User Experience, Performance, and Social Acceptability: Usable Multimodal Mobile Interaction

Julie R. Williamson

Doctor of Philosophy
University of Glasgow
January 2012

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Julie R. Williamson*)

Abstract

This thesis explores the social acceptability of multimodal interaction in public places with respect to acceptance, adoption and appropriation. Previous work in multimodal interaction has mainly focused on recognition and detection issues without thoroughly considering the willingness of users to adopt these kinds of interactions in their everyday lives. This thesis presents a novel approach to user experience that is theoretically motivated by phenomenology, practiced with mixed-methods, and analysed based on dramaturgical metaphors.

In order to explore the *acceptance* of multimodal interfaces, this thesis presents three studies that look at users' initial reactions to multimodal interaction techniques: a survey study focusing on gestures, an on-the-street user study, and a follow-up survey study looking at gesture and voice-based interaction. The investigation of multimodal interaction *adoption* is explored through two studies: an *in situ* user study of a performative interface and a focus group study using *experience prototypes*. This thesis explores the *appropriation* of multimodal interaction by demonstrating the complete design process of a multimodal interface using the performative approach to user experience presented in this thesis.

Chapter 3 looks at users' initial reactions to and acceptance of multimodal interactions. The results of the first survey explored location and audience as factors the influence how individuals behave in public places. Participants in the on-the-street study described the desirable visual aspects of the gestures as playful, cool, or embarrassing aspects of interaction and how gestures could be hidden as everyday actions. These results begin to explain *why* users accepted or rejected the gestures from the first survey. The second survey demonstrated that the presence of familiar spectators made interaction significantly more acceptable. This result indicates that performative interaction could be made more acceptable by interfaces that support collaborative or social interaction.

Chapter 4 explores how users place interactions into a usability context for use in real world settings. In the first user study, participants took advantage of the wide variety of possible performances, and created a wide variety of input, from highly performative to hidden actions, based on location. The ability of this interface to support flexible interactions allowed users to demonstrate the the purposed of their actions differently based on the immediately co-located spectators. Participants in the focus group study discussed how they would go about placing multimodal interactions into real world contexts, using three approaches: relationship to the device, personal meaning, and relationship to functionality. These results demonstrate how users view interaction within a usability context and how that might affect social acceptability.

Chapter 5 examines appropriation of multimodal interaction through the completion of an entire design process. The results of an initial survey were used as a baseline of comparison from which to design the following focus group study. Participants in the focus groups had similar motives for accepting multimodal interactions, although the ways in which these were expressed resulted in very different preferences. The desire to use technology in a comfortable and satisfying way meant different things in these different settings. During the 'in the wild' user study, participants adapted performance in order to make interaction acceptable in different contexts. In some cases, performance was hidden in public places or shared with familiar spectators in order to successfully incorporate interaction into public places.

Contributing Publications

Williamson, J. R. *Send Me Bubbles: Multimodal Performance and Social Acceptability*. In Extended Abstracts (Student Research Competition) of ACM CHI 2011 (Vancouver, Canada). ACM Press.

Rico, J., and Brewster, S. *Gesture and Voice Prototyping for Early Evaluations of Social Acceptability in Multimodal Interfaces*. In Proceedings of ICMI 2010 (Beijing, China). ACM Press.

Rico, J., Jacucci, G., Reeves, S., Hansen, L., and Brewster, S. *Designing for Performative Interactions in Public Spaces*. In Proceedings (Workshops) of Ubicomp 2010. ACM Press.

Rico, J. and Brewster, S. *Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability*. In Proceedings of ACM CHI 2010 (Atlanta, GA, USA), ACM Press Addison-Wesley, pp 887-896.

Rico, J. *Evaluating the Social Acceptability of Multimodal Mobile Interactions*. In Extended Abstracts (Doctoral Consortium) of ACM CHI 2010 (Atlanta, USA), ACM Press Addison-Wesley.

Rico, J. and Brewster, S. *Gestures all around us: differences in social acceptability perceptions of gesture based interfaces*. In Proceedings of MobileHCI 2009 (Bonn, DE). ACM Press.

Acknowledgements

I would like to thank my advisor Stephen Brewster, who has always given me excellent guidance in my research and supported me in a wide variety of research endeavours. I am constantly inspired by Steve's enthusiasm and sheer talent in his work. This thesis would not be possible without his supervision and guidance. I would also like the thank my second supervisor Matthew Chalmers for all the interesting and challenging discussions that helped to shape this research.

I would like the thank the thesis committee Steve Benford, Quintin Cutts, and Ron Poet, for all of their time spent reviewing this thesis and their guidance to improve the final submission.

I could not have completed this research with the financial support from the National Science Foundation Graduate Research Fellowship, and generous donations from Nokia and HP Labs.

Many thanks go to Paul Dourish for giving me my first taste for research and helping me to find the topics that are most interesting to me and to the LUCI Lab for lots of good times. I thank my colleagues at HP Labs in Bangalore who welcomed me and guided and supported my research into cross cultural comparisons of social acceptability.

I have also been lucky to have an amazing mentor who encouraged me, gave me confidence, and helped me see the good things about myself. André van der Hoek has been an amazing mentor since my undergraduate years and been a wonderful influence on me throughout my postgraduate life.

I would like to thank the friends and colleagues at the University of Glasgow that make working and living here so amazing. For all my friends in the Department/School of Computing Science, thank you for being so awesome, having such great chat, and commiserating when necessary.

I would like to thank John Williamson, who has given me love and support throughout this process and who willingly read this entire thesis more than once.

I am thankful for my family and their long distance support through the good times and the bad, especially my mother.

In all of this, I must thank the series of miraculous circumstances that brought me to where I am today, from a series of broken cameras that changed my life forever, the fortunate random path that brought me to Informatics at UCI, to the introduction from Paul that brought me to Glasgow and helped me find myself and my happiness.

Contents

List of Figures

Chapter 1

Introduction

1.1 Multimodal Interaction in Everyday Life

Performative technologies, such as public displays, interactive installations, and whole body gestural interfaces, give users the opportunity to express themselves in new ways through technology, immerse their whole body into an interactive experience, and share interaction by performing in social settings. These exciting extensions to user experience have led to the creation of a huge variety of performative interfaces, such as the WaveWindow interface (Perry *et al.*, 2010), crowd-based performative interactions (Sheridan *et al.*, 2011), the *Worlds of Information* large public display (Jacucci *et al.*, 2010), and the Bridge project that allowed users to 'walk on water' (Cross, 2011). However, many of these performative interfaces include a specific 'stage' where performance is bounded by a region defined as the 'performance space.' What happens when performative interactions leave the stage and become integrated into everyday life, occurring naturally and sporadically throughout the day? This thesis focuses on such everyday performances, investigating a series of mobile multimodal interactions in everyday life.

1.2 Motivation

Previous work in multimodal interaction has mainly revolved around issues of recognition and detection, as well as advancements in sensing technologies. These previous works, however, do not fully address the issue that multimodal interfaces require users to *perform* new and possibly embarrassing behaviours in public places. In order to design multimodal interfaces to provide a satisfying and enjoyable experience, interface designers need to consider the social acceptability of their interfaces and the user experience of the interactions they create. However, social acceptability is not simply a static issue of 'acceptable' or 'unacceptable,' but a dynamic decision process that occurs in different social contexts at different stages of experience. This work seeks to better understand this process and effectively incorporate these considerations into the design of multimodal interfaces.

1.3 Thesis Statement

This thesis demonstrates the importance of evaluating user experience and the social acceptability of interacting with multimodal interfaces while mobile or in public settings. Because multimodal interfaces require users to 'perform' as part of interaction, the design of these interfaces should be led by user experience and social acceptability rather than technology and hardware. This thesis presents a theoretically motivated methodology and set of analysis techniques and demonstrates the usefulness of this approach in a variety of settings.

1.4 Contributions

With respect to understanding user experience and social acceptability, this thesis organises a series of user studies into three increasing levels of usage; acceptance, adoption, and appropriation. These studies combine quantitative and qualitative research methods in a novel way in order to address the specific issues of user experience and social acceptability. These studies also make use of a variety of prototypes and deployment systems, demonstrating how interaction techniques can be evaluated at different phases of the development process from early design to longitudinal deployment with a range of implementation costs.

This thesis makes significant contributions to the understanding of performative user experience and its effect on the social acceptability of multimodal interfaces. This is demonstrated through a series of evaluations in laboratory and real world settings. These studies demonstrate some of the factors that influence *how* users make decisions about multimodal interaction at different stages of development and use, such as location and audience, performativity, age, and cultural background.

With respect to theoretical motivations and methodology, this thesis presents a novel performative approach to user experience with a theatrical basis in phenomenology. This approach makes use of mixed methods with results that are analysed using dramaturgical metaphors. These metaphors are based on theatre, incorporating concepts such as stages, performances, and characters to understand how individuals present themselves and behave in public places. The success of this methodology is demonstrated at various levels of sophistication through the course of this thesis.

1.5 A Short Review of Multimodal Interfaces

It is important to note that the term 'multimodal' is widely debated, with a variety of definitions in the literature. If using a mouse means that I can see the mouse, feel it moving across my desk and hear the sound it makes, is this a 'multimodal' interaction? Although this interaction does involve multiple modalities, it has not been purposefully designed to *exploit* those modalities for increased experience or usability. One definition of multimodal interfaces states that "multimodal systems process two or more combined user input modes – such as

speech, pen, touch, manual gestures, gaze, and head and body movements – in a coordinated manner with multimedia system output." (Oviatt, 2003). However, this seems a rather strict definition which does not give the same weight to output as to input. A gesture-based system using vibrotactile and audio output would not be considered a "multimodal" interface by this definition. For the sake of this work, multimodal interfaces will be those interfaces that specifically exploit the capabilities and affordances of more than one modality, either used together or separately as part of one interface. Individually, these modalities, such as gesture, speech, visuals, audio, and vibrations, are still aspects of a multimodal interface even if they are investigated or used individually alongside other modalities.

The design of multimodal interfaces has come a long way since the revolutionary 'Put That There' system was introduced in 1980 (Bolt, 1980). This system utilised a large display, voice commands, and deictic pointing gestures to create one of the first multimodal systems. Using this system, users could make a command such as "create a blue square there" and point to a place on the screen to complete the command. Users could also perform pointing gestures and voice commands to move existing objects around the screen. Names could be given to objects on the screen simply by saying "call that an appointment" and pointing to an object on the screen. Many multimodal systems have focused on speech and simple 2D gestures such as pointing gestures or screen based gestures (Oviatt, 2003). The CUBRICON system (Neal *et al.*, 1989) also used speech input, but used pointing gestures with the mouse rather than the hand. The CUBRICON system focused on providing multimodal feedback using a variety of techniques such as visualisations on a colour graphics display, a monochrome display, and speech output by the system. These multimodal techniques were used to allow users to navigate through a large information base of tactical Air Force mission information.

The Iconic system began to move away from simple deictic gestures towards more complex depictive ones (Koons and Sparrell, 1994). This system used data gloves to collect information continuously about hand shape, orientation, movement and location while the user spoke. The speech is analysed using a speech recognition system while the gestures were analysed and segmented. Together, speech and gesture are interpreted and represented visually for the user. This system was used similarly to the 'Put That There' system described above, in that it allowed the user to describe and manipulate objects through speech and gesture. The Iconic system, however, allowed the user to describe objects in much more detail using gestures beyond the simple deictic gestures. For example, users could command the system to "tilt the floor like this" and demonstrate the amount of tilt using a two handed gesture. These kinds of gestures incorporate more conversational interactions than previous systems.

As multimodal systems become more advanced, the realistic possibility of creating multimodal mobile systems brings forward a number of new issues for interaction design. Lemmelä *et al.* (2008) propose a specific set of design principles that must be examined when designing multimodal systems for a mobile context. This approach involves identifying usage limitations of different contexts, creating usage scenarios for multimodal techniques to create an application design, creating modality-specific designs, rapid prototyping, and evaluation of the prototype in naturalistic situations. These five design principles stress how important context is to a user's ability and willingness to interact with a mobile multimodal interface. Oviatt stresses that multimodal interfaces should be designed to support fluid changes in the cogni-

tive load on users during multitasking or complicated interactions (Oviatt *et al.*, 2004). Oviatt demonstrates that individuals can utilise multimodal interaction techniques when tasks are difficult or multitasking if required. Willingness to interact in public and cognitive load are important factors to consider when designing mobile multimodal interfaces.

