perpendicular to r_{act} is proportional to the angular velocity of the system. This force is present only in the sensor coordinate system, not in the inertial coordinate system (Luinge 2002).

B.4.4 Magnetometers

A magnetometer is a sensor used to measure the strength of the earths magnetic field. The earth has a magnetic field which resembles that of the simple bar magnet with field lines originating at the south pole and and terminating at the north pole. The field lines have slightly varying strength

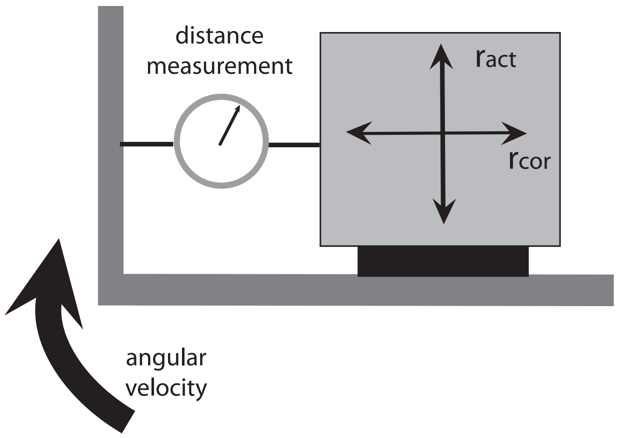


Figure B.3: A vibrating mass gyroscope: a mass is actuated in the direction given by r_{act}. If the box is rotated with an angular velocity perpendicular to the plane, it will experience an apparent force in the direction perpendicular to the angular velocity and momentary mass speed. The displacement of the mass in the direction perpendicular to r_{act} is then proportional to the angular velocity of the system. Figure adapted from (Luinge 2002)

and direction at different points around the earth but at a local level we may think of these fields as being constant and use them as a reference, given a suitable calibration. A typical IMU will usually contain three 'vector magnetometers' which have the ability to measure the component of the magnetic field in a particular direction, in this case along the x, y or z axes. The use of three orthogonal vector magnetometers allows us, in theory, to calculate the magnetic field strength, inclination and declination of our mobile device. In reality though, this is hampered by the rapidly varying fields in a typical urban environment due to large metal structures perturbing the local field. We may achieve a reliable magnetic heading from our magnetometers after a suitable calibration for the local environment.

The most common kind of magnetometers used in mobile applications are fluxgate magnetometers. A fluxgate magnetometer consists of a small, magnetically susceptible, core wrapped by two coils of wire. A current is passed through one coil, causing an alternating cycle of magnetic saturation. This creates an electrical field in the other coil, which is measured. If the magnetic background is neutral the input and output currents will match but if there is a magnetic field present the current will be magnetised in alignment with that field, giving us a way of measuring that field.

B.5 Sensor Output

When we take raw data from an accelerometer what are we actually seeing? Is this pure acceleration information? The straight answer is 'no'. What we are actually seeing is the combination of acceleration, systematic errors and noise which are characteristic of any physical measurement. As was mentioned before it is possible to think of an accelerometer as a mass suspended inside a box by a spring with the displacement of the mass, with respect the edge of the box, thought of as being proportional to the difference between the inertially referenced acceleration and the gravitational acceleration acting along the accelerometers sensitive axis. The difference between the inertially referenced acceleration and the gravitational acceleration is referred to here as the specific force (Roth 1999).

The accelerometer may be thought of as producing an output o_a, modeled as being equal to:

$$o_a = S_a a^a_{SF} + b_a + e_a \tag{B.9}$$

where the vector a_{SF} is the specific force vector at the origin of the *navigation-frame* and the term $S_a a^a_{SF}$ reflects the ideal linear response of the accelerometers. The matrix S_a is called the accelerometer response matrix. It is a diagonal matrix:

$$S^a = \begin{vmatrix} S_{ax} & 0 & 0 \\ 0 & S_{ay} & 0 \\ 0 & 0 & S_{az} \end{vmatrix}$$

where the elements S_{ax}, S_{ay} and S_{az} are the ideal linear scale factors of the x, y and z accelerometers respectively. The vector b_a is the accelerometer bias vector which describes the offsets that may be present in the output components from the sensor and is determined at calibration. The vector e_a is the accelerometer noise vector. These include errors from non-linearities or hysteresis in the accelerometer responses and errors due to the fact that the accelerometers do not measure at exactly the origin on the *accelerometer-frame* (Roth 1999).

Output from the gyroscopes is defined in a similar way to that of the accelerometers. The gyroscopes may be thought of as producing an output

o_g, modeled as:

$$o_g = S_g \omega_{ib}^g + b_g + e_g \tag{B.10}$$

where the vector o_g is the output of our three gyroscopes and the vector ω_{ib} is the angular velocity of the *body-frame* with respect to the earth centred *inertial-frame*. $S_g \omega_{ib}^g$ describes the ideal linear response of the gyroscopes and as with the accelerometer, the matrix S_g is called the gyroscope response matrix:

$$S^g = \begin{vmatrix} s_{gx} & 0 & 0 \\ 0 & s_{gy} & 0 \\ 0 & 0 & s_{gz} \end{vmatrix}$$

where s_{gx}, s_{gy}, s_{gz} are the response factors of the x, y and z gyroscopes, defined in a similar way to that of the accelerometers. The vector b_g is the gyroscope bias vector and describes the biases which exist in the gyroscopes. The vector e_g is the gyroscope noise vector, which is similar to the accelerometer noise vector, however, unlike the the accelerometers there are no errors introduced from the fact that all the gyroscopes do not measure from the same point, since we know that the angular velocity is the same at all parts of a rotating rigid body.

B.5.1 Sensor Placement

When building an inertial measurement unit it is necessary to consider the effects that placing the sensors in different areas of the IMU will affect the output. In almost all formulations of the reference frames required to describe a typical INS, accelerometers are *theoretically* placed at the origin of the *accelerometer frame*. This is obviously always an approximation since the finite sizes of the MEMS accelerometers stop each sensor from measuring at exactly the same point but it is a solid approximation to make. There exists though a fundamental difference between accelerometer and gyroscope positioning. The point in our mobile device where the gyroscopes measure is actually of no relevance since from first principles we know that the angular rate at any point inside a rotating rigid body has the same angular velocity and we may think of our mobile device as a rigid body. This is not the case for acceleration so the point from which we are measuring motion is of great importance and must be known. What if, for reasons of limited real estate and size restrictions on the sensor housing etc. we wish to move the accelerometers off of the sensor 'origin' by some known distance? What affect does this have on our measurements?

Rigid Body Kinematics

We may thnk of the rotation of a mobile device as the 1D rotation of a rigid body and we may define the angular velocity of a point on a the rotating body as ω. This value doesn't change from point to point on the rotating rigid body. When we have determined the angular velocity of our point on the body, the velocity is simply

$$v = \omega \times r \tag{B.11}$$

where $\mathbf{r}$ is the position vector of the point considered with respect to the origin of our reference frame, i.e. the centre of our rotating body. The acceleration of any point in the rigid body is then obtained by taking the derivative of equation 1. Thus,

$$a = \dot{\omega} = \dot{\omega} \times r + \omega \times \dot{r} = \alpha \times r + \omega \times (\omega \times r) \tag{B.12}$$

where α is the angular acceleration vector. So we see that the acceleration at any point in the rigid body is determined by $\mathbf{r}$ as ω and α are constant. So as we increase $\mathbf{r}$, the distance from the origin, the measured acceleration is increased. To examine the effects of increasing $\mathbf{r}$ we consider the rotation around one axis and examine the varying theoretical accelerations We see from figure B.4 shows that as the distance, r from the origin is increased, areas of higher accelerations are amplified somewhat whereas lower accelerations are relatively unchanged.

B.6 Basic Algorithms

We introduce now some basic tools required for working with these kind of sensors. One of the main things to consider is calibration.

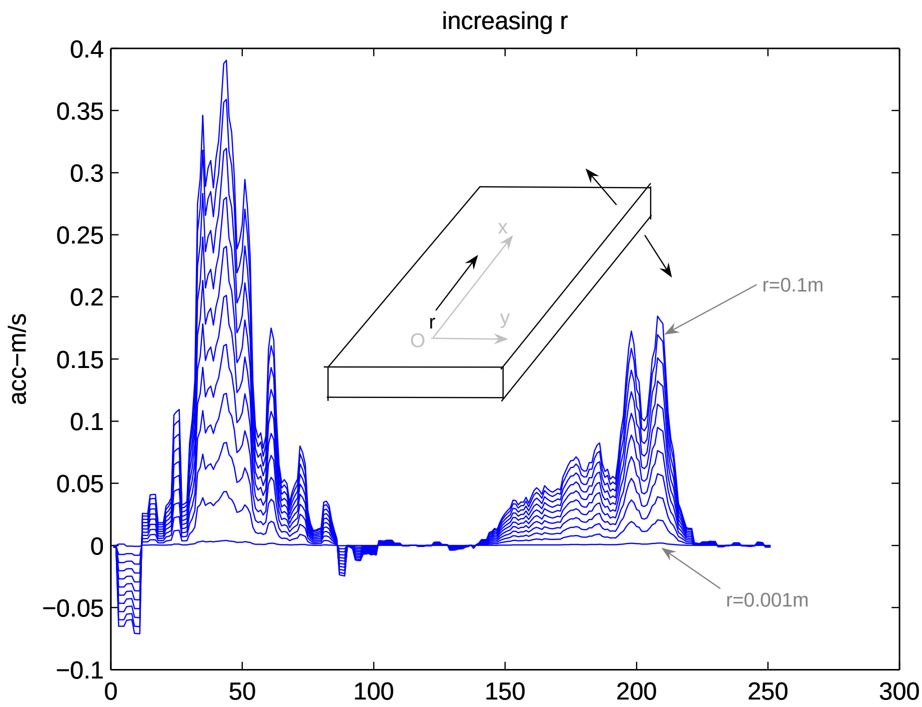


Figure B.4: As the distance r, the distance of the accelerometers from the origin of the body-frame is increased the measured acceleration becomes increasingly amplified

B.6.1 Calibration

Before it is possible to work with any sensor data it is necessary to perform some simple calibrations. It is necessary to calibrate separately the accelerometers, gyroscopes and magnetometers.

B.6.2 Accelerometer Calibration

Calibration of the accelerometers is not necessary in all situations. For a gesture recognition application it may actually be better to work with raw accelerometer data where as for a tilt application, if we are not working with the full derived strapdown equations, the data needs to be quickly calibrated or 'zero'd' at the beginning of each use. Zeroing the data essentially just involves defining the rest point of the device i.e. the values from the accelerometer where the device is flat in the user's hand. Any tilting of the device will then give a deviation from these 'zero' values which are then used as indicators that the device is being tilted. A more rigorous way to

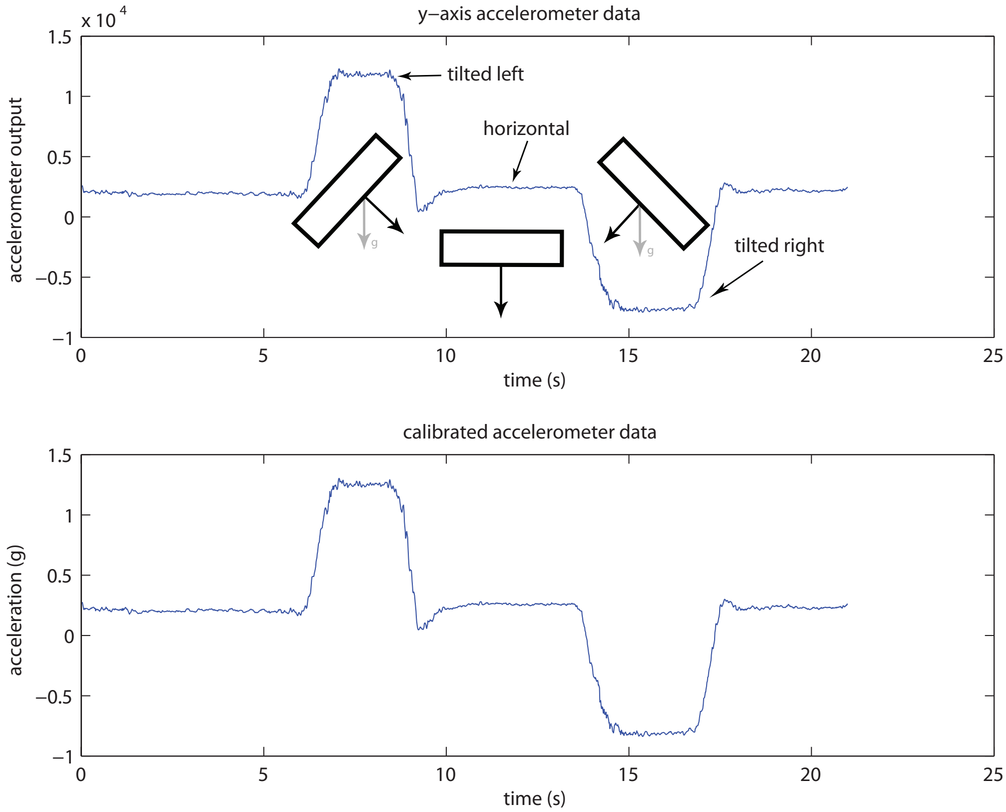


Figure B.5: accelerometer data for a device first tilted left then back to horizontal, then tilted right

calibrate accelerometers, in order to take an actual acceleration value from them is to first measure a value for gravity, **g**. We can then divide the output value from the accelerometer in order to achieve a value for acceleration as a function of **g**. Measuring a value for **g** needs to be performed for each of the x, y and z accelerometers since there may exist slight differences in the output for each. This simply involves holding the device in the appropriate rotation for your chosen accelerometer, as illustrated in figure B.6, and noting the output value, this value is then the value for **g**.

B.6.3 Gyroscope Calibration

We calibrate the gyroscopes in order to gain estimated angular rates in radians per second from the raw sensor data. The calibration of the gyroscopes involves rotating the device through 360 degrees, on a flat surface and examining the output from the x, y or z gyroscope. Figure B.7 shows

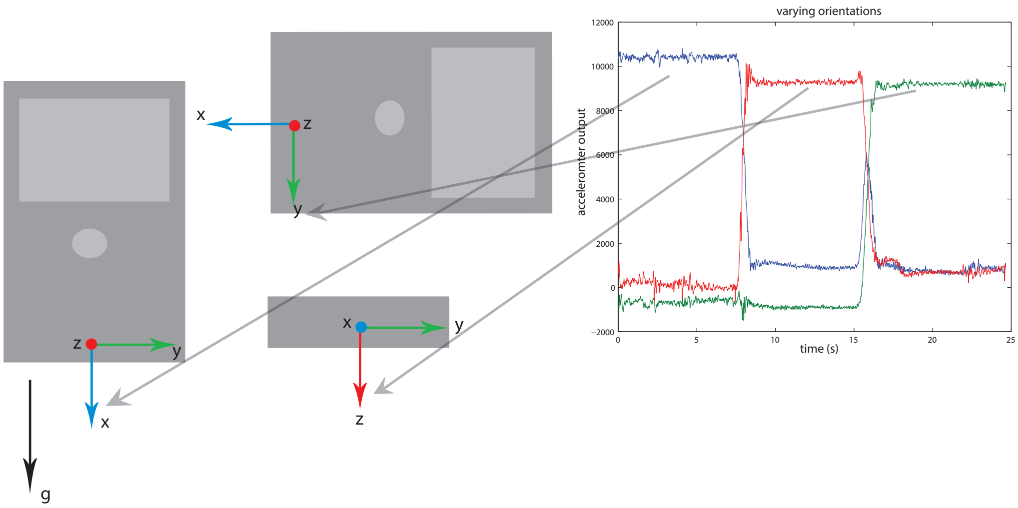


Figure B.6: the device must be held in different orientations for each of the x, y and z accelerometers in order to achieve a value for $\mathbf{g}$.

gyroscope data for 4 different rotations of our device, each with varying speed.