1.5.1 Gesture at the Interface

Throughout this thesis, gesture is often the focus of the multimodal interactions evaluated. However, the lack of a clear or unified definition of gesture has been cited as a challenge for research into gesture-based interfaces (Wexelblat, 1998). Kendon gives a general definition of gestures as voluntary and expressive movements of the body (Kendon, 1997). This definition succeeds at capturing the common understanding of 'gesture' but does not describe 'gesture' completely. Specifically, this definition fails to address how context, perception, meaning, and relationship to speech affect how an action is commonly identified as a gesture. Kendon does discuss the role that perception plays in determining whether a given action is a gesture (Kendon, 1986) even though this isn't included in this definition. Kendon states that individuals are able to understand gestures and identify expressive intent simply by watching how the action is performed. This results in the circular definition that a gesture is simply an action that is considered a gesture by others. In order to narrow down the definition of gesture, Kendon's work only considers those gestures that are used along with speech and are perceived by observers as part of the meaning of the speech. Accidental gestures and fidgeting are not included in his analysis. This difference between the general definition of gestures and the applied definition gesture can be seen in many of the following examples.

Cassell defines gestures as hand movements occurring during speech (Cassell, 1998), although she criticises this limited definition. Multimodal systems often use gestures outside of the context of speech, but Cassell states that many systems using gesture focus on 'gesture languages' rather than gestures that would ever naturally occur with speech. Cassell argues that the communicative elements of gestures that occur with speech are important when creating 'natural' user interfaces and are lost with the use of 'gesture languages.' Väänänen and Böhm define gestures as "body movements which are used to convey some kind of information from one person to another" (Väänänen and Böhm, 1993). This definition is then further refined, given that the information conveyed by gestures is easily understood by observers but is vague and implicit in nature. For example, an individual might use a gesture alongside speech when saying "I only want a little bit of milk in my coffee." A gesture may reinforce that a small amount of milk is desired, but will not communicated exactly how much milk is considered 'a little bit' by the speaker. For the purpose of their system, Väänänen and Böhm highlight that gestures used in human computer interfaces must have explicitly defined meanings, which is in direct opposition with how gesture are used in daily life. In order to accommodate the technical restrictions of a gesture-based interface, a limited definition of gestures was used so that gestures mapped directly to specific meanings or functionalities. For Väänänen and Böhm's final system, a set of hand positions were implemented, which included finger pointing, an open palm, and a fist. While these are easy to perform, they are necessarily arbitrary movements and fall short of the original definition that revolved around implicit information exchange.

For the gestures used throughout this thesis, the term 'gesture' will be used as it was in Väänänen and Böhm's work described above. Even though the design and selection of gestures will revolve around communication, this thesis accepts that gestures must necessarily map to more explicitly defined meanings when used in an interface as opposed to everyday life. Additionally, gesture classifications from the literature will be used throughout this thesis that organise gestures into more specific groups that have been adapted for applicability to multimodal interaction. Although the scope and specificity of the classification schemes available in the literature is highly variable, this thesis incorporates classifications and categories based on context of use (Kendon, 1986), meaning with respect to speech (McNeill, 1996), and how meaning is learned (Nespoulous *et al.*, 1986).

1.5.2 Performative Interactions

There is a strong body of previous work that builds on sociological and anthropological foundations to explore the issues of performance and experience as part of an interface. This previous work varies in its theoretical foundations, the primary focus of inquiry, and the methodologies used when evaluating performative interactions in different settings.

Understanding interactive technologies as a performance can be understood from the perspective of traditional theatre and performance studies. Jacucci *et al.* (2005) discuss a series of case studies that are analysed based on anthropological studies of experience and theatre as well as ethnographic studies of performance. Jacucci *et al.* (2005) present a manifesto on how concepts and practices from theatre and performance can be practically applied to the design process of interactive technologies by changing the way these technologies are viewed by designers. Jacucci *et al.* (2005) state that one of the implications of this stance with respect to traditional human-computer interaction research means "configuring and staging space instead of measuring and simulating." By reframing interaction as performance, evaluation moves away from studies of usability towards the creation and design of interactive experiences.

Dalsgaard and Hansen (2008) also build on performance theory to develop the concept of "performing perception," describing in great detail the experience of performing with respect to the roles users must adopt throughout an interactive experience. Dalsgaard and Hansen's work also incorporates concepts from Goffman to understand how performative interactions relate to focused or unfocused social interactions, where focused interactions are those in which all the participants are expected to have a single focus of attention and unfocused interactions are those in which there is no single focus of attention but where all of the participants are aware of each other's presence and act accordingly (Goffman, 1966). The concepts of focused and unfocused interactions change the way performative actions should be viewed and analysed given the changing relationships between spectators, performers, and the technology itself in those different settings.

Other work has looked at performance not from the user's or performer's perspective, but from the viewpoint of the spectator. Reeves *et al.* (2005) describe how the spectator experience can be characterised based on the visibility of the performer's manipulations and the resulting effects. Performances can then be described as magical, expressive, secretive, or suspenseful

from the perspective of the spectator. Reeves *et al.* (2005) also goes on to discuss additional facets of the spectator experience such as the changing dynamics of spectating when interaction is possible and considerations for the transitions between spectating and performing. Montero *et al.* (2010) built on this work by looking at how spectators determined the social acceptability of gesture-based interactions based on the visibility of performers' manipulations and the resulting effects. This research discussed the importance of visible effects as a way that spectators make sense of actions.

There is also a wide variety of research that focuses on both spectators and performers through qualitative or ethnographic evaluations. Sheridan *et al.* (2007) describe how performance frames can be used to understand participation in digital live art performances. This research revolves around the concept of "wittingness," where individuals' awareness of the performance frame influences how they participate in a performative experience. Gardair *et al.* (2011) looked at how performative spaces are defined for street performers, examining how passers-by transition into audience members in this setting. (Benford *et al.*, 2006) also discuss the fluid relationship between audience members, witting or otherwise, and bystanders through an evaluation of a mobile performative game. This project sought specifically to blur the boundaries of digital and physical aspects of the game to encourage performative activities in public places and implicate passersby as unwitting participants.

Placed within this divergent area of work in performative interaction, this thesis focuses on a complementary line of enquiry that builds on several aspects of the previous works while exploring performativity through a novel methodology. Based on the theoretical foundations of this thesis, as described in Chapter 2, this research is clearly focused on performers and their individual perceptions of performance. However, this forms a natural complement to the spectator focused research described above, where both explore the same phenomenon but from different perspectives. Research from both of these perspectives contributes equally to the evaluation and design of performative interfaces since users often switch fluidly from performer to spectator and back throughout an interaction (Leong *et al.*, 2011).

With respect to methodology, this thesis moves away from much of the previous work that often involved case studies, ethnographic methods, or analysis from the perspective of ethnomethodology. The methodology described in Chapter 2 and used throughout this thesis involves a mixed-methods approach where analysis is based on both qualitative and quantitative data. In many cases, traditional ethnographic techniques would be inappropriate or impractical given the highly mobile and sporadic nature of the interactions of interest to this thesis. In this case, observation of sparse interactions is often impractical. Additionally, this approach is based on Goffman's dramaturgical metaphors (Goffman, 1990) as opposed to the performance theory or frame analysis commonly used in the previous work. These metaphors provide a compelling way to analyse and organise the repeated observation, measurement, or self-reported description of the everyday sporadic interaction explored in this thesis. This is in contrast to the fine grained analysis on a smaller subset of denser interactions that are observed or recorded as seen in frame analysis. For those interactions where the data is dense and rich in quality, frame analysis or conversation analysis may be more appropriate in order to fully exploit the data. These two approaches are not in opposition, but simply present two different methods for analysing different kinds of interactive experiences and the data they

produce.

1.6 Using Multimodal Interfaces: Acceptance, Adoption, and Appropriation

This thesis presents studies in an increasing order of depth with respect to the experience of the user. This organisation is also based on how much control the experimenter has over the experience and whether or not the interactions are placed into a usability context, that is attached to a specific functionality or application scenario.

Acceptance – This phase investigates the most basic and early perceptions of multimodal interactions before those interactions are put into a usability context. The initial perceptions and reactions to a given interaction technique represent the first step in the decision process for users when considering the acceptance of a new technology.

Adoption – This phase evaluates interaction techniques that have been placed into a usability context but are still evaluated using prototypes or relatively controlled user studies.

Appropriation – This phase investigates interaction techniques by evaluating multimodal interaction 'in the wild.' Here, the fully implemented interaction techniques are placed into a usability context and evaluated in real world settings.

These three phases of usage are based roughly on the technology acceptance model (TAM) as presented by Davis (1989). Interface acceptance is similar to 'perceived usefulness' and 'perceived ease of use.' This thesis argues that the consideration of 'social acceptability' is also important in this initial phase. This thesis explores acceptance specifically outside of a usability context to limit the bias that functionality or recognition accuracy might impose on initial results. For example, if a new multimodal interaction technique is presented within the context of a mobile email browser, users' pre-existing opinions about mobile email are likely to influence their opinions about the interaction technique even though issues relating to data connectivity and email usability are unrelated to the new interaction technique. In order to minimise this kind of influence, acceptance is explored before interactions are placed into a usability context. The next phase, interface adoption, is similar to the TAM's 'behavioural intention to use' with some notable differences. Firstly, this is the first time when the usability context is considered. Rather than simply 'making a plan' for usage, this second phase of adoption is based on imagining or actually using a technology with the additional information about context and functionality to inform decisions. Secondly, this thesis also argues that some level of experience is useful before decisions about adoption are be made so that the 'behavioural intention to use' is based on some amount of experience. Interface appropriation is similar to TAM's 'actual system use' where users are actively using the system in their own daily lives where this thesis places a specific emphasis on extended use in real world settings.

1.7 Challenges in Social Acceptability Research

When investigating the social acceptability of any new interaction technique, the inevitable question arises: what about greater social change? The investigation of possible new interaction techniques can appear to be heavily affected by greater social change. Are the preferences that users describe going to change if 'everyone else' accepts this interaction technique? What if this interaction technique is supported by popular media and becomes acceptable? A common example of a greater social change in opinion can be seen in the adoption of Bluetooth headsets, which led to the relatively new habit of seemingly 'talking to yourself' in public. If those interactions were once unacceptable, how did they become acceptable and how could that have been predicted?

This thesis addresses this issue in three ways. Firstly, the focus of this work is on *how* users go about incorporating new interaction techniques into their daily lives, not *what* users choose to accept. The goal of this work is not to identify a list of 'acceptable' interaction techniques, but rather to understand the aspects of these techniques that users most often consider and the process by which users decide to accept them. Secondly, this thesis works within the scope of subjective user experience, which is inherently an individual and personal investigation. Because the theoretical foundation of this work is strongly based in phenomenology, the main focus is on individual subjective experiences, where larger social influences are only addressed from the perspective of the individual.

Finally, this thesis does not try to simply suggest that greater social change is out of the scope of this work, but accepts the duality of structure as described by (Giddens, 1986). By investigating a topic that involves decisions in social settings, this work influences greater social change while also being influenced by it. Giddens's Structuration theory describes in great detail the recursive structure of society, attempting to remove the emphasis on humanist and positivist sociological traditions by removing focus from individual experience and larger social structures, but focusing on the way in which these are both reproduced by *social practices*. Giddens states that "the basic domain of study of the social sciences, according to the theory of structuration, is neither the experience of the individual actor, not the existence of any form of societal totality, but the social practices ordered across space and time. Human social activities, like some self-reproducing items in nature, are recursive." (Giddens, 1986). Lewis presents a perspective on how social practices become a part of widespread *convention*. (Lewis, 2002). Lewis describes how problems with "coordination" are often solved using conventions, which come about based on explicit agreements, social contracts, and accepted norms and standards. With respect to multimodal interfaces, Lewis's work helps to explain how individuals negotiate the acceptability of new interactions but does not address where those interactions might come from.

This thesis attempts to incorporate these perspectives on greater social change while remaining true to its subjective roots. This work subscribes to the humanist tradition of sociology, and values individual and subjective experience as the primary source for investigation. This thesis incorporates some of the concepts of Structuration theory by going beyond simply identifying, describing and understanding experience and looking at those social processes that drive this experience.

1.8 Thesis Structure

The thesis is written such that each chapter builds on the previous ones, but each chapter may also be read independently. The thesis begins with the description of the methodology used throughout this thesis, followed by three chapters dedicated to acceptance, adoption, and appropriation, and finishing with a concluding chapter. The appendices include all of the materials used through the course of this thesis.

The second chapter presents in detail the theoretical approach used, the rationale behind the methodology, and the techniques for analysis used in this thesis. The goals of this section are to present a sound theoretical motivation for this methodology of user experience research and to expose and make accessible the analysis techniques.

The third chapter begins by introducing in more detail the 'acceptance' phase of multimodal interface use and presents three studies that examine user reactions and acceptance of multimodal interfaces. These studies include a survey on the acceptability of gesture based interaction, an on-the-street study, and a survey on the acceptability of gesture and voice based interaction. Large portions of this chapter are based on work published in CHI 2010 (Rico and Brewster, 2010b).

The fourth chapter goes on to present in more detail the 'adoption' phase of use. This chapter presents two studies that explore two areas of adoption. The first study looks at how users develop gesture and voice interactions *in situ* in order to make performance both acceptable and enjoyable. This section is based on work published in CHI 2011 Extended Abstracts (Williamson, 2011). The second study, a series of focus groups, looks at how users evaluate interaction techniques and place them into a usability context. This section is based on work published in ICMI 2010 (Rico and Brewster, 2010a).