To obtain the average angular velocity, ω_{av}, achieved for this rotation we simply divide 2π by the time taken for the rotation, t_{rot}.

$$\omega_{av_x} = 2\pi/t_{rot} \tag{B.13}$$

We then divide this value by the difference between the max, g_{max} and min, g_{min} values from each rotation to obtain the gyroscope calibration value g_{cal}.

$$g_{cal} = \omega_{av_x}/(g_{max} - g_{min}) \tag{B.14}$$

We may then use this value to obtain an angular rate from our gyroscope data by simply multiplying the value for g_{cal} by the raw value from our sensor. Obviously the value achieved using this approach is an approximate value since we are using raw sensor data which we know is not entirely composed of angular rate information. This method though is acceptable since it may be performed 'on the fly' in any situation making it ideal for the everyday use of a mobile device.

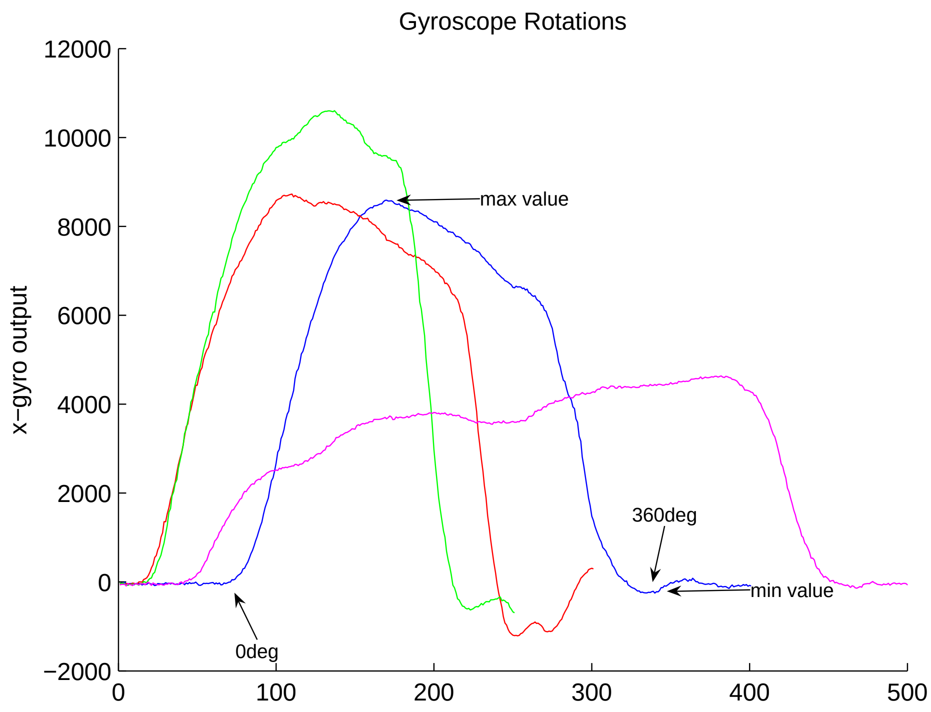


Figure B.7: Gyroscope data for 4 different rotations of our device with varying speeds for each. It can be seen from this data that the rotations were not perfectly smooth.

B.6.4 Magnetometer Calibration

Calibration of our magnetometer data is used to achieve accurate heading determination. Calibration involves the rotation of our device around all three axes in order to determine maximum and minimum magnetometer readings on each axis. This allows us to calculate a value for the 'compass bias', which is a constant vector that the magnetic field of the local environment adds to the measurement. We can also calculate the 'scale factor', which is the apparent magnetic field strength of the Earth.

If we look at figure B.8 we see that the plot of x-axis data against y-axis data for three different positions in the same room produces a circle. In an ideal world these would be perfect circles and would all have the same radius but these circles are not perfect for a number of reasons. One reason is the external magnetic interference mentioned previously and another is that the device was not held completely horizontal as it was rotated around the z-axis in the x-y plane. In terms of figure B.8, we may define the *compass*

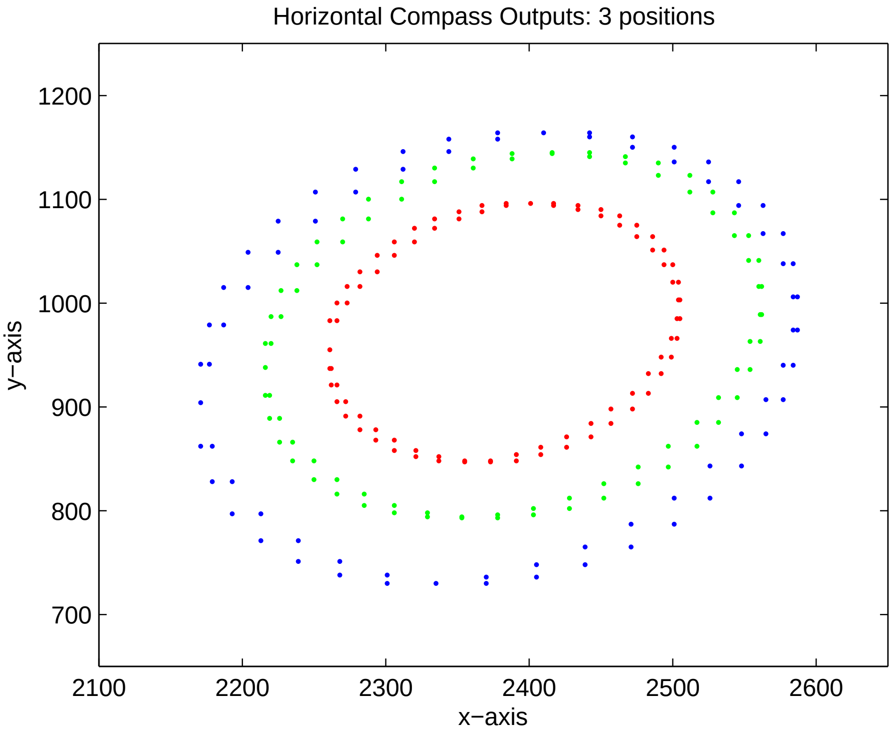


Figure B.8: Plots of x-axis magnetometer data against y-axis magnetometer data for our device rotated in the horizontal plane in three different positions in the same room

bias as a vector pointing to the centre of the circle and the *magnetic scale factor* as the radius of the circle. As with the gyroscopes it is very difficult to obtain perfect results from this type of handheld calibration but they are sufficient. As can be seen from figure B.8 different circles/ellipses with differing scale factors are produced even in the same room. This implies the need for constant recalibration in differing magnetic environments.

Using the min and max values from each axis we can determine separate bias and scale factors for each axis.

$$b_x = (min_x + max_x)/2 \tag{B.15}$$

$$b_y = (min_y + max_y)/2 \tag{B.16}$$

$$b_z = (min_z + max_z)/2 \tag{B.17}$$

$$s_x = (min_x - max_x)/SCALE \tag{B.18}$$

$$s_y = (min_y - max_y)/SCALE \tag{B.19}$$

$$s_z = (min_z - max_z)/SCALE \tag{B.20}$$

where s represents the magnetometer scale factor and b is the compass bias. SCALE is simply a constant value indicating the number of output units per earths magnetic field, usually set at 512. The raw magnetometer data, for the x-axis, may then be calibrated as follows:

$$B_{x,cal} = (B_x - b_x) * (s_x/SCALE) \tag{B.21}$$

where $B_{x,cal}$ is the calibrated magnetometer data for the x-axis, similarly for the y and z axes. This calibrated data is then ready to be used in the heading calculation but first we require to determine the tilt of our device.

B.6.5 Tilt Determination From Accelerometer Data

It is possible to calculate the tilt of our system, pitch θ and roll ϕ, from accelerometer data alone. The gravity vector in our *navigation-frame*, $\mathbf{g}^n$, is related to its *body-frame* coordinates, g^b, by the expression

$$g^b = C_n^b g^n \tag{B.22}$$

It can be shown that eqn(8) may be written as:

$$g^c = \begin{vmatrix} -\sin\theta \\ \cos\theta \sin\phi \\ \cos\theta\cos\phi \end{vmatrix} g$$

dividing both sides of this equation by g and solving for θ and ϕ results in

$$\theta = \arcsin\left(-\frac{g_{x_b}}{g}\right) \tag{B.23}$$

and

$$\phi = \arctan\left(\frac{g_{y_b}}{g_{z_b}}\right) \tag{B.24}$$

equation(10) is not accurate when g_{z_b} is equal to zero, which occurs when $\theta = \pi/2$ or $\theta = -\pi/2$

What we wish to measure is the specific force vector in the *body-frame* frame, f^b, which when our device is at rest is simply a measure of the gravity vector in the *body-frame*.

$$g^b = -f^b \tag{B.25}$$

If this is the case we can calculate the tilt of the device by first using the accelerometer vector output, o_a, in the *accelerometer-frame* to obtain an estimate of the specific force in the *body-frame* from which we can make estimates of the pitch and roll.

If we wish to obtain an estimate of f^b from the accelerometer vector output, o_a, the latter must be converted to m/s^2 and then corrected for axis misalignments and sensor biases according to:

$$\hat{f}^b = \hat{C}_a^b \hat{S}_a^{-1} o_a - \hat{b}_{a,equ}^b \tag{B.26}$$

where $\hat{C}_a^b$ is the coordinate transformation matrix that transforms a vector from the accelerometer frame into the mesh frame, $\hat{S}_a^{-1}$ is the assumed inverse of the accelerometer response matrix and $\hat{b}_{a,equ}^b$ is the assumed equivalent accelerometer bias vector in *body-frame* coordinates.

To reduce the effect of noise in the accelerometer vector output on the estimated specific force vector $\hat{f}^b$, a time average of the accelerometer vector output should be used for the estimation.

A full discussion of accelerometer outputs and various issues associated with this output and the details of tilt determination can be found in (Roth 1999).

B.6.6 Heading Calculation

Determining the compass heading from magnetometer data is a trivial procedure if we may assume that the pitch, θ and roll, ϕ angles are both 0 i.e. that our device is sitting in the horizontal plane. In this case we may simply calculate the compass heading using only the calibrated x and y data from the magnetometers. If it is the case that the device is tilted slightly with respect to the horizontal plane, which in reality is most likely, we first need to transform the magnetometer data back into the horizontal plane, that is the plane perpendicular to the Earth's gravitational vector (Caruso 1999), using the following equations:

$$X_h = X\cos\phi + Y\sin\theta\sin\phi - Z\cos\theta\sin\phi \tag{B.27}$$

$$Y_h = Y\cos\theta + Z\sin\theta \tag{B.28}$$

which take advantage of the pitch and roll angles calculated in the previously and all of the calibrated x, y and z magnetometer data. The final tilt compensated heading may then be calculated as follows:

$$\text{if}(X_h < 0), heading = 180 - \arctan(Y_h/X_h)$$
$$\text{if}(X_h > 0, Y_h < 0), heading = -\arctan(Y_h/X_h)$$
$$\text{if}(X_h > 0, Y_h > 0), heading = 360 - \arctan(Y_h/X_h)$$
$$\text{if}(X_h = 0, Y_h < 0), heading = 90$$
$$\text{if}(X_h = 0, Y_h > 0), heading = 270$$

We may then use the calculated value for the heading as our azimuth or yaw, ψ.

Bibliography

A., Köhne and Wößner M. (2007). GPS - explained. URL: `http://www.kowoma.de/en/gps/`.

Addlesee, M., R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward and A. Hopper (2001). Implementing a sentient computing system. *Computer* **34**(8), 50–56.

Ängeslevä, J., I. Oakley, S. Hughes and S. O'Modhrain (2003*a*). Body Mnemonics. In: *Physical Interaction (PI03) Workshop on Real World User Interfaces, Mobile HCI Conference 2003.*

Ängeslevä, J., I. Oakley, S. Hughes and S. O'Modhrain (2003*b*). Body Mnemonics: Portable device interaction design concept. In: *proceedings of UIST'03, Vancouver, Canada*.

Ashbrook, D. and T. Starner (2003). Using GPS to learn significant locations and predict movement across multiple users. In: *Personal Ubiquitous Computing.*. Vol. 7(5). pp. 275–286.

Athans, M. and D. Kendrick (1973). Control theory and economics: A survey, forecast, and speculations. *Automatic Control, IEEE Transactions on* **19**, 518–524.

Aylward, R. and J. A. Paradiso (2006). Sensemble: a wireless, compact, multi-user sensor system for interactive dance. In: *NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression.* IRCAM Centre Pompidou. Paris, France, France. pp. 134–139.

Bahl, P. and V. N. Padmanabhan (2000). RADAR: An in-building RF-based user location and tracking system.. In: *IEEE INFOCOM (Tel-Aviv, Israel, Mar. 2000)*. pp. 775–784.

Baillie, L., H. Kunczier and H. Anegg (2005). Rolling, rotating and imagining in a virtual mobile world. In: *MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*. ACM Press. New York, NY, USA. pp. 283–286.

Bain, P. (1993). A combined clinical and neurophysiological approach to the study of patients with tremor. *Journal of Neurolology* **56(8)**, 839–844.

Bajaj, R., S. L. Ranaweera and D. P. Agrawal (2002). Gps: location-tracking technology. *Computer* **35**(4), 92–94.

Barkhuus, L. and A. K. Dey (2003). Is context-aware computing taking control away from the user? Three levels of interactivity examined. In: *Proceedings of UbiComp 2003*. Springer. Seattle, Washington. pp. 150–156.

Baskett, B. (2000). Aeronautical design standard performance specification handling qualities requirements for military rotorcraft. Technical report. US Army Aviation and Missile Command Aviation Engineering Directorate, Alabama.

Beale, R. and A. Edwards (1999). Gestures and Neural Networks in Human Computer Interaction.. In: *Proceedings of the IEE Colloquium on Neural Nets in Human Computer Interaction*. IEE Digest. pp. 7/1–7/3.

Bederson, B. B. (1995). Audio augmented reality: a prototype automated tour guide. In: *CHI '95: Conference companion on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 210–211.

Bell, M., M. Chalmers, L. Barkhuus, M. Hall, S. Sherwood, P. Tennent, Barry Brown, Duncan Rowland and Steve Benford (2006). Interweaving mobile games with everyday life. In: *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press. New York, NY, USA. pp. 417–426.

Benford, S., A. Crabtree, M. Flintham, A. Drozd, R. Anastasi, M. Paxton, N. Tandavanitj, M. Adams and J. Row-Farr (2006). Can you see me now?. *ACM Transactions on Computer-Human Interaction* **13**(1), 100–133.

Benford, S., H. Schnädelbach, B. Koleva, R. Anastasi, C. Greenhalgh, T. Rodden, J. Green, A. Ghali, T. Pridmore, B. Gaver, A. Boucher, B. Walker, S. Pennington, A. Schmidt, H. Gellersen and A. Steed (2005). Expected, sensed, and desired: A framework for designing sensing-based interaction. *ACM Trans. Comput.-Hum. Interact.* **12**(1), 3–30.