The fifth chapter completes the discussion of multimodal interface usage by looking at the 'appropriation' phase of use. This chapter showcases the methodology proposed in this thesis from start to finish, ending with a longitudinal study completed in different cultural settings. This chapter discusses an initial survey of existing user practices, the results of a focus group looking at preferences and usability scenarios, and concludes with a longitudinal user study that examines appropriation and usage in everyday life.

The final chapter summarises this thesis, discussing the most important contributions of this work and reflecting on the success of the methodology. The conclusions also give a final analysis of the challenges addressed in this thesis and directions for future work.

Chapter 2

User Experience: A Performative Perspective on Multimodal Interaction

2.1 Introduction

The evaluation of user experience has become an increasingly important aspect of usability and human computer interaction research. However, as a relatively new area of research it has lacked a clear and unified theory, methodology and understanding (Law *et al.*, 2008). Given the multidisciplinary nature of user experience research, which incorporates designers, psychologists, sociologists, computing scientists, and researchers from the mixed background of human computer interaction and beyond, it is not surprising that the community is having trouble finding common ground. The diversity of definitions and approaches has led to a rather fragmented community lacking both common understandings and comparability of results.

There have been a variety of activities and special interest groups in the research community tackling these disparities by discussing appropriate methods for user experience (Obrist *et al.*, 2009), developing shared understandings or definitions of user experience, (Law *et al.*, 2008) and discussing the theory, if any, behind user experience, (Obrist *et al.*, 2011). While some important overarching aspects of user experience have been identified (Law *et al.*, 2009) there is still a lack of clear direction in the community as a whole.

Within the user experience research community, there are three main groups: those focusing on the user, the interaction, and the product [1]. The approach presented in this thesis uses performance to combine these themes through subjective individual experiences by incorporating interactions as performances and products as part of the visible aspects of the performance

[1] These themes have been discussed at the Special Interest Group on UX Theories (Obrist *et al.*, 2011) although they do not appear in the SIG description.

itself. This performative approach to user experience stems from the *embodied interaction* tradition, as described by Dourish (2004). However, where Dourish's embodied interaction builds on the combination of social computing and tangible computing, this performative perspective on user experience builds on social computing and whole body interaction. This slightly different foundation leads to less of a focus on *embodiment*, and more on *performance*. This approach revolves around the evaluation of *performative interaction*, which is considered by this approach to be any interaction that requires action in public places.

This chapter presents the theory, methodology, and analysis techniques of this novel performative approach to user experience. In combining social computing and whole body interaction, this approach to user experience is based on phenomenology, practiced using mixed methods from sociology and computing science, and analysed using dramaturgical metaphors. Although this approach focuses on 'performativity,' the key concepts of user experience, performance, and social acceptability presented here are not limited to 'performative interaction' but have a more general applicability. For example, interfaces such as social networks that require 'gestures' as in symbolic gestures of kindness, inclusion, or friendship might not be considered 'performative' but can effectively be evaluated through the same lens.

2.2 User Experience, Performance, and Social Acceptability

In order to clearly present this novel approach to user experience, some key concepts must first be described. Although user experience plays a significant role in this work, user experience is just one of three concepts used in this approach alongside performance and social acceptability. Performance will be used throughout this thesis to mean the 'performance' of everyday life in a much broader sense than the traditional understanding of performance as in a 'theatrical' performance. Social acceptability will be taken as the concept that encompasses user experience, performance, and decision making.

2.2.1 User Experience

User experience is "a person's perceptions and responses that result from the use or anticipated use of a product, system, or service" as defined by ISO 9241-210:2010 (9241-210:2010, 2010). However, successfully evaluating user experience requires a much greater understanding of such perceptions and responses than this simple definition might suggest. For example, how should a researcher go about gathering a person's perceptions and responses? How do users generate such perceptions and responses, and which of these perceptions are important and interesting for researchers? These are the questions that divide the user experience community most clearly, where different communities have very different ideas about how to answer these questions.

In general, the majority of the community focuses on a user centred approach (Obrist *et al.*, 2011). This may not be surprising given that user experience should necessarily involve the user, but there are those who would focus on interaction or products as the key facet of user

experience research. It can be argued that a thorough approach to user experience should incorporate all three of these factors. Although this approach primarily centres on users and their subjective experiences, interactions and products are also incorporated. Interactions play a part in how users develop perceptions through feedback from and interactions with others. Products play an important role in user experience because a device or interface is often the focus of interaction. The product is also an important visible aspect of users' presentation of self and influences how they use the product while in public and how they interpret the reactions of spectators. Thus, users, interactions, and products each have an important role in user experience.

There are some important aspects of user experience that are more widely agreed upon. Firstly, user experience is fluid and dynamic since the personal moods, social context, and past experiences of users are ever changing. These clearly have a significant effect on the current experience (Vermeeren *et al.*, 2010). In any given experience, individual users have very different perceptions that are influenced by a wide variety of factors both internal, such as mood, and external, such as social context. The dynamic changes in user experience over time vary from very short term changes, such as a change in physical context, to long term changes brought about and developed by repeated experiences within a changing social group or context. Secondly, user experience is inherently an individual and personal experience, albeit at times heavily influenced by spectators and social groups (Law *et al.*, 2009). Because an individual user is the one creating the perceptions and responses that user experience is so interested in, this is where the experience exists. In order to better understand how these perceptions are developed, we can look at presentation of self and experience as a *performance.*

2.2.2 Performance

The idea of interaction as a performance (Jacucci, 2004) provides a way of understanding interaction as the presentation of self and the experience of interacting in front of others. Previous works in the design and evaluation of *performative* interfaces have looked at a variety of performance contexts, such as musical performances (Benford, 2010) and digital art performances (Sheridan *et al.*, 2007). With respect to multimodal interaction, this performative perspective seems especially appropriate because multimodal interfaces often require users to adopt new behaviours that can be highly visible or audible in public spaces. The work of Goffman discusses in great detail the use of dramaturgical metaphors in understanding behaviour in social and public settings (Goffman, 1990). An important aspect of this work is the definition of everyday performance and how this affects our understanding of action in public places.

The definition of performance that will be used throughout this thesis is that of 'everyday performance,' where nearly every action of everyday life is considered a kind of performance. As we present ourselves to others in daily life, our actions are like performances of our desired 'character,' or the impression we hope to create, and we are constantly adjusting that performance as we gather feedback from others. Goffman describes a wide range of performances, from implicit performances of everyday action and impression management to explicit performances such as giving a formal presentation to an audience. These concepts can be further refined as impressions or performances *given* and *given off*, where impressions *given* relate to

those explicit performances and impressions *given off* relate to implicit performances (Goffman, 1990). Implicit performances might be actions that are performed without being explicitly aware of them, but which are unconsciously adjusted constantly throughout the day as feedback is gathered from spectators. For example, when running late one might constantly look at the time and fidget while waiting for a train. These actions might be unconsciously done, but they still communicate a state of agitation and are still considered a performance under this definition. More explicit performances carry with them significantly more intention from the performer and more clearly defined performer/spectator roles. For example, giving a presentation is an explicit performance where the presenter stands at the front of the room and talks and the audience gives their attention to the presenter. The roles are predefined and, hopefully, already recognised and understood by everyone present.

This thesis uses of the term performance to encompass a wide variety of activities, many of which might not be considered 'performance' in the traditional sense. This approach considers both intentional and unintentional performances as part of a spectrum from simple performance to *performativity*. If we consider every action in public to be a performance of some kind, then those actions that are carried out with intention and awareness of spectators can be described as *performative*. Indeed, much of the focus of this work will look at performative interactions with respect to how these actions are hidden or exaggerated, how performativity affects experience and decisions about usage, and how performance is achieved by users in everyday settings.

2.2.3 Social Acceptability

The motivation for looking at user experience is to determine how users might accept multimodal interactions into their everyday lives. Because multimodal interfaces require users to adopt new and often strange behaviours, understanding the willingness of users to adopt these behaviours in their everyday lives is an important consideration. The term *social acceptability* encapsulates these concerns as the social factors which affect user experience and the acceptance of new interaction techniques.

Social acceptability is a user's continuous decision process that is influenced by the experiences gathered while performing. Social acceptability is not a simple matter of acceptable or unacceptable, but is a continuous decision process that develops over time. Users perform a given action, collect feedback from the present or imagined spectators, and adjust the performance as appropriate. The experience that is remembered is composed of the actual personal and physical experiences and the higher level interpretation of the social experience. These memories influence individual decisions on an individual scale. These in turn influence and are influenced by larger scale social changes. This duality of structure, as described by Giddens (1986), represents an important issue in any examination of society and behaviour. This balance between individual and social change is a theme that comes up repeatedly in this thesis. The evaluations in this thesis focus on *individual* change while accepting that this is both a limitation of the research and a necessity when focusing on subjective experience.

There are a few examples of previous work looking at social acceptability, although these stud-

ies are not centred on user experience and performance and lead to very different discussions. Previous studies have looked at how interactions are influenced by and can be designed for spectators (Reeves *et al.*, 2005). Montero et al. ran a focus group study that looked at how the social acceptability of gestures is evaluated based on the relationship between the visibility of actions and their effects (Montero *et al.*, 2010). These visible aspects of gesture-based interaction play a significant role in how spectators evaluate the social acceptability of these interactions. For example, if the gesture is highly visible and the effects are not, spectators have difficulty understanding the meaning or motivation behind a gesture. However, if both gesture and effect are visible, spectators can associate the gesture with the effect to make sense of the performance. Another previous study, focusing on interaction, was completed by Ronkainen *et al.* (2007). This study used video prototypes as part of a survey to examine user willingness to perform the given gestures within a specific context for a specific task. For each video, possible responses about willingness to use the gesture included: Yes, it's fun / Yes, it's useful / Yes (other reason) / No, it looks silly / No, it's not useful / No (other reason) (Ronkainen *et al.*, 2007). This survey focused on interaction, where users' responses might be based as much on how the gesture fits with the interaction as on how comfortable they would feel performing that gesture in public. For example, if a user selects the response "No, it's not useful," do they mean the gesture is a poor fit for the functionality, or do they simply not use the functionality shown in the video? However, this focus on interaction was useful for successfully ruling out those gestures that were unacceptable early in the design process. This survey proved to be a useful design tool, but did not address user experience with significant depth.

These examples of previous work show different perspectives on what is important when evaluating for social acceptability. However, the work of this thesis is clearly user centred and focuses on user experience as a subjective individual experience. This leads to work that centres on individual perceptions of personal experiences, rather than the descriptions or perceived meanings of others' actions. The following sections will build on the concepts described here to present a novel approach to user experience research. This approach begins with the theoretical foundations of this user experience research, the resulting mixed methods methodology that combines qualitative and quantitative methods, and the analysis techniques based on dramaturgical metaphors.

2.3 A Novel Approach to User Experience

This approach to user experience begins with a theoretical foundation in phenomenology. Similar to embodied interaction, a performative perspective on user experience values experience itself as a subjective phenomenon that is constantly interpreted by the experiencer. The methods of this approach are essentially mixed, with methods from humanistic tradition of sociology and computing science combined to collect a variety of qualitative and quantitative data about experience. The analysis of this data is based on dramaturgical metaphors, where actions, decisions, and perceptions are organised using themes from traditional theatre that can be applied to everyday life.

2.3.1 Theoretical Foundations: Phenomenology and Experience

From the perspective of phenomenology, social life exists within the subjective experience of individuals where an 'objective' world or experience is of little importance or interest and arguably may not exist as an object that can or should be studied. Phenomenology stems from the humanistic branch of sociological theory that also includes topics such as ethnomethodology and symbolic interactionism. Two key assumptions separate these humanistic sociologies from traditional or positivist sociological theories (Benson and Hughes, 1983). First, that humans as social beings are not simply controlled by external social forces and factors but are constantly interacting and actively creating their own social reality. Second, that specialised methods are required to access the experiences, perceptions, and intentions that are central to humanistic sociological inquiries (Benson and Hughes, 1983). Within the field of sociology, this meant a shift from the traditional methods of looking at records, reports, and statistics to the development of participant observation, refined use of the interview technique, and the adoption of ethnographic methods. With respect to computing science, adopting this theoretical stance suggests a move away from traditional lab based studies of interface accuracy and speed to studies that examine experience, usability, and perceptions as they are developed in the real world.

Simply described by Luckmann (1978), "the goal of phenomenology is to *describe* the *universal* structures of the *subjective* orientation in the world, not to explain the *general* features of the *objective* world." The subjective orientation of the world refers to the subjective experience of individuals going through their daily lives. For Schütz (1967), the experience of everyday life was the natural focus of inquiry and the source of information for understanding social worlds, called the *lifeworld* (Benson and Hughes, 1983). The lifeworld is simply the world where everyday life is taken as objective and the reality of that experience is taken for granted. In the lifeworld, individuals take on the natural attitude (Lewis and Staehler, 2011), where one can assume that the world exists outside of our perception and assume that world is objective. In phenomenological studies, the objective study of subjective experience relies heavily on the phenomenological *epoché* (Lewis and Staehler, 2011), or the ability to removes one's self from the natural attitude and into the reflective and analytic stance that strips away, as much as possible, one's own cultural, temporal, and social bias. This attitude is one of reflection, questioning, and interpretation where experiences must be analysed critically through *description*, *reduction*, and *interpretation* (Lanigan, 1988). The phenomenological epoché allows researchers to look beyond experience as it presents itself in the lifeworld by questioning the lifeworld itself, exploring the natural attitude, and understanding how individuals makes sense of experience in order to go about their everyday lives.

This thesis builds a methodology of user experience based on phenomenological theory. There are two key aspects of this theory that influence the practice of this methodology. Firstly, that individual memories and perceptions are of key importance to understanding social experiences. Gaining access to such personal thoughts presents difficulties, which will be discussed in detail in the next section. Secondly, that analysis will require some suspension of the natural attitude to understand where perceptions come from and why such beliefs are held. However, the application of these key concepts will be limited by the practical challenges of completing research in social situations. As Attig (1985) states, a philosopher is "ever aware of the need

for, yet sensitive to the limits of, the application of static concepts and ideas to a flowing experience of a dynamic reality." In the carefully crafted and designed experience of multimodal interfaces, the application of these phenomenological theories is indeed both necessary and limited.

2.3.2 Methodology: Creating and Studying Experiences

The methods that are used throughout this thesis address the key concerns of the theoretical foundations presented here. Firstly, there is a significant focus on self reported experience and subjective user accounts. Secondly, the analysis of these accounts revolves around everyday experiences that might otherwise be taken for granted and requires a change in attitude with respect to questioning these everyday experiences. When evaluating the user experience of an interaction, we are often trying to design for a specific experience or, perhaps more often, trying to identify and understand the experience that we've already created. By adopting qualitative methods, we can gain access to experience through observations and interviews. Quantitative methods can also be used to gather objective data about experience. For example, accelerometer data can be collected throughout an interaction with a system as one measure of the observable aspects of that interaction. This can be combined with qualitative data to gain additional perspectives on the experience and in some way combine qualitative sociological methods and quantitative computing methods of evaluation harmoniously. Therefore, the methodology presented here is one of mixed methods that aims to create experiences and gather quantitative data as much as qualitative data for an analysis that draws from both.

Research as a Social Encounter

An important aspect of user experience research that must not be taken for granted is that this research is completed in the context of a social encounter that comes with predefined roles, expectations, and norms and standards. The social encounters that have been created for the purpose of user studies, whether completing an interview, an observation session, or even the use of quantitative sensing to record behaviours remotely, must be designed to deal with the issues associated with the social encounter. For example, the manner, appearance, and style of questioning of the interviewer has a significant impact on the interview itself (Schensul *et al.*, 1999). The role of experimenter and participant in a user study changes how participants behave because of the 'authority' of the experimenter, with the classic Milgram experiment of obedience as an extreme example of this effect (Milgram, 1963). In less extreme situations, it is still clear that simple observation is likely to affect users' behaviour (McCarney *et al.*, 2007). These issues are important to consider when designing an experiment in order to prevent this influence as much as possible and when completing analysis in order to take into account any influences which might have affected the results. The identification and acceptance of the influence of a social encounter as part of research is an important consideration for rigorous qualitative studies.

Another important issue is that the experience or phenomenon that is the focus of study often isn't practical or accessible to be studied directly in the context in which it appears. For

example, during a longitudinal deployment of a multimodal mobile system it is not practical or ethical to constantly shadow a participant in order to observe interaction as it naturally and sporadically occurs throughout the day. Additionally, it is often the case that the experience itself is not easily observable or that the observations themselves aren't enough to understand the experience thoroughly. For example, observations alone may not be sufficient to understand why a user chose *not* to use a gesture-based interaction while commuting. In many cases, the study of user experience must be completed outside the actual context where that experience occurs through interviews and discussions after the fact. Only through thorough analysis of self-reported motivations and experiences can the researcher gain a complete understanding of observed behaviours.

Surveys

Surveys have been used throughout this work, focusing on gathering initial reactions to and acceptance of multimodal interaction techniques. Particularly, these surveys have used video prototypes and multiple choice style questions. This format can be useful for gathering initial responses because a video prototype can be developed early in the design process and therefore allow researchers to explore a divergent set of designs and ideas. The content of the questions used in this thesis has been based on social factors that influence behaviour in public places such as location, audience, and appearance. Successful examples of surveys used for social acceptance can be seen in Rico and Brewster (2010b) and Ronkainen *et al.* (2007).

Focus Groups

The general design and practice of focus groups has been discussed in detail elsewhere (Stewart *et al.*, 2007). The focus groups used in this work have been specifically designed to create an experience and gather quantitative and qualitative data about that experience. A successful focus group looking at user experience should have the following elements: initial familiarisation with the interaction techniques through the use of experience prototypes (Buchenau and Suri, 2000), collection of quantitative data relating to acceptance of interaction techniques, and discussion about the use of these interactions in real world settings with respect to how the interaction techniques might be adopted. Collecting quantitative data about acceptance of the interaction techniques is an important part of the first phase of a user experience focus group. This should be done individually before any in depth discussion of the interaction techniques with the group. Typically, this should consist of Likert scale style ratings, accept/reject responses, and rankings of the interaction techniques. The collection of this quantitative data should be followed by the qualitative discussions about the interaction techniques. A good example of this style of focus group can be seen in Rico and Brewster (2010a).

Interviews

The interview method is often used in sociological research from a variety of related theoretical perspectives, including phenomenology, symbolic interactionism, and ethnomethodology.

The techniques (Schensul *et al.*, 1999) and critiques (Benson and Hughes, 1983) of this method are described in detail elsewhere. Most importantly, these works describe the design of interview questions, the manner and appearance of the interviewer and the effect this has on the success of the interview itself as a social encounter.

Throughout the course of this thesis, *semi-structured* interviews were used (Schensul *et al.*, 1999). Although the title might suggest these interviews lack organisation, this term simply means that the interview is administered from a series of set questions with open ended responses. Often, the interviewer will probe users' responses with further questions to flesh out the detail in the responses and follow up any interesting comments. The design of the questions provides the order and structure to the interview while still allowing the interviewer to temporarily leave the set questions when necessary. One of the great advantages of interviews is that they can easily be completed at most stages of development, in order to look at acceptance, adoption, and appropriation.

User Studies

Designing and deploying user studies is a good way to look at appropriation and actual system use, but can also be completed in the early stages of design using prototypes to investigate acceptance and adoption. User studies that focus on user experience come in a variety of forms, from using simple experience prototypes (Buchenau and Suri, 2000), to completing controlled 'on-the-street' studies (Rico and Brewster, 2010b), to relatively uncontrolled longitudinal deployments (Williamson *et al.*, 2011). In all of these studies, however, the goal is to create an experience for users and gather as much data as possible through sensing technology, interviews, questionnaires, and observations. This experience can then be studied through the variety of qualitative and quantitative data collected throughout the experiment.

This thesis, like Sherwood *et al.* (2009), encourages the evaluation of user experience in the wild. However, user studies in the wild present difficulties. Evaluations using a variety of sensors, known as instrumented usability (Crossan *et al.*, 2008), use the powerful technique of remote observation when it isn't practical to physically observe system use. By instrumenting the user with sensors it is possible to detect some aspects of the current context of use, such as whether the participant is walking, riding public transport, or still. A good example of an instrumented usability study looking at user experience can be seen in Murray-Smith *et al.* (2007). This study used accelerometers in order to detect gait phase during mobile phone calls to identify conversation synchronisation.

2.3.3 Analysis: Understanding Through Metaphors

When interaction is viewed as performance, experience can be understood and analysed with respect to performers, characters, audiences, and props. When trying to understand, evaluate, and analyse user experience, the use of dramaturgical metaphors presents a compelling approach. These metaphors provide an intuitive method for designing questions, organising themes in data, and understanding behaviour. Goffman's use of dramaturgical metaphors, as

described in (Goffman, 1966), (Goffman, 1986) and (Goffman, 1990), has been widely adopted in social science and human computer interaction research. Although Goffman's work has been heavily influenced by phenomenology, especially Schutz's work (Lanigan, 1988), and is often associated with ethnomethodology (Smith, 2005), it is in fact a-theoretical and critics claim that Goffman fails to analyse his results to the full depth required in traditional phenomenology (Lanigan, 1988). Goffman did not claim to be a phenomenologist, but used only specific aspects of the theory to guide his work. Like Goffman, this approach will take the theoretical benefits of phenomenology and apply them practically to the method and analysis of user experience.

Dramaturgical Metaphors

Using dramaturgical metaphors to understand interaction means taking concepts from theatrical performance or explicit performance and applying them to everyday situations. In order to organise these concepts, Goffman describes the ideas of *fronts* and *regions* (Goffman, 1990) as ways in which we gather information about our context in order to perform appropriately. These traditionally theatrical concepts describe everyday surroundings and how decisions are made based on those surroundings. *Fronts* contain physical and social aspects of the immediate surroundings that help to define the situation for individuals. *Regions* describe more abstract places and their associated perceptions and norms and standards. Together, fronts and regions provide individuals with information needed to define the situation and determine the acceptability of their actions. By identifying and organising in this way the aspects of everyday life that influence behaviour, researchers can begin to understand experience through the performative lens.

Fronts – Goffman defines fronts as "that part of an individual's performance which regularly functions to define the situation ... the expressive equipment of a standard kind intentionally or unwittingly employed." (Goffman, 1990). This *expressive equipment* includes the physical front, the personal front, and the social front. The physical front is the actual setting where performances occur, which includes the furniture and objects that fill a given space to make up a physical setting, similar to a stage. Depending on the intended use of the 'stage', the physical setting may provide props or scenery that facilitate a given performance in real life as they do in theatrical productions. In an office, a desk and chairs will create a more appropriate setting than a hammock. The setting will inform an individual's definition of the situation and help to determine appropriate action. For example, mobile phone users in an office might take advantage of cubicle walls to muffle a personal conversation. The absence of cubicle walls might influence users to reject mobile phone calls from friends at their desk. The personal front includes aspects of appearance and manner of an individual, as costume and makeup would for actors. The appearance and manner of the performer will influence how they feel their actions will be perceived and influence how they interpret and make sense of spectator reactions. The social structures that exist in a given situation are referred to as the social front. The social front includes the learned set of expectations and behaviours associated with a social status, and norms and standards, which influence how we make sense of the world around us. For example, a manager in their place of business might dress more formally than at their home and social norms and standards dictate that a certain level of respect is expected

from others. Both manager and workers know the expected norms and standards and present themselves and treat others appropriately. Social fronts are used in order to understand what might be expected in a given situation and how appearance and manner should fit together. Just as these existing social structures are used to act out stories that are familiar and make sense, these social fronts are used by individuals to make sense of everyday life.

Regions – Goffman defines regions as "any place that is bounded by some degree by barriers to perception." (Goffman, 1990). For example, an office might be separated into distinct regions by cubicle walls, where my desk is a distinctly different area than your desk. A region includes a front region, politeness and decorum, and the back region. The front region is like the front stage and refers to the place where the performance occurs, similar to the physical front. The front region includes the physical setting and the objects that fill the space and defines the setting physically. Politeness and decorum define the region cognitively or emotionally. Politeness refers to the expected behaviours when an individual is actively engaged with others. Decorum refers to the expected behaviours when an individual is present in the region but not actively engaged with others. For example, it may be unacceptable to answer a phone call while engaged in conversation at a party, but perfectly acceptable if standing off to the side. In a different setting, such as a movie theatre, it is unacceptable to answer a phone call even if you are not personally watching the movie. Politeness and decorum reinforce region by dictating the actions that are expected in that situation whether one is actively engaged with that situation or just simply present. The back region is similar to backstage, which refers to any place where the individual is not 'performing.' The back region is a place where the decorum and politeness of the region are more relaxed or completely ignored. For an office, the break room can be viewed as the back region of the workplace. While in the break room, there is no pressure on employees to look busy or partake in work related activities. Regions organise performance spaces by defining clear areas for performance or non-performance and defining the type of 'work' that is required to demonstrate and reinforce the region.

Key Concepts

In order to effectively apply these dramaturgical metaphors in practice, this chapter distills these metaphors into concepts that can be used to guide the design, execution, and analysis of user experience studies. These key concepts provide guidance and purpose when completing both qualitative and quantitative user experience research from this performative perspective. From designing survey questions, to completing interviews and coding results, dramaturgical metaphors can guide this work and help to organise results. The key concepts that this methodology draws from dramaturgical metaphors are described in detail below.