Beuter, A., A. Glass, C. Mackey and M. Titcombe (2003). *Nonlinear Dynamics in Physiology and Medicine.* second ed.. Springer. New York.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford University Press.

Björk, S., J. Falk and R. Hanssonand P. Ljungstrand (2001). Pirates! - using the physical world as a game board. In: *Interact 2001, IFIP TC.13 Conference on Human-Computer Interaction.*

Blakelock, J. H. (1991). *Automatic Control of Aircraft and Missiles.* John Wiley.

Blankertz, B., G. Dornhege, M. Krauledat, M. Schroder, J. Williamson, R. Murray-Smith and K-R. Muller (2006). The Berlin brain-computer interface presents the novel mental typewriter hex-o-spell. In: *3rd International BCI Workshop and Training Course.*

Blankertz, B., G. Dornhege, M. Krauledat, M. Schroder, J. Williamson, R. Murray-Smith and K-R. Muller (2007). A Note on Brain Actuated Spelling with the Berlin Brain-Computer Interface. In: *HCI International.*

Brewster, S. (1997). Using non-speech sound to overcome information overload. In: *Special Issue on Multimedia Displays.* pp. 179–189.

Brewster, S. A., P. C. Wright and A. D. N. Edwards (1993). An evaluation of earcons for use in auditory human-computer interfaces. In: *CHI '93: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 222–227.

Brewster, S., J. Lumsden, M. Bell, M. Hall and S. Tasker (2003). Multimodal Eyes Free interaction techniques for wearable devices.. In: *Proceedings of ACM CHI 2003 (Fort Lauderdale, FL)*. ACM Press, Addison-Wesley. pp. 463–480.

Britting, K. R. (1971). *Inertial Navigation Systems Analysis*. Wiley-Interscience. New York.

Brooks, R. and S. Iyengar (1998). *Multi-sensor fusion: fundamentals and applications with software*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Brummit, B., B. Meyers, J. Krumm, A. Kern and S. Shafer (2000). Easyliving: Technologies for intelligent environments. In: *Handheld and Ubiqitous Computing, 2nd International Symposium*. pp. 12–29.

Burrell, J. and G. K. Gay (2002). E-graffiti: evaluating real-world use of a context-aware system. *Interacting with Computers* **14**(4), 301–312.

Card, Stuart K., Allen Newell and Thomas P. Moran (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc.. Mahwah, NJ, USA.

Caruso, M. (1999). Applications of magnetic sensors for low cost compass systems. Technical report. Honeywell, SSEC.

Caudell, T. and D. Mizell (1992). Augmented reality: an application of heads-up display technology tomanual manufacturing processes. In: *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*. Vol. 2. pp. 659–669.

Chalmers, M., M. Bell, B. Brown, M. Hall, S. Sherwood and P. Tennent (2005). Gaming on the edge: Using seams in ubicomp games. In: *ACM Advances in Computer Entertainment (ACE'05)*. pp. 306–309.

Chang, A. and C. O'Sullivan (2005). Audio-haptic feedback in mobile phones. In: *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 1264–1267.

Chen, G. and D. Kotz (2000). A survey of context-aware mobile computing research. Technical report. Dartmouth College. Hanover, NH, USA.

Cheverst, K., K. Mitchell and N. Davies (2001). Investigating context-aware information push vs information pull to tourists. In: *Proceedings of Mobile HCI 01*.

Cheverst, K., N. Davies, K. Mitchell, A. Friday and C. Efstratiou (2000). Developing a context-aware electronic tourist guide: some issues and experiences. In: *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 17–24.

Cho, S., J. Koo Oh, W. Bang, W. Chang, E. Choi, Y. Jing, J. Cho and D. Yoon Kim (2004). Magic wand: A hand-drawn gesture input device in 3-d space with inertial sensors. In: *IWFHR '04: Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04)*. IEEE Computer Society. Washington, DC, USA. pp. 106–111.

Cho, S. J., R. Murray-Smith, C. Choi, Y. Sung, K. Lee and Y-B. Kim (2007). Dynamics of tilt-based browsing on mobile devices. In: *Extended abstracts of ACM SIG CHI Conference, San Jose*.

Choi, E., W. Bang, S. Cho, J. Yang, D. Kim and S. Kim (2005). Beatbox music phone: gesture-based interactive mobile phone using a tri-axis accelerometer. In: *Industrial Technology, 2005. ICIT 2005. IEEE International Conference on*. pp. 97– 102.

Cohen, M. and L. F. Ludwig (1991). Multidimensional audio window management. *Int. J. Man-Mach. Stud.* **34**(3), 319–336.

Cook, P. R. and S. Lakatos (2003). Using dsp-based parametric physical synthesis models to study human sound perception. In: *IEEE Workshop* *on Applications of Signal Processing to Audio and Acoustics*. pp. 75–78.

Costanza, E., S. A. Inverso and R. Allen (2005). Toward subtle intimate interfaces for mobile devices using an emg controller. In: *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 481–489.

Counts Media (2007). Yellow Arrow. Project homepage: `http://www.yellowarrow.net`.

Craik, K. J. W. (1947). Theory of the human operator in control systems. *British Journal of Psychology* **38**, 56–61.

Crossan, A. and R. Murray-Smith (2004). Variability in Wrist-Tilt Accelerometer Based Gesture Interfaces. In: *Mobile Human-Computer Interaction MobileHCI 2004: 6th International Symposium* (S. A. Brewster and M. D. Dunlop, Eds.). Vol. 3160 of *Lecture Notes in Computing Science*. Springer-Verlag. Glasgow, Scotland. pp. 144–155.

Crossan, A., R. Murray-Smith, S. Brewster, J. Kelly and B. Musizza (2005). Gait phase effects in mobile interaction. In: *CHI '05: extended abstracts on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 1312–1315.

Degani, A. (2004). The grounding of the royal majesty. In: *Taming HAL: Designing Interfaces Beyond 2001* (A. Degani, Ed.). pp. 100–121. Palgrave Macmillan.

Deuschl, G., P. Bain and M. Brin (1998). Consensus statement of the Movement Disorder Society on tremor. *Movement Disord* **13 (Suppl. 3)**, 2–23.

Drozd, A., S. Benford, N. Tandavanitj, M. Wright and A. Chamberlain (2006). Hitchers: Designing for cellular positioning. In: *UbiComp 2006: Ubiquitous Computing, 8th International Conference*. Vol. 4206. Springer. pp. 279–296.

El-Rabbany, A. (2002). *Introduction to GPS: the Global Positioning System.* Boston: Artech House.

Elliott, G. T. and B. Tomlinson (2006). Personalsoundtrack: context-aware playlists that adapt to user pace. In: *CHI '06: CHI '06 extended abstracts on Human factors in computing systems.* ACM Press. New York, NY, USA. pp. 736–741.

Ellis, W. H. B, A. Burrows and K. F. Jackson (1953). Presentation of air speed while decklanding: comparsion of visual and auditory methods.. In: *Technical Report 841, UK RAF Flying Personnel Research Committee.*

Eslambolchilar, P. (2007). Making Sense of Interaction Using a Model-Based Approach. PhD thesis. Hamilton Institute, NUI Maynooth.

Eslambolchilar, P, A Crossan and R Murray-Smith (2004). Model-based target sonification on mobile devices. In: *Interactive Sonification Workshop, Bielefeld* (T.Hermann and A.Hunt, Eds.).

Eslambolchilar, P. and R. Murray-Smith (2006). Model-Based, Multimodal Interaction in Document Browsing. In: *Invited paper, 3nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms, MLMI.* Washington DC., USA. pp. 1–12.

Espinoza, F., P. Persson, A. Sandin, H. Nyström, E. Cacciatore and M. Bylund (2001). Geonotes: Social and navigational aspects of location-based information systems. In: *UbiComp '01: Proceedings of the 3rd international conference on Ubiquitous Computing.* Springer-Verlag. London, UK. pp. 2–17.