Understanding Location – Identifying how a user understands, perceives, and feels about the location of interaction is an importance aspect of experience that is emphasised by the dramaturgical approach. Where did the interaction occur? How does the user describe the location? What norms and standards does the user describe when talking about different locations? How familiar is the user with the different locations discussed? These questions are important for understanding the setting, as perceived by the user, where interaction occurs. Physical fronts and regional fronts have a strong influence on the presentation and behaviour

of individuals in public places and must be clearly identified by the researcher.

Perceptions about Performance – Subjective experience and individual opinions about how the performance looks and feels are important aspects of experience. What aspects of the performance do users like or dislike? What visual aspects of the performance are identified when discussing preferences? How do users describe comfort or discomfort when talking about performance? How do users feel that they look while performing? How do users describe their own personality when discussing performance? These kinds of internal perceptions about the enjoyability, acceptability, and overall experience of performing are important for the researcher to identify in order to understand the personal front of the performer.

Reactions from Spectators – How users make sense of the reactions of spectators and how these influence future behaviour are an important aspect of experience. Who are the spectators that are present during a given interaction? What do users think about spectators in the different locations where interaction occurs? What did the spectators do during a given interaction? How do users describe the meaning of spectator reactions? Because spectators have a clear influence on individual behaviour, understanding how the individual makes sense of spectator reactions helps to make sense of motivation and decision making.

Recalling Past Experiences – The way in which users tell stories about past experiences, the opinions and values that are associated with past experiences, and the self-reported reasons for overall judgements about past experiences are important. What experiences do users most often discuss? How do users discuss these experiences as evidence for their opinions? Which aspects of a past experience are the most important for users? Gathering self-reported data about past experiences allows researchers to understand what information users are building on when making decisions in public places and how behaviours are affected by this.

Coding and Analysis

When coding any transcripts or observation logs, there are a wide variety of approaches one may adopt when deciding which themes to pull out, how to discover interesting content, and which parts of the transcript require coding. A typical analysis involves three steps: open, axial, and selective coding. These coding processes are described in great detail in Strauss and Corbin (1998). In essence, open coding represents the first stage of analysis where tags or labels are applied throughout a transcript. During open coding, the concepts described above represent the types of content which are interesting and should be identified. The second stage, axial coding, involves organising all of the labels and tags. This organisation is led by the concepts and metaphors laid out by Goffman. The final stage, selective coding, takes those organised concepts and pulls out the overall themes and the generalizable results of the study.

Another important concept in coding and analysis is *strips* (Goffman, 1986). These form an important part of any qualitative analysis as the starting or entry point into a stream of activity that is the focus of analysis. Goffman defines strips as "any arbitrary slice or cut from the stream of ongoing activity ... A strip is not meant to reflect a natural division made by the subjects of inquiry or an analytical division made by students who inquire." (Goffman, 1986). A strip can refer to the collected ethnographic observations or the self-reported experiences of

an individual for whatever activity or experience being studied.

2.4 Conclusions

This chapter presents a novel performative approach to user experience research based on phenomenology and dramaturgical metaphors. Although this perspective on user experience focuses on multimodal interaction in this thesis, it also has a much broader applicability because of the focus on 'everyday performance.' A wide variety of interactions can be viewed through this lens, such a mobile interfaces, the presentation of online personas through social networks or avatar based games, or interaction with public displays. The motivation of these explorations of user experience is to better understand how users accept, adopt, and appropriate new technology into their everyday lives, especially where interaction occurs in public or social settings. With respect to multimodal interfaces, where performance plays a significant role in user experience, these kinds of evaluations are important because the social acceptability of multimodal interaction techniques affects how it can be used in everyday life. By evaluating user experience at different stages of use, this approach can lead to more usable, enjoyable, and acceptable multimodal interfaces.

The performative perspective on user experience begins with theoretical foundations in phenomenology. The application of this theory to user experience research means a significant focus on subjective experience and the analysis of the everyday actions that might be taken for granted by performers themselves but be of great interest to researchers. The methodology argues for a mixed methods approach, combining traditional ethnographic methods from the humanist tradition of sociology with quantitative usability methods from computing science. This approach allows for gathering a wide variety of data for a thorough and robust analysis. Finally, this approach relies heavily on dramaturgical metaphors as a way of understanding performance in everyday life and how interaction with technology fits into existing practices. These metaphors guide data gathering and analysis by organising results in terms of performers, stages, props, scenarios, and norms and standards.

The approach described in this chapter will be demonstrated in this thesis with increasing complexity and sophistication in each chapter. The third chapter discusses how this approach can be used to determine initial reactions to multimodal interactions using very highly controlled prototype experiences. The fourth chapter goes on to discuss interaction acceptance and adoption in more detail, asking users to imagine how interactions could be used in daily life and creating relatively controlled experiences in real world settings. The fifth chapter explores interaction appropriation in the wild, demonstrating the methodology from beginning to end to create a relatively uncontrolled experience in real world settings. This complete demonstration of the methodology shows how it can be used at different stages of design and development and rigorously tests it in different settings.

Chapter 3

Reacting to and Accepting Multimodal Interfaces

3.1 Introduction

Davis (1989) argues that the first step towards accepting a new interaction technique or technology is understanding the 'perceived usefulness' and 'perceived ease of use' for that new technology. For those new interaction techniques and technologies that also require users to accept more performative interactions, such as interfaces using gestures that require users to perform highly visible actions, social factors also play an important role in technology acceptance. Users might ask themselves "will someone notice me doing this gesture?" or "will someone think this is a strange thing to do?" Davis's original Technology Acceptance Model did not include social factors, but others have identified this missing element and extended the original model (Malhotra and Galletta, 1999) to take these factors into account. User experience evaluations allow designers to understand how personal and social factors affect the acceptance and use of technology.

This initial acceptance phase is simply a user's first assessment of a new technology and acceptance of the possibility of its use in the future. In this chapter, evaluations of acceptance focus on interaction techniques by removing these interactions from a specific usability context. For example, gesture-based commands are presented to users without these commands being given any particular purpose or functionality within a specific interface or application. The usability context is removed in order to minimise the influence other external factors, such as the usefulness of the *application area* or the *appropriateness* of the technique for that application. Removing as many factors external to acceptance as possible and evaluating interaction techniques out of an interaction context allows for exploration of the baseline of acceptability. This approach provides a solid basis for looking at acceptance independently of other influencing factors, where more in-depth evaluations looking at adoption and appropriation can investigate interaction techniques in more realistic contexts. This chapter focuses on those initial reactions, specifically at early stages of the design process.

This chapter explores this initial acceptance phase through three studies. Firstly, a survey that explored social factors of acceptance using video prototypes and multiple choice questions. This survey focused on location and audience as social factors that affect technology acceptance. Secondly, an 'on-the-street study' that required users to perform a set of gestures in public in order to understand why users liked or dislike these interactions. Finally, a second survey study that explored acceptance based on social factors using more complex social settings. This study examined how different spectators in specific locations affect technology acceptance. This chapter concludes with a reflection on these results and on the methodology itself.

3.2 Gathering Initial Reactions to Gesture-Based Interactions

In order to gather some initial data on which interaction techniques users might accept or reject, an Internet survey was completed using video prototypes and multiple choice questions. The video prototypes portrayed gesture-based interaction techniques, and the multiple choice questions investigated possible situations where those interaction techniques would be acceptable. Gestures were chosen as the focus of this first study because of their highly visible nature and the ease of portraying these interactions on video. The multiple choice questions investigated location and audience as factors that influence behaviour in public places (Goffman, 1990) and provide a useful basis from which to begin exploring social acceptability. The results of this survey provide an initial insight into social acceptance of gestures with respect to the locations where they might be used and how different kinds of spectators influence social acceptability.

3.2.1 The Survey

This survey examined a set of eighteen gestures that included both *device-based* and *body-based* gestures. A body-based gesture refers to any gesture that directly involves movements of the body without directly manipulating a mobile device. For example, head nodding and foot tapping are both gestures that are performed by moving the body without directly manipulating a device. A device-based gesture refers to any gesture that involved directly touching or moving a typical mobile device, in this case a mobile phone. For example, device shaking and squeezing both involve directly manipulating a mobile device. The gestures in the survey were chosen based on their usage in existing gesture-based interfaces (Brown and Williamson, 2007), (Crossan *et al.*, 2008), (Strachan *et al.*, 2007), (Williamson *et al.*, 2007) and their potential to be used in future interfaces. The names and descriptions are given in Figure 3.1. The full list of survey questions can be seen in Appendix A.1 and the videos are available in Appendix B.

Before beginning the survey, respondents were told the aim of the study was to assess the social acceptability of these gestures for use as part of a mobile interface. For each gesture, survey respondents watched a video of the gesture being performed and answered multiple-choice questions. Each video was displayed with the gesture name and a short description of

Device-Based Gestures	
Device Whip	Whip-like movement of the device in the hand.
Device Shake	Shaking the device in the hand, side to side.
Pocket Tap	Tapping the device (in the pocket) with the hand.
Device Squeezing	Squeezing the device in the hand.
Table Tap	Tapping the device (on a table) with the hand.
Shoulder Tap	Tapping the shoulder with the device.
Device Rhythm	Shaking the device in the hand up and down with an even rhythm.
Body-Based Gestures	
Shoulder Rotation	Rotation of the shoulder forwards and backwards.
Hand Proximity	Measuring a space between the hands.
Nose Tap	Tapping the nose with the hand.
Head Nodding	Nodding the head side to side.
Shifting Balance	Shifting balance side to side while sitting.
Wrist Rotation	Rotation of the wrist side to side.
Clapping	Clapping the hands.
Foot Tapping	Tapping one foot while sitting or standing.
Belt Tapping	Tapping the belt with the hand.
Finger Proximity	Measuring a space between the fingers.
Arm Squeezing	Squeezing of the forearm with the opposite hand.

Figure 3.1: Names and descriptions of all the gestures used in the survey.

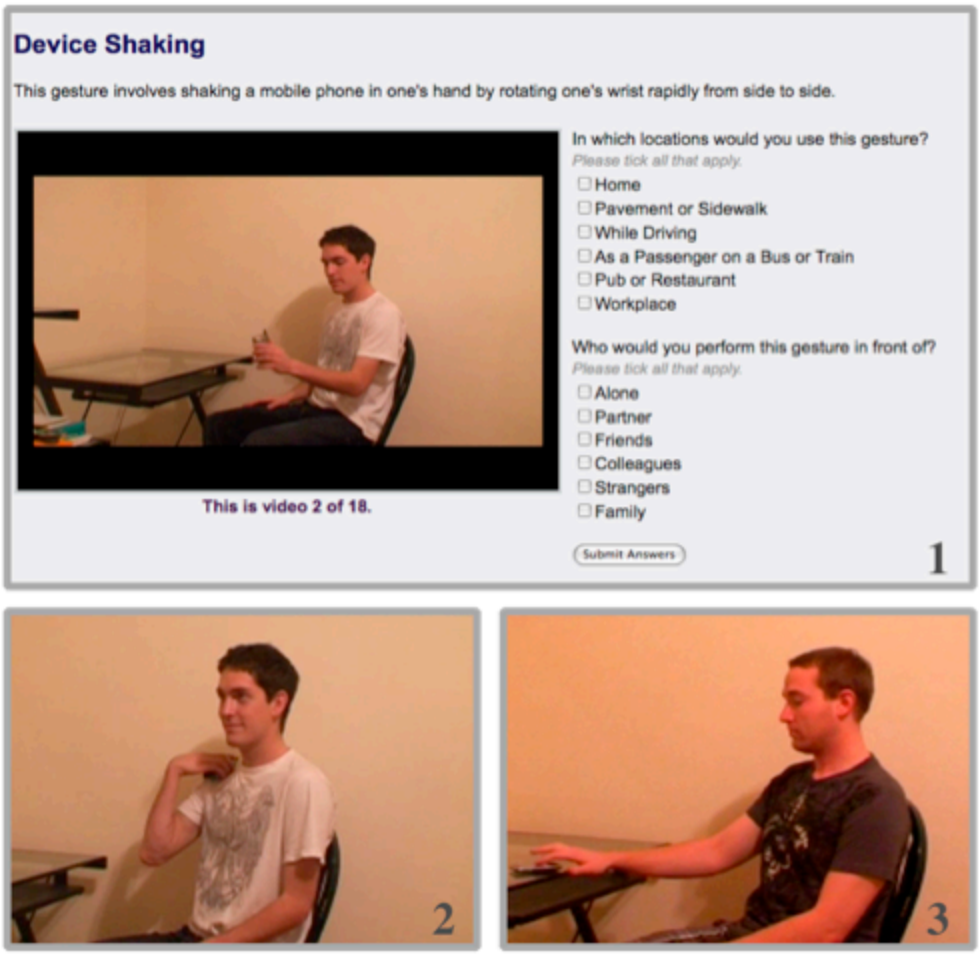

Figure 3.2: Screenshots and video frames from the gesture survey. 1. Survey Questionnaire Layout. 2. Frame from Shoulder Tapping video. 3. Frame from Table Tapping video.