Etter, R. (2005). Melodious walkabout - implicit navigation with contextualized personal audio contents. In: *Adjunct Proceedings of the Third International Conference on Pervasive Computing.* Vol. 191(204).

Fallman, D. (2002). Wear, point and tilt. In: *Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques.* ACM Press. pp. 293–302.

Feiner, S., B. Macintyre and D. Seligmann (1993). Knowledge-based augmented reality. *Commun. ACM* **36**(7), 53–62.

Feiner, S., B. MacIntyre, T. Hollerer and A. Webster (1997). A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In: *Wearable Computers, 1997. Digest of Papers., First International Symposium on*. pp. 74–81.

Fels, S. S. and G. E. Hinton (1990). Building Adaptive Interfaces With Neural Networks: The Glove-Talk pilot study. In: *Human Computer Interaction: Interact '90*. pp. 683–688.

Fishkin, K. P., A. Gujar, B.L. Harrison, T. Moran and R. Want (2000). Embodied user interfaces for really direct manipulation. In: *Communications of the ACM*. Vol. 43, Issue 9. ACM Press. pp. 74–80.

Foody, J., D. Kelly, D. Kumar, T. Fitzgerald, B. Caulfield, T. Ward and C. Markham (2006*a*). A Real Time Motion Capture System, Using USB Based Tri-Axis Magnetic and Inertial Sensors.. In: *Irish Signals and Systems Conference, Dublin, Ireland*. New York, NY, USA.

Foody, J., D. Kelly, D. Kumar, T. Fitzgerald, T. Ward, B. Caulfield and C. Markham (2006*b*). A Prototype Sourceless Kinematic-Feedback Based Video Game for Movement Based Exercise.. In: *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering Revolution In BioMedicine*. New York, NY, USA.

Furness, T. A. (1986). The Super Cockpit and its Human Factors Challenges. In: *Proc. Human Factors Soc*. pp. 48–52.

Gaver, W. (1986). Auditory icons: using sound in computer interfaces. *Human Computer Interaction* **2**, 167–177.

Google Inc. (2007). DodgeBall. `http://www.dodgeball.com`.

Gustafsson, F. and F. Gunnarsson (2005). Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. *Signal Processing Magazine, IEEE* **22**(4), 41–53.

Gwon, Y., R. Jain and T. Kawahara (2004). Robust indoor location estimation of stationary and mobile users. In: *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*. pp. 1032–1043.

Harling, P. A. and A. D. N. Edwards (1997). Hand tension as a gesture segmentation cue. In: *Proceedings of Gesture Workshop on Progress in Gestural Interaction*. Springer-Verlag. London, UK. pp. 75–88.

Harrison, S. and P. Dourish (1996). Re-place-ing space: the roles of place and space in collaborative systems. In: *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*. ACM Press. New York, NY, USA. pp. 67–76.

Hazas, M., C. Kray, H. Gellersen, H. Agbota, G. Kortuem and A. Krohn (2005). A relative positioning system for co-located mobile devices. In: *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM Press. New York, NY, USA. pp. 177–190.

Henrysson, A., M. Billinghurst and M. Ollila (2005). Virtual object manipulation using a mobile phone. In: *ICAT '05: Proceedings of the 2005 international conference on Augmented tele-existence*. ACM Press. New York, NY, USA. pp. 164–171.

Hermersdorf, M., H. Nyholm, J. Perkiö, V. Tuulos, J. Salminen and H. Tirri (2006). Sensing in rich Bluetooth environments. In: *Workshop on World-Sensor-Web, in Proc. SenSys.*

Hightower, J., C. Vakili, G. Borriello and R. Want (2001). Design and calibration of the spoton ad-hoc location sensing system. UW CSE Technical Report: `http://seattle.intel-research.net/people/jhightower//pubs/hightower2001design/hightower2001design.pdf`.

Hinckley, K., J. Pierce, E. Horvitz and M. Sinclair (2005). Foreground and background interaction with sensor-enhanced mobile devices. *ACM Trans. Comput.-Hum. Interact.* **12**(1), 31–52.

Hinckley, K., J. Pierce, M. Sinclair and E. Horvitz (2000). Sensing techniques for mobile interaction. In: *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology*. ACM Press. New York, NY, USA. pp. 91–100.

Hodges, S., L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur and K. Wood (2006). Sensecam: A retrospective memory aid.. In: *Ubicomp*. pp. 177–193.

Holland, S., D. R. Morse and H. Gedenryd (2002). AudioGPS: Spatial audio navigation with a minimal attention interface.. *Personal and Ubiquitous Computing* **6**(4), 253–259.

Hollander, A. (1994). An exploration of virtual auditory shape perception. Master's thesis. University of Washington.

Holmquist, L. E., H. Gellersen, G. Kortuem, A. Schmidt, M. Strohbach, S. Antifakos, F. Michahelles, B. Schiele, M. Beigl and R. Mazé (2004). Building Intelligent Environments with Smart-Its. *IEEE Comput. Graph. Appl.* **24**(1), 56–64.

Holovaty A. (2007). Chicago Crime. Project homepage: `http://www.chicagocrime.org`.

Isard, M. and A. Blake (1998). Condensation - conditional density propagation for visual tracking. In: *Journal of Computer Vision*. Vol. 29(1). pp. 5–28.

Jagacinski, R. J. and J. M. Flach (2003). *Control Theory For Humans - Quantiative Approaches to Modeling Human Performance*. Lawrence Erlbaum Associates. London.

Jones, M. and S. Jones (2006). The music is the message. *interactions* **13**(4), 24–27.

Jones, M., S. Jones, G. Bradley and G. Holmes (2006). Navigation by music: an initial prototype and evaluation.. In: *Proceedings of the International Symposium on Intelligent Environments*. Microsoft Research (ISBN: 1-59971-529-5). pp. 1849–1852.

Jung, Y., P. Persson and J. Blom (2005). Dede: design and evaluation of a context-enhanced mobile messaging system. In: *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 351–360.

Kallio, S., J. Kela, J. Mäntyjärvi and J. Plomp (2006). Visualization of hand gestures for pervasive computing environments. In: *AVI '06: Proceedings of the working conference on Advanced visual interfaces*. ACM Press. New York, NY, USA. pp. 480–483.

Kela, J., P. Korpipaa, J. Mantyjarvi, S. Kallio, G. Savino, L. Jozzo and Di Marca (2006). Accelerometer-based gesture control for a design environment. *Personal Ubiquitous Comput.* **10**(5), 285–299.

Kelley, C. R. (1968). *Manual and Automatic Control: A Theory of Manual Control and Its Applications to Manual and to Automatic Systems*. Academic Press.

Kelly, J. (2004). Video of tilt-menu for mobile phone emulator. `http://www.dcs.gla.ac.uk/~rod/Videos/JamesMenu2.avi`.

Kjeldsen, R. and J. Kender (1996). Toward the use of gesture in traditional user interfaces. In: *FG '96: Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96)*. IEEE Computer Society. Washington, DC, USA. p. 151.

Kontkanen, P., P. Myllymäki, T. Roos, H. Tirri, K. Valtonen and H. Wettig (2004). *Probabilistic methods for location estimation in wireless networks*. Chap. 11. Kluwer Academic Publishers.

Körding, K. P. and D. Wolpert. (2004). Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247.

Kramer, J. and J. Leifer (1988). The talking glove: An expressive and Receptive 'verbal' communications Aid for the Deaf, Deaf-Blind and Non-Vocal. In: *SIGCAPH 39* . pp. 12–15.

Krumm, J. and E. Horvitz (2006). Predestination: Inferring destinations from partial trajectories. In: *Ubicomp 2006: Ubiquitous Computing, 8th International Conference*. Vol. 4206 of *LNCS*. Springer. pp. 243–260.

Laerhoven, K. Van and H. Gellersen (2006). Fair dice: A tilt and motion-aware cube with a conscience.. In: *26th International Conference on Distributed Computing Systems Workshops (ICDCS 2006 Workshops), 4-7 July 2006, Lisboa, Portugal*. IEEE Computer Society. p. 66.