Location	Audience
Home	Alone
Pavement or Sidewalk	Partner
Driving	Friends
Passenger on a Bus or Train	Colleagues
Pub or Restaurant	Strangers
Workplace	Family

Figure 3.3: Locations and audiences for multiple choice survey questions.

the gesture portrayed. The videos lasted from one to four seconds and were played on a loop while survey respondents answered the associated questions. These videos did not include any audio. Each video portrayed a single gesture being performed by a male actor sitting at a desk in front of a plain background. Figure 3.2 shows screenshots of the survey questions as well as frames from two of the gesture videos. Because survey respondents were asked to imagine the locations and audiences where they might perform these gestures, the videos were designed to focus solely on the gesture itself. The videos used in this survey intentionally portrayed a plain scene without a defined context so that the setting would not distract viewers from evaluating the gesture. Since an important role of the videos was to make gestures easy to observe, they were performed using clearly visible movements. For example, wrist rotation was performed with the arm held in front of the performer rather than at the side of the body. While these constraints did not always reflect real world usage, this was necessary to give respondents a clear view of the gesture being performed.

After watching each video, respondents were asked to select from a list all of the locations where they would be willing to perform the given gesture as part of a mobile interface. Respondents were then asked to select from a list all of the types of audiences they would be willing to perform the gesture in front of. These responses intentionally focused on personal experiences, asking respondents to imagine *themselves* in these settings rather than how they would feel about someone else performing the given gesture. The locations and audiences used in the survey are given in Figure 3.3. These options were selected to address a broad cross-section of common situations without providing an overwhelming number of options for survey respondents. Removing the usability context helped to remove unwanted influences from responses that might not be helpful in determining acceptance. For example, a video prototype could demonstrate a gesture being used in order to open the phone's camera application. If respondents state that they would not use this gesture, it is not possible to know if this is because they find the gesture unacceptable, they find the gesture inappropriate for the application, or they simply do not use that application. Removing the usability context helps to remove these unwanted external influences. While this style of survey does not describe *why* the gesture is unacceptable, it does provide a useful method for gathering early reactions to new interaction techniques at a low development cost.

Location	p value	Audience	p value
Home	0.276	Alone	0.152
Pavement	**< 0.0001**	Partner	**< 0.0001**
Driving	**< 0.0001**	Friends	**< 0.0001**
Passenger	**< 0.0001**	Colleagues	**< 0.0001**
Pub	**< 0.0001**	Strangers	**< 0.0001**
Workplace	**< 0.0001**	Family	**< 0.0001**

Figure 3.4: p values calculated using the Cochran Test for significant differences based on gestures for locations and audiences. $p < 0.003$ are significant, adjusting for Bonferroni's correction.

3.2.2 Results

These results include 55 survey responses from respondents ranging from age 22 to 55. With respect to current geographic location, 43% of respondents were living in the United Kingdom, 45% in the United States, and 12% declined to state or were the only respondent from their country. 29% of the respondents were female and 71% were male. Respondents were recruited through university email lists over the span of six weeks. The survey was administered over the Internet.

The results show that location and audience played a significant role in determining the social acceptability of performing gestures. There were significant differences between acceptability rates of gestures for all locations except home and for all audiences except alone, as shown in Figure 3.4. The non-parametric Cochran Test was used for testing this binary response data of related samples (Conover, 1971). Location and audience make a significant difference when respondents decided to accept or reject a given gesture, demonstrating how important gesture design is to acceptance and eventual usage.

Looking at the data in greater detail, a level or rating of 'acceptability' was determined for different groupings of the responses. For example, survey responses were grouped by location, audience, and gesture type. An acceptability rating was calculated for each respondent using the ratio of positive responses to negative responses for a given grouping. The Friedman Test was used for the overall comparison of three or more treatments of ordinal non-parametric data (Cramer, 1994). The Wilcoxon Signed Rank Test was used for pairwise comparisons for related samples of non-parametric data (Cramer, 1994). Effect sizes were calculated using the Z values of the pairwise tests (Corder and Foreman, 2009). All of the significance value thresholds were adjusted using Bonferroni's correction.

Where? Locations and Gesture Usage

The awareness of one's surroundings and the ability to take advantage of the physical setting are important aspects of a performance. By comparing average acceptability rates between locations, these results demonstrate which settings were the best facilitators of gesture usage

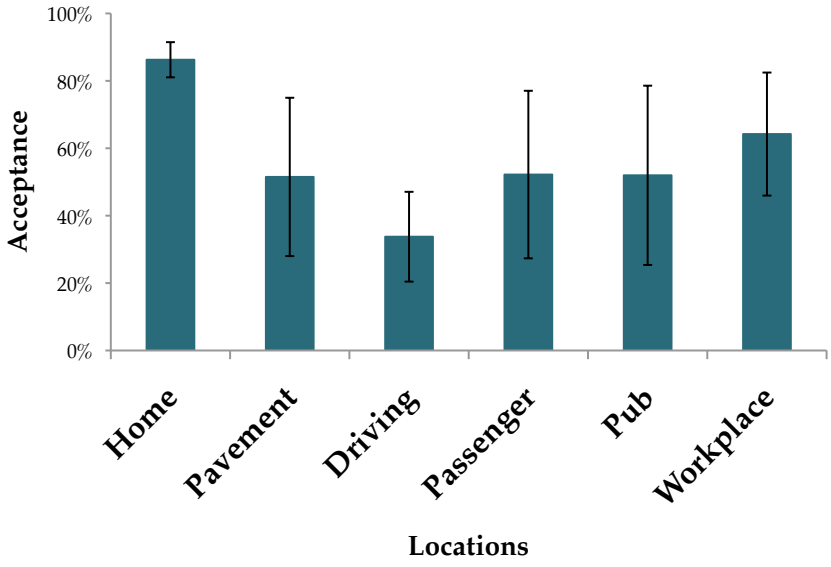

Figure 3.5: Average percentage of gesture acceptability by location. Error bars show one standard deviation.

	Home	Pavement	Driving	Passenger	Restaurant	Workplace
Home		**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Pavement			**0.0002**	0.4354	0.6745	**< 0.0001**
Driving				**< 0.0001**	**0.0002**	**< 0.0001**
Passenger					0.9442	**< 0.0001**
Restaurant						**< 0.0001**
Workplace						

Figure 3.6: p values for pair-wise comparisons of locations. $p < 0.003$ are significant, adjusted for Bonferroni's correction.

and which settings were the most controversial. Figure 3.5 shows the average acceptance rate and standard deviation for gestures in each location.

Figure 3.6 gives the details of pair-wise comparisons for statistical significance of locations and Figure 3.7 shows the effect sizes. The majority of locations were significantly different from every other location, with the exceptions in the combinations of passenger and pavement, pub and pavement, and passenger and pub. Home, workplace, and driving were each significantly different from every other location. These differences show that location played an important role in how respondents determined the acceptability of a given gesture. For example, the significant difference between pavement and workplace shows that these places were socially different for survey respondents with respect to acceptability. Gestures were 12% more likely to be used in the workplace than on the pavement and 34% more likely to be used at home. The pavement, as a public location, provides few opportunities for privacy and requires a more restricted set of social norms due to the necessity of sharing the space with strangers. Home, on the other hand, offers many opportunities for privacy and relatively relaxed social norms as it is a familiar place shared with family members. The workplace is semi-public, where

	Home	Pavement	Driving	Passenger	Restaurant	Workplace
Home		**0.8**	**0.8**	**0.8**	**0.8**	**0.8**
Pavement			**0.6**	--	--	**0.6**
Driving				**0.6**	**0.6**	**0.7**
Passenger					--	**0.6**
Restaurant						**0.6**
Workplace						

Figure 3.7: Effect sizes for pair-wise comparisons. Values 0.5 and over are considered large effect sizes. Effect sizes for non-significant comparisons are not reported.

colleagues may be familiar but the space is not private. In each location, individuals consider social acceptability based on social norms and standards as well as privacy affordances to determine if they will accept a given gesture.

An important aspect of the data is the differing levels of variance between the gestures' acceptability for each location. Higher levels of variance indicate that survey respondents had more varied opinions about which gestures were acceptable and which were not in a given location. Comparing home, with a standard deviation of 5%, to pub, with a standard deviation of 27%, shows a major difference in the way these locations were perceived. While home was relatively uniform in its perception, pub was very divergent with some gestures being highly acceptable and others highly unacceptable. Acceptance rates varied from as low as 14% for shoulder rotation to as high as 95% for table tapping. These differences show how important gesture design is in order to create systems that are usable in a variety of locations.

Who's There? Audience and Gesture Acceptability

The average acceptance rates and standard deviations for each audience are shown in Figure 3.8. Each audience type was significantly different from every other type except for the combinations of friends and family, partners and family, and partners and friends as shown in Figure 3.9. The effect sizes for these pair-wise comparisons are shown in Figure 3.10.

Familiarity with the audience plays a significant role in gesture acceptability. Strangers, the least familiar audience with an average acceptability rate of 51%, were significantly lower than partners, with an average acceptability of 77% and alone with an average acceptability of 88%. This shows that more familiar audiences gave performers more confidence in using gestures than less familiar audiences. For example, 60% of respondents were willing to use the shoulder rotation in front of their partner, 54% were willing to use it in front of their friends, and only 20% were willing to use this in front of strangers. Familiar audiences provide performers with more freedom to try new things and more opportunities to explain their behaviour. Friends or family members may have previous experiences witnessing an individual's usage of a particular gesture while a stranger may only be present during one isolated event. The presence of spectators is an important aspect of experience and performance because spectators provide feedback to users that help them make decisions about use. Reeves *et al.* (2005) discuss how the visual aspects of performative interaction play an important role in the spectator experience.

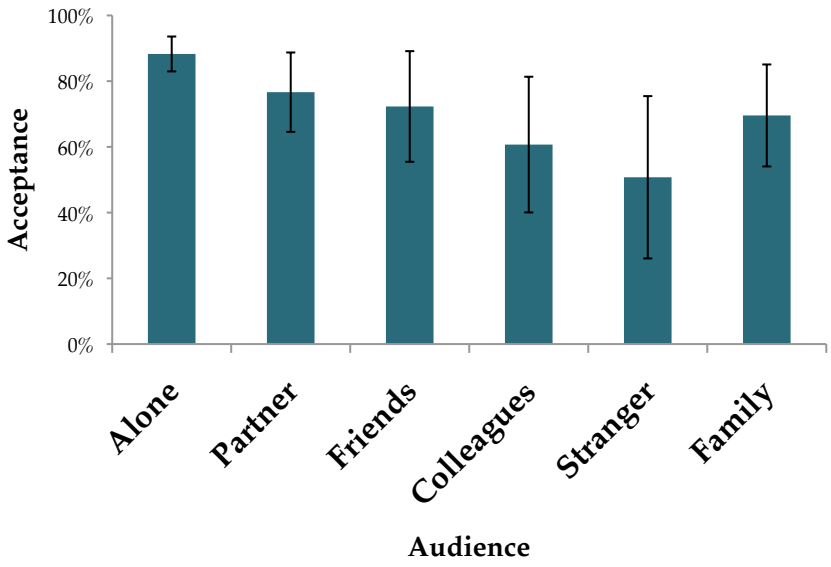

Figure 3.8: Average percentage of gesture acceptability by audience. Error bars show one standard deviation.

	Alone	Partner	Friends	Colleagues	Strangers	Family
Alone		**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Partner			0.04	**<0.0001**	**<0.0001**	0.004
Friends				**< 0.0001**	**<0.0001**	0.0344
Colleagues					**0.0004**	**0.0016**
Strangers						**< 0.0001**
Family						

Figure 3.9: p values for pair-wise comparisons of audiences. $p < 0.003$ are significant, adjusted for Bonferroni's correction.

	Alone	Partner	Friends	Colleagues	Strangers	Family
Alone		**0.6**	**0.7**	**0.8**	**0.8**	**0.7**
Partner			--	**0.7**	**0.7**	--
Friends				**0.7**	**0.7**	--
Colleagues					**0.5**	**0.5**
Strangers						**0.6**
Family						

Figure 3.10: Effect sizes for pair-wise comparisons. Values 0.5 and over are considered large effect sizes. Effect sizes for non-significant comparisons are not reported.

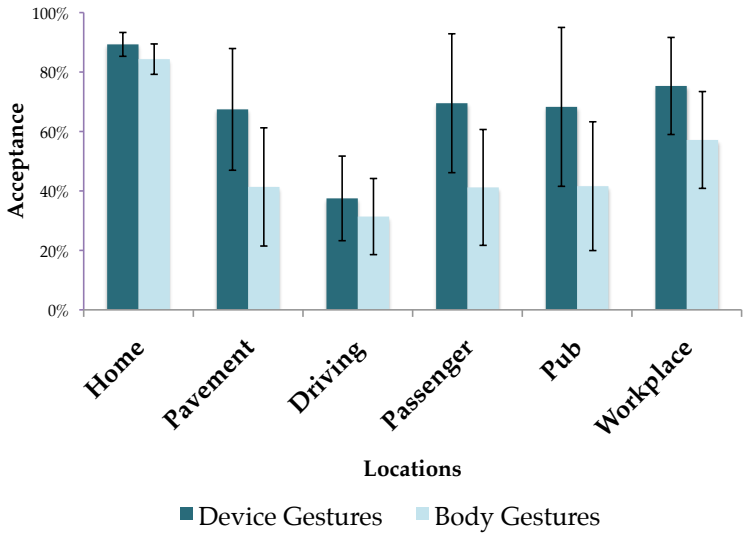

Figure 3.11: Average acceptability of device and body-based gestures by location. Error bars show one standard deviation of acceptability for each gesture type at the given location.

These visual aspects of gesturing also play an important role in the individual user experience since users make decisions about acceptability based on the responses of spectators.

Device and Body-Based Gestures

The survey demonstrates a significant difference between the social acceptability of device-based gestures versus body-based ones. Pair-wise comparisons, shown in Figure 3.12, demonstrate that device-based gestures were significantly more acceptable at every location except home and driving and every audience type. Figure 3.13 shows the effect sizes for these comparisons. The average acceptance rates and standard deviations of the device and body-based gestures are shown in Figure 3.11.

It is possible that respondents found device-based gestures more acceptable because these include the visibility of the device itself. Device-based gestures provide clear indicators for audience members that a performer's actions are directed towards the mobile interface as part of some interaction. However, there were some body-based gestures with high levels of acceptability. Foot tapping, for example, had an acceptance rate of 88% based on audience. This gesture utilised subtle, everyday movements. The ability to disguise gestures as everyday activities is one way that body-based gestures can become more acceptable.

3.2.3 Discussion

This survey demonstrated that individuals take into account location and audience when deciding to interact with a gesture-based interface, having a significant effect on which gestures were acceptable based on these factors. The results indicate that location and audience, which have already been used to describe behaviour in public places (Goffman, 1990), have a signif-

Location	p value	Audience	p value
Home	0.01	Alone	**0.002**
Pavement	**< 0.0001**	Partner	**< 0.0001**
Driving	0.03	Friends	**< 0.0001**
Passenger	**< 0.0001**	Colleagues	**< 0.0001**
Pub	**< 0.0001**	Strangers	**0.0002**
Workplace	**< 0.0001**	Family	**0.0002**

Figure 3.12: p values for pair-wise comparisons of device versus body-based gestures. $p < 0.003$ are significant, adjusted for Bonferroni's correction.

Location	Effect Size	Audience	Effect Size
Home	--	Alone	0.4
Pavement	**0.7**	Partner	**0.6**
Driving	--	Friends	**0.7**
Passenger	**0.7**	Colleagues	**0.7**
Pub	**0.7**	Strangers	**0.7**
Workplace	**0.6**	Family	**0.6**

Figure 3.13: Effect sizes for pair-wise comparisons for device versus body-based gestures. Values 0.5 and over are considered large effect sizes. Effect sizes for non-significant comparisons are not reported.

icant influence on gesture usage and acceptance as part of a mobile interface. Locations that provided respondents with more privacy were more likely to have higher acceptability rates than those that did not. With respect to audience, respondents were more likely to perform gestures in front of familiar audiences. Gesture type also influenced respondents' willingness to accept gestures as part of a mobile interface. Device-based gestures, which provided visual cues to spectators about the performers' actions, were more likely to be used than body-based gestures. Because the device is visible during these gestures, respondents felt more comfortable with the idea of performing these gestures because their actions would be more easily understood by spectators. The ability to demonstrate a gesture or action is part of a mobile device plays in important role in gesture acceptability. Although this work shows that a visible device is good at 'demonstrating' the interaction, other methods of demonstrating interaction could be developed to improve the acceptability of new interactions that do not require users to hold a device. For example, interactions that are reactions to some noticeable action, like a ringtone, could demonstrate the purpose of the interaction without using the visibility of the device.

This survey also demonstrated the fact that the social acceptability of gesturing in certain locations is often heavily coupled to the typical audiences expected in those locations. People expect the pavement to be full of strangers and expect their partner or family members to be in the home. These issues indicate that survey questions should present a more clearly defined situation by combining location and audience. Looking at these factors together in different combinations would allow survey questions to explore more thoroughly the effect that spectators have on interaction. For example, the second survey described in this chapter looks at how familiar audiences might make respondents more comfortable performing gestures in highly public locations. If you are performing a gesture on the pavement, does having your friends with you make this significantly more comfortable? When you are performing a gesture at home, does being alone versus with family make a significant difference to the acceptability of that gesture? Another interesting issue comes out of more unusual combinations of locations and audiences, such as unfamiliar audiences in highly private or personal locations. For example, how acceptable is it to perform some gesture at home alone versus at home when entertaining work colleagues? The combination of different locations and audiences studied together allows for a richer understanding of context and social norms and standards.

While these survey results provide an interesting insight into the social acceptability of gestures, they are based on imagined situations rather than real ones. The survey method is a useful tool for gathering early responses, but respondents are asked to imagine a given context without having the chance to actually perform the gestures in those contexts. Additionally, while these results provide some interesting insights into which gestures users find acceptable, this quantitative survey does not examine *why* some gestures are acceptable while others are not. The following study builds on these results in order to explore the aspects of gestures that were liked or disliked, the experience of performing in different locations, and how preferences developed through multiple experiences.

Body Based Gestures		Device Based Gestures	
Wrist Rotation	Acceptable	Shaking	Acceptable
Foot Tapping	Acceptable	Screen Tapping	Acceptable
Head Nodding	Unacceptable	Rhythm	Unacceptable
Nose Tapping	Unacceptable	Shoulder Tapping	Unacceptable

Figure 3.14: Gestures used in the on-the-street user study.

3.3 On-the-Street: Bringing interactions into realistic settings

Eight gestures from the previous gesture survey were investigated in this 'on-the-street' user study. These were chosen to include four body-based gestures and four device-based gestures. For both gesture categories, two acceptable and two unacceptable gestures were chosen in order to collect responses about why users liked and disliked the interaction techniques. Acceptability was determined using the acceptance rate for each gesture from the initial gesture survey. Unacceptable gestures were those with comparatively low acceptance rates as compared to other gestures of the same type. These are shown in Figure 3.14.

3.3.1 The Study

Study participants each attended three individual user study sessions. These sessions were spaced one week apart and lasted roughly one hour each. These repeated sessions were completed in order to better understand how gesture usage over time affected participants' opinions with respect to gesture preferences, emotional comfort and sustainability of interest. During each session, participants were asked to perform repetitions of each gesture in both a private indoor and a public outdoor setting. These settings were chosen to reflect two significantly different locations with respect to gesture acceptance rates from the previous gesture survey. The indoor setting consisted of a private room in the university with a desk and chairs. For the outdoor setting, participants stood in a public location on the pavement near a bus stop and an underground station on a busy city street, as shown in Figure 3.15. After completing the experiment in each setting, each participant completed a semi-structured interview in the indoor setting. Gesture order and setting order were randomised for each participant.

During each session, participants used an application that prompted gesture performance and logged accelerometer data. The accelerometer data was used to determine the amount of energy, in the form of acceleration, used during each gesture repetition. This information was collected to better understand how objective measures of energy compared to user preferences. The application used for this study ran on a Nokia N95[1], a standard mobile phone. It utilised the internal accelerometer to log device-based gesture movements and an external SHAKE sensor pack[2] to log accelerometer data for the body-based gestures. The external sensor pack,

[1]More information: http://europe.nokia.com/support/product-support/nokia-n95
[2]More information: http://code.google.com/p/shake-drivers/

Figure 3.15: Outdoor setting for user study. Experimenter on left, participant on right on a busy city street.

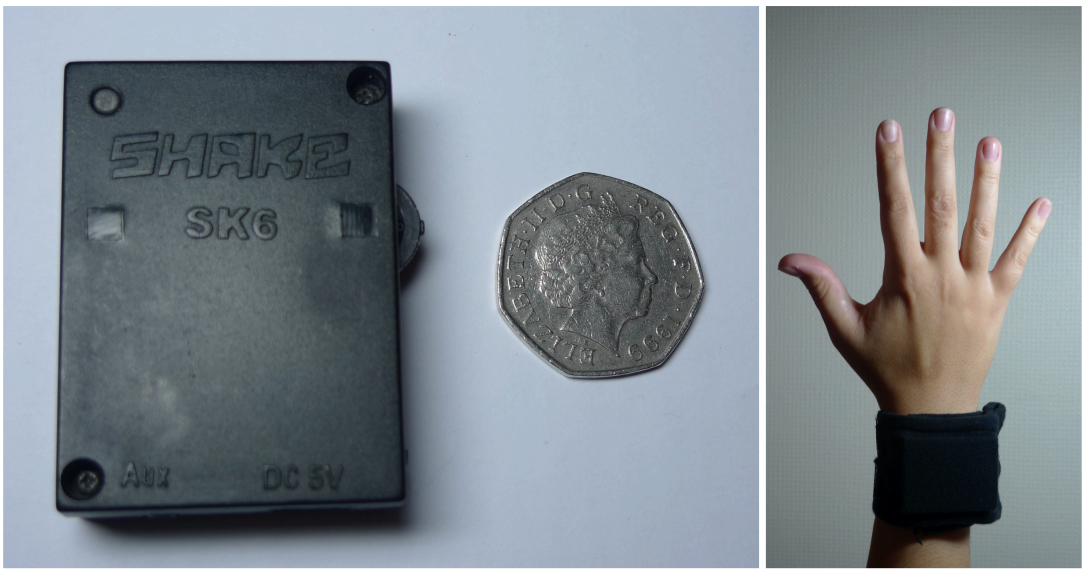

Figure 3.16: Left: SHAKE device shown with a fifty pence piece. Right: SHAKE worn on the wrist.

roughly the size of a matchbox, contains an accelerometer, magnetometer and gyroscope that can be accessed through a Bluetooth connection (Figure 3.16, left). The SHAKE was attached to the wrist for the wrist rotation and nose tapping gestures (Figure 3.16, right), attached to the shoe for the foot tapping gesture, and attached to headphones worn during the head nodding gesture.

For each gesture, the phone application played a short video demonstrating how the gesture should be performed. Once the video was complete, the name and description of the current gesture appeared on screen until the participant was ready to begin the gesture repetitions. The names and descriptions for the subset of gestures used in this on-the-street study were taken from the text used in the first survey study, as shown in Figure 3.1. Between repetitions, scrolling text was used to provide users with a short reading task, lasting roughly 10-15 seconds. Participants were told this was meant to appear as though they were reading text messages or checking email on the phone. Additionally, this gave participants a simple task to complete between repetitions so that gestures were not completed rapidly in quick succession. The application would prompt participants to perform a repetition of the current gesture by flashing a white screen with the gesture name. After completing five repetitions of the gesture, participants would continue on to the next gesture until all eight were completed. During the study, participants were told to perform the gesture based on their interpretation of the video and the textual descriptions. The experimenter offered no advice on 'correct' performance, always advising the participants to view the video or read the description if they had questions about a gesture.

3.3.2 Results

Throughout the course of the study, the results were collected in the form of accelerometer log files, observations, and semi-structured interviews. Eleven participants completed this study for a total of 33 participant hours. Participants were aged from 21 to 28, comprised of 4 females and 7 males.

Why Gesture? Preferences and Rationales

This study aimed to better understand why users might accept or reject gestures for use in everyday life. During the interviews following each user study session, participants described the reasons why they liked and disliked gestures. These results provide an important insight into how users might accept gesture-based interaction into their daily lives, listed in order from most to least commonly discussed.

Subtle Movement – Gestures that were small or unobtrusive such as phone tapping or foot tapping, were more comfortable for study participants. The ability to perform visually small gestures and hide performance was the most popular reason for finding a gesture acceptable. For example, one participant stated that "considering what we were doing, the most important part was that it wasn't drawing unnecessary attention to what we were doing, some of the other gestures were."

Looks or Feels Similar to Everyday Actions – Those gestures that were familiar in feeling or appearance were more accepted by study participants. This included actions that are literally like everyday movements, such as foot tapping, and actions that are metaphorically like everyday movements, such as shaking a phone. For example, one participant stated that "foot tapping looks very similar to what you do normally anyway. You'd probably be tapping your foot if you were listening to music or something." Another example was a gesture that *felt* like an everyday action, such as the shaking or rhythm gestures. One participant described these as "natural movements, they were things you'd do anyway, like shaking a juice bottle to mix it up." This result shows how everyday actions could make acceptable mobile phone interactions.

Similar to Existing Technology – Some of the gestures, such as screen tapping and rhythm, were described as similar to touchscreen and tilt-based interfaces. This ability to relate the performance to existing technology made these gestures more acceptable to use. For example, one participant stated that "tapping the phone is quite comfortable because we use touch phones anyway, it's more understandable." Even though the phones used in this study were not touch sensitive devices, tapping the screen was still an acceptable gesture. With respect to tilt-based interfaces, one participant stated that "I haven't used them [iPhone applications], I have seen people using them. I suppose for both the iPhone and things like the [Nintendo] Wii, they are not necessarily used on the phone, they're kind of, part of public consciousness of ways you can interact with something." Being able to relate interaction to existing technology in order to explain the interaction to spectators was an important aspect of acceptable performance.

Enjoyable Movement – Gestures that provided a high amount of satisfaction from the actual

feeling or appearance of the gesture itself were more acceptable and even desirable to perform. One participant described how enjoyable the rhythm gesture was, stating that "I don't know, I just fell for it." This participant went on later to describe the rhythm gesture as "quite funky, and it's quite cool, to dance or something, move your hand up and down... The business men and well off people, I don't know if they would be able to use all of those gestures, but cool people, they would, and they would be happy to use it." This result shows that an element of enjoyment or playfulness in a gesture can make it more acceptable and motivate users to perform gestures rather than hide them.

Participants also described gestures that they disliked and provided reasons for their opinions. The following reasons were reported for feeling uncomfortable about gestures, listed in order from most to least frequent.

Looks Weird or Attention Seeking – Gestures that required the participant to perform large or noticeable actions were the most commonly disliked gestures. For example, one participant stated that "the shoulder tap, nose tap, and head nodding were all quite attention seeking. I was actually aware that someone looked at me when I was tapping my nose, I thought that must seem a bit strange." The participants' perceptions about spectators undesirably taking notice of their actions made gestures much less comfortable to perform.

Physically Uncomfortable – Some gestures, including head nodding, foot tapping, and wrist rotation, required the participant to move their body in a way that was described as uncomfortable. For example, one participant stated that "the shoulder tap I did not like that one at all. That was uncomfortable. It was like the foot tapping as well, it was uncomfortable." Although it is clear that uncomfortable gestures are undesirable, it was surprising which gestures were described as uncomfortable.

Interferes with Communication – Participants disliked the head nodding gesture because it might be distracting or confusing to use during a conversation. One participant stated that "I think the head nodding is particularly weird because the head is the primary means of communicating with other people so people look at your head first of all." Gestures that might interfere with communication were uncomfortable because of undesirable reactions from spectators.

Uncommon Movement – Gestures that do not naturally occur in daily life, such as the shoulder tapping, were disliked. For example, one participant stated that "it [shoulder tap] was just a very unusual gesture, I can't imagine a context in which the shoulder would make any sense at all." These kinds of gestures were undesirable, however they are often the gestures used in interfaces because of the lowered risk of false positive gesturing occurring in daily life.

These motivations for accepting and rejecting gesture-based interaction provide some insight into the results of the previous survey. In many cases, participants focused on the visual aspects of gestures when determining the acceptability of a given gesture. Many of the reasons participants described for both liking and disliking gestures were based on the imagined reactions of spectators. For example, participants liked gestures that were subtle, not easily noticed by spectators, and did not draw attention the the performer. In other cases, participants liked gestures if the visible aspects of that gesture could be blended into everyday actions, such as

foot tapping.

Less frequently, participants cited personal aspects of the experience of performing as the reasons for liking or disliking gestures. For example, the enjoyment of performing gestures playfully was one of the few reasons for liking a gesture that focused on the personal experience but represents an important consideration. As these results show, the physical experience of performing can make interactions *fun* through playfulness, *cool* for early adopters, or *sneaky* for those trying to hide their actions.

Public and Private: How Audience Affects Gesture Usage

Following from the survey, this on-the-street study looked more specifically at *why* participants feel so different in public versus a private locations. During each interview, participants were questioned about their opinions of the indoor compared to the outdoor setting. While three participants reported that they felt indifferent about these settings, the remaining eight said the settings were very different. Of these eight, seven specifically reported that the outdoor setting was less comfortable than the indoor one.

Participants discussed their awareness of the audience around them in the outdoor setting, sometimes mentioning a general discomfort performing gestures outside, or mentioning specific instances where they became aware of others watching them. For example, one participant stated that "being inside is more comfortable, you don't have people wondering what you're doing." Another participant stated that "there was one time when somebody turned around and looked at me when I was tapping my nose." These experiences with spectators were important because participants gathered feedback from the reactions of spectators and adjusted their behaviour based on those reactions.

Participants discussed how the awareness of others around them affected the way they chose to perform the gestures. Because this user study only logged gesture performance without completing gesture recognition or providing feedback about 'correctness,' participants were able to change their performances depending on their own preferences or comfort. For example, one participant stated that "the nose tapping, when I was outside, I kind of disguised it as a scratch or pushing my glasses up." While some participants took advantage of the ability to disguise movements while outside, others actually changed their manner of performance by completing the gesture more quickly or with smaller movements. For example, one participant stated that "I would tend to do them a bit more quickly, just to get it done. Not to look like an idiot too much." These adjustments made users feel more comfortable, and represent a need for flexibility in gesture recognition in different contexts. Some participants reported that the outside location was more uncomfortable for specific gestures. For example, one participant stated that "it felt a bit more conspicuous outside, especially when you are doing the head nodding gesture." This demonstrates the importance of selecting appropriate gestures for different contexts.

Gesture Name	Survey Ranking	Study Ranking	Energy Measured
Screen Tapping	**1**	**1**	24.3
Shaking	**2**	**2**	93.5
Foot Tapping	**3**	5	196.0
Wrist Rotation	4	**3**	201.2
Rhythm	5	4	111.2
Shoulder Tap	**6**	**7**	67.5
Nose Tapping	**7**	**6**	74.0
Head Nodding	**8**	**8**	87.5

Figure 3.17: Gesture rankings from both the survey and the on-the-street study and energy measures. Energy units are arbitrary but consistent between measurements.

Energy and Gesture Ranking

Each gesture was given two rankings: one based on the previous survey results and another on the user study, ranging from 1 to 8, as shown in Figure 3.17. The survey ranking was determined using the average acceptance of the eight gestures based on audience. Rankings from the user study were determined by subtracting the number of negative mentions from the number of positive mentions for each gesture from interview transcripts. By comparing the top three and bottom three ranked gestures from each study, the results show that these are nearly identical, with the exception of rank 3, which was held by foot tapping in the survey and wrist rotation in the user study. These similarities show that the survey was effective in predicting highly acceptable and unacceptable gestures. Even though the survey revolved around imagined situations, respondents were able to provide accurate responses when shown videos and questions, as verified by the reported preferences when participants performed gestures in real settings in the on-the-street study.

An objective measure of energy was calculated using the accelerometer data for each gesture based on the derivative of acceleration over the time spent performing the gesture. The energy measurements from the on-the-street study are shown in Figure 3.17. Surprisingly, these results show that there is no correlation between energy and ranking, where gestures requiring high amounts of energy and movement could be more acceptable than those requiring little movement. The subtlety of a gesture was not necessarily a matter of energy required to complete it, but also depended on the visual appearance of the gesture, the perceived meaning, and the ability to perform the gesture in an acceptable manner. For example, screen tapping was the most popular gesture and had the lowest energy level, but foot tapping, the third most popular gesture, had the second highest energy. It is not the energy required to perform a gesture that makes it acceptable, but a variety of factors such as visibility, perceived meaning, or physical comfort.

Completing surveys of social acceptability provides useful insights that can be incorporated into early stages of design. For example, a relatively simple survey can rule out unacceptable

gestures before time is spent on the implementation of gestures that might never be used due to poor acceptance. Video prototypes can be created early in the design process to test a divergent set of interactions before any implementation has begun. This survey is best used as a preliminary evaluation because it can only provide limited ranking information, with more in depth qualitative studies required to better understand why a given interaction is liked or disliked.

Changes Over Time: Developing Gesture Preference

Multiple trials were completed in order to better understand how opinions of the gestures changed over time and which experiences influenced those opinions. After each session, participants were asked to compare their experiences to previous sessions. After the second session, nine of eleven participants stated that they felt the gestures were more acceptable than their first experience. Eight participants stated that they felt greater levels of comfort, and one reported greater confidence during the second session. Interestingly, participants described these greater levels of comfort when objective measures of energy indicated that gestures were performed in a similar manner across repeated trials. One participant reported that "I was more comfortable, now I'm used to doing gestures. I know what to expect." Participants felt more comfortable during the second session because they remembered positive experiences from the previous session. These experiences helped participants feel more confident that they would not become embarrassed during the subsequent trials. For example, one participant stated that they had "a greater familiarity with them [gestures], more comfortable doing them as well. It's not so new and unknown from last time. I know what it's like to touch my nose in public surroundings." When asked why the second session was different, another participant stated that "I think having the experience of actually performing the gestures. I knew it wouldn't cause me any undue concern, it was just a more comfortable experience overall." Participants performed an action, gathered feedback from the responses of others, and therefore had more information to work with in subsequent performances. After just one experience, participants were more comfortable using the gestures on a busy city street.

Gesture preferences also changed in the second session compared to the first session. One participant stated that there was a "clearer distinction between which [gestures] I liked best and which I disliked." Other participants stated that specific gestures became more acceptable. For example, one participant stated that "some of them, like the shoulder tapping or nose tapping, felt less awkward this time than they did last time. Maybe if I kept using them they wouldn't feel as strange." The continued exposure to these gestures allowed participants to develop clear preferences. After the third session, ten of eleven participants reported positive benefits from repeating the gestures, with all of these participants specifically reporting greater comfort in performing the gestures.

3.3.3 Discussion

This user study explored the initial reactions the participants developed when given the opportunity to perform gestures in different settings. Although the rationales discussed are

helpful in understanding more about acceptance and rejection, there is a significant overlap between these reasons where some gestures could be described as both acceptable and unacceptable. For example, the desire for enjoyable, playful movements is in direct opposition to the desire for subtle movements. With the enjoyable movements, there is an element of taking pleasure in the display of the gesture and a desire for others to notice. Similarly, the need for gestures to be familiar opposes the need for gestures not to interfere with communication. Many of the gestures that users might be familiar with are those that commonly occur within the context of speech, such as head nodding. A head nod directed towards a mobile interface could easily be mistaken by spectators as part of the current conversation and cause confusion. Designers should begin with a divergent set of possible interactions and complete an initial evaluation based on the criteria here. In this way, these reactions and rationales provide designers with a basic criteria for evaluating gestures early in the design process.

The rationals described above might help designers choose more acceptable gestures, but often present challenges for implementation. For example, participants discussed how subtle movements were much more acceptable but these movements are inherently harder to detect using both inertial and visual gesture recognition. Another issue described by participants was the poor acceptability of actions that do not occur in everyday life. However, from an implementation perspective, these are highly desirable actions because of the reduced risk of accidentally performing the gesture. These kinds of gestures are often used when dealing with the segmentation issue (Strachan *et al.*, 2007): knowing when intentional gesturing begins and ends. This is often dealt with by using a gating gesture, which is usually an uncommon movement in order to avoid false positive recognitions. These issues highlight considerations designers must make to create a system that is both acceptable and robust.

Based on the results of this user study, two design considerations are recommended when completing user studies of social acceptability. Because differences were observed between the first and second trials, this thesis recommends completing at least two trials when evaluating for social acceptability because users will develop preferences and change their acceptance rates after multiple trials. Additionally, this thesis suggests that such trials be completed in real world settings, rather than lab ones. Study participants based the vast majority of their opinions and feelings about the gestures based on the spectators and experience of the outdoor setting. Gathering the experiences of performing gestures in real settings allowed participants to report more in-depth opinions of the gestures, using real social experiences to explain their opinions. Completing multiple trials in real world settings provided insights into experience and acceptability that would not have been possible in a single trial lab evaluation.

This study was designed to look at gestures specifically outside of the context of an actual application or activity. This was beneficial to the results because feedback provided by participants was not influenced by factors such as gesture appropriateness in the application or usefulness of the application. For example, if the study had used a web browsing application, the results could be influenced by participants' familiarity with mobile browsing, their desire to use browsing on the go, and the appropriateness of the gesture to the application. These issues represent a very different problem than simple acceptance of a gesture, but rather look at *adoption* of the gesture within the greater context of an application. Although these issues are certainly important, and will be investigated in great detail in the next chapter, this study

was looking specifically at acceptance and rejection. One problem with this decision is that these results did not look at how motivation can affect social acceptability. For example, if a phone starts ringing at an undesirable moment such as during a meeting, would users be more willing to do a less acceptable gesture if that guaranteed the phone would immediately go to silence? The role of motivation is an important aspect to adoption, but was not investigated in these studies of acceptance.

The design of this user study successfully