Lantz, V. and R. Murray-Smith (2004). Rhythmic interaction with a mobile device. In: *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*. ACM Press. New York, NY, USA. pp. 97–100.

Linjama, J. and T. Kaaresoja (2004). Novel, minimalist haptic gesture interaction for mobile devices. In: *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*. ACM Press. New York, NY, USA. pp. 457–458.

Lipscomb, J. S. (1991). A Trainable Gesture Recogniser. In: *Pattern Recognition*. Vol. 24(9). pp. 895–907.

Loomis, J. M., R. G. Golledge and R.L. Klatzky (1998). Navigation system for the blind: Auditory display modes and guidance. Presence: Teleoperators and Virtual Environments, 7, 193-203.

Lorensen, W. E., H. Cline, C. Nafis, D. Altobelli and L. Gleason (1993). Enhancing reality in the operating room.. In: *IEEE Visualization*. pp. 410–415.

Ludford, P. J., D. Frankowski, K. Reily, K. Wilms and L. Terveen (2006). Because i carry my cell phone anyway: functional location-based reminder applications. In: *CHI '06: Proceedings of the SIGCHI confer-*

ence on Human Factors in computing systems. ACM Press. New York, NY, USA. pp. 889–898.

Luinge, H. J. (2002). Inertial Sensing Of Human Movement. PhD thesis. University Of Twente.

Lyons, K., M. Gandy and T. Starner (2000). Guided by voices: An audio augmented reality system. In: *International Conference on Auditory Display, Atlanta, GA, April 2000.*

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press. available from http://www.inference.phy.cam.ac.uk/mackay/itila/.

Makino, H., I. Ishii and M. Nakashizuka (1996). Development of navigation system for the blind using GPS and mobile phone combination. In: *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*. New York. pp. 506–507.

Marentakis, G. and S. A. Brewster (2005). A comparison of feedback cues for enhancing pointing efficiency in interaction with spatial audio displays. In: *MobileHCI '05: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*. ACM Press. New York, NY, USA. pp. 55–62.

Marmasse, N. and C. Schmandt (2002). A user-centered location model. *Personal Ubiquitous Comput.* **6**(5-6), 318–321.

McCullough, M. (1998). *Abstracting Craft: The Practiced Digital Hand*. MIT Press.

McCullough, M. (2005). *Digital Ground: Architecture, Pervasive Computing and Environmental Knowing*. MIT Press.

McLeigh, G. (2007). An investigation into the use of gesture interaction to capture information in a scientific laboratory environment. Master's thesis. Dept of Computing Science, University of Glasgow.

Microsoft Research (2007). Wasp. Project homepage: `http://research.microsoft.com/sendev/projects/wasp/`,.

Milnes-Walker, N. (1971). A study of pursuit and compensatory tracking of auditory pitch. *Ergonomics* **14**, 479–486.

Mitchell, W. J (2006). electronic lens. Project homepage: `http://mobile.mit.edu/elens/`.

Mohring, M., C. Lessig and O. Bimber (2004). Video see-through ar on consumer cell-phones. In: *ISMAR '04: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'04)*. IEEE Computer Society. Washington, DC, USA. pp. 252–253.

Mynatt, E. D., M. Back, R. Want and R. Frederick (1997). Audio aura: Light-weight audio augmented reality. In: *ACM Symposium on User Interface Software and Technology*. pp. 211–212.

National Imagery and Mapping Agency (NIMA) (1991). Department of defense world geodetic system 1984, its definition and relationships with local geodetic systems.. Technical Report 8250.2. NIMA.

National Transportation Safety Board (1995). Grounding of the Panamanian passenger ship Royal Majesty on Rose and Crown shoal near Nantucket, Massachusetts. Technical Report NTSB Number MAR-97/01; NTIS Number PB97-916401. National Transportation Safety Board.

Nemirovsky, P. and G. Davenport (1999). Guideshoes: navigation based on musical patterns. In: *CHI '99: CHI '99 extended abstracts on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 266–267.

Oakley, I., J. Ängeslevä, S. Hughes and S. O'Modhrain (2004). Tilt and feel: Scrolling with vibrotactile display. In: *EuroHaptics 2004*. pp. 316–323.

Oliver, N. and F. Flores-Mangas (2006). MPTrain: a mobile, music and physiology-based personal trainer. In: *MobileHCI '06: Proceedings of* *the 8th conference on Human-computer interaction with mobile devices and services*. ACM Press. New York, NY, USA. pp. 21–28.

Partridge, K., S. Chatterjee, V. Sazawal, G. Borriello and R. Want (2002). TiltType: Accelerometer-Supported Text Entry for Very Small Devices. In: *UIST'02: Proceedings of the 15th annual ACM symposium on User interface software and technology*. ACM Press. Paris, France. pp. 201–204.

Patten, J., H. Ishii, J. Hines and G. Pangaro (2001). Sensetable: a wireless object tracking platform for tangible user interfaces. In: *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 253–260.

Patwari, N., J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses and N. S. Correal (2005). Locating the nodes: cooperative localization in wireless sensor networks. *Signal Processing Magazine, IEEE* **22**(4), 54–69.

Perng, J. K., B. Fisher, S. Hollar and K. S. J. Pister (1999). Acceleration Sensing Glove (ASG). In: *ISWC*. pp. 178–180.

Peters, J. W. (2004). *Complete Idiot's Guide to Geocaching*. Alpel Publishing.

Piekarski, W. and B. Thomas (2002). Arquake: the outdoor augmented reality gaming system. *Commun. ACM* **45**(1), 36–38.

Pirhonen, A., S. Brewster and C. Holguin (2002). Gestural and Audio Metaphors as a Means of Control for Mobile Devices. In: *Proceedings of ACM CHI2002 (Minneapolis, MN)*. ACM Press, Addison-Wesley. pp. 291–298.

Pister, K. S. J., J. M. Kahn and B. E. Boser (1999). Smart dust: Wireless networks of millimeter-scale sensor nodes. *Highlight Article in 1999 Electronics Research Laboratory Research Summary.*

Poulton, E. C. (1974). *Tracking skill and manual control*. Academic press.

Powers, W. T (1973). *Behavior: The Control of Perception*. Aldine, Hawthorne, NY.

Preece, J., Y. Rogers and H. Sharp (2002). *Interaction Design: Beyond Human Computer Interaction*. John Willey.

Priyantha, N. B., A. Chakraborty and H. Balakrishnan (2000). The cricket location-support system. In: *MOBICOM*. pp. 32–43.

Priyantha, N. B., A. K. L. Miu, H. Balakrishnan and S. J. Teller (2001). The cricket compass for context-aware mobile applications. In: *Mobile Computing and Networking*. pp. 1–14.

Purcell, P. (1985). Gestural input to interactive systems. *Computer Bulletin* pp. 3–7.

Raento, M., A. Oulasvirta, R. Petit and H. Toivonen (2005). Contextphone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing* **4**(2), 51–59.

Randell, C. and H. Muller (2001). Low cost indoor positioning system. In: *Ubicomp 2001: Ubiquitous Computing* (Gregory D. Abowd, Ed.). Springer-Verlag. pp. 42–48.

Rasmussen, C. E. and C. K. I. Williams (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

Rath, M. and D. Rocchesso (2005). Continuous sonic feedback from a rolling ball. *IEEE MultiMedia* **12**(2), 60–69.

Rekimoto, J. (1996). Tilting operations for small screen interfaces. In: *ACM Symposium on User Interface Software and Technology*. pp. 167–168.

Rekimoto, J. (2001). Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In: *ISWC '01: Proceedings of the 5th IEEE International Symposium on Wearable Computers*. IEEE Computer Society. Washington, DC, USA. p. 21.

Riecke, B. E., J. Schulte-Pelkum, F. Caniard and H. H. Bülthoff (2005). Towards lean and elegant self-motion simulation in virtual reality. In: *VR '05: Proceedings of the 2005 IEEE Conference 2005 on Virtual Reality*. IEEE Computer Society. Washington, DC, USA. pp. 131–138.

Rocchesso, D., R. Bresin and M. Fernstrom (2003). Sounding objects. *Multimedia, IEEE* **10**(2), 42–52.

Roth, F. (1999). Strapdown inertial navigation for ground penetrating radar data acquisition: Theory and experiments. Master's thesis. Colorado School of Mines, USA.

Rubine, D. (1991). Specifying Gestures By Example. In: *Computer Graphics*. Vol. 25.

Rukzio, E., K. Leichtenstern, V. Callaghan, P. Holleis, A. Schmidt and J. Shiaw-Yuan Chin (2006). An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning.. In: *Ubicomp*. pp. 87–104.

Savidis, A., C. Stephanidis, A. Korte, K. Crispien and K. Fellbaum (1996). A generic direct-manipulation 3d-auditory environment for hierarchical navigation in non-visual interaction. In: *Assets '96: Proceedings of the second annual ACM conference on Assistive technologies*. ACM Press. New York, NY, USA. pp. 117–123.

Sayed, A. H., A. Tarighat and N. Khajehnouri (2005). Network-based wireless location: challenges faced in developing techniques for accurate wireless location information. *Signal Processing Magazine, IEEE* **22**(4), 24–40.

Schaal, S., J. Peters, J. Nakanishi and A. Ijspeert (2004). Learning Movement Primitives. In: *International Symposium on Robotics Research (ISRR2003)*. Springer.

Schilit, B., Norman A. and R. Want (1994). Context-aware computing applications. In: *IEEE Workshop on Mobile Computing Systems and Applications*. Santa Cruz, CA, US.

Schmidt, R. A. and T. Lee (2005). *Motor Control and Learning-4th Edition - A Behaviourial Emphasis*. Human Kinetics.

Sheridan, T. B. and W. R. Ferrell (1974). *Man-Machine Systems: Information, Control, and Decision Models of Human Performance*. MIT press.

Smith, L. (1997). The maintenance of uncertainty. In: *International School of Physics 'Enrico Fermi, Course CXXXIII,*. pp. 177–246.

Söderkvist, J. (1994). Micromachined gyroscopes. In: *Sensors and Actuators A*. Vol. 43. pp. 65–71.

Spinellis, D. (2003). Position-annotated photographs: A geotemporal web. *IEEE Pervasive Computing* **2**(2), 72–79.

Starner, T., B. Leibe, B. Singletary and J. Pair (2000). MIND-WARPING: towards creating a compelling collaborative augmented reality game. In: *Intelligent User Interfaces*. pp. 256–259.

Steed, A. (2004). Supporting mobile applications with real-time visualisation of GPS availability.. In: *Mobile HCI*. pp. 373–377.

Strachan, S. and R. Murray-Smith (2004). Muscle tremor as an input mechanism. In: *UIST 2004*.

Strachan, S., P. Eslambolchilar, R. Murray-Smith, S. Hughes and S. O'Modhrain (2005). GpsTunes: controlling navigation via audio feedback. *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services* pp. 275–278.

Strachan, S., R. Murray-Smith, I. Oakley and J. Ängeslevä (September 13-16, 2004). Dynamic Primitives for Gestural Interaction. In: *Mobile Human-Computer Interaction MobileHCI 2004: 6th International Symposium, Glasgow, UK*. Springer-Verlag. pp. 325–330.

Sutherland, I. E. (1968). A head-mounted three dimensional display. In: *Fall Joint Computer Conference, AFIPS Conference Proceedings*. Vol. 33. ACM Press. New York, NY, USA. pp. 757–764.

Thimbleby, H. (1990). *User interface design.* ACM Press. New York, NY, USA.

Titterton, D. H. and J. L. Weston (2004). *Strapdown Inertial Navigation Technology, 2nd Edition.* The Institution of Electrical Engineers.

Trimble Navigation Ltd. (2002). *Lassen Sq GPS Module, Low-power, micro-size GPS solution for mobile products.* Trimble.

Tsukada, K. and M. Yasamura (2002). Ubi-finger: Gesture input device for mobile use. In: *APCHI 2002.* Vol. 1. pp. 388–400.

Tuulari, E. and A. Ylisaukko-oja (2002). Soapbox: A platform for ubiquitous computing research and applications. In: *Pervasive '02: Proceedings of the First International Conference on Pervasive Computing.* Springer-Verlag. London, UK. pp. 125–138.

Čapkun, S., M. Hamdi and J. Hubaux (2001). GPS-free positioning in mobile ad-hoc networks. In: *HICSS '01: Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 9.* IEEE Computer Society. Washington, DC, USA. p. 9008.

Vijaykumar, S. and S. Schaal (2000). Locally Weighted Projection Regression: An O(n) Algorithm for Incremental Real Time Learning in High Dimensional Space. In: *Seventeenth International Conference on Machine Learning (ICML2000), Stanford, California.* pp. 1079–1086.

Wagner, D. and D. Schmalstieg (2005). First steps towards handheld augmented reality. In: *Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on.* pp. 127–135.

Wang, Z. and B. Jezekiel (1996). Conveying visual information with spatial auditory patterns. In: *Speech and Audio Processing, IEEE Transactions on.* Vol. 4. pp. 446–455.

Want, R., A. Hopper, V. Falcao and J. Gibbons (1992). The active badge location system.. *ACM Trans. Inf. Syst.* **10**(1), 91–102.

Ward, D. J., A. F. Blackwell and D. J. C. MacKay (2000). Dasher - a data entry interface using continuous gestures and language models. In: *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology*. ACM Press. New York, NY, USA. pp. 129–137.

Warren, N., M. Jones, S. Jones and D. Bainbridge (2005). Navigation via continuously adapted music. In: *CHI '05: CHI '05 extended abstracts on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 1849–1852.

Wexelblat, A. (1998). Research challenges in Gesture: Open issues and unsolved problems. In: *Gesture and sign language in Human-Computer Interaction*.

Wigdor, D. and R. Balakrishnan (2003). TiltText: Using Tilt for Text Input to Mobile Phones. In: *UIST'03: Proceedings of the 16th annual ACM symposium on User interface software and technology*. ACM Press. Vancouver, Canada. pp. 81–90.

Williamson, J. (2006). Continuous Uncertain Interaction. PhD thesis. Department of Computing Science, University Of Glasgow.

Williamson, J. and R. Murray-Smith (2004). Pointing without a pointer. In: *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 1407–1410.

Williamson, J. and R. Murray-Smith (2005*a*). Dynamics and probabilistic text entry. In: *Hamilton Summer School on Switching and Learning in Feedback systems* (R. Murray-Smith and R. Shorten, Eds.). Vol. 3355 of *Lecture Notes in Computing Science*. Springer-Verlag. pp. 333–342.

Williamson, J. and R. Murray-Smith (2005*b*). Sonification of probabilistic feedback through granular synthesis. *IEEE Multimedia* **12**(4), 45–52.

Williamson, J., R. Murray-Smith and S. Hughes (2007). Shoogle: Multimodal Excitatory Interaction on Mobile Devices. In: *Proceedings of ACM SIG CHI Conference*.

Williamson, J., S. Strachan and R. Murray-Smith (2006). It's a long way to Monte Carlo: probabilistic display in GPS navigation. In: *MobileHCI '06: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*. ACM Press. New York, NY, USA. pp. 89–96.

Zhai, S. and P. Kristensson (2003). Shorthand writing on stylus keyboard. In: *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press. New York, NY, USA. pp. 97–104.

Zimmerman, T. G., J. Lanier, C. Blanchard, S. Bryson and Y. Harvill (1987). A Hand Gesture Interface Device. In: *Human Factors in Computing Systems*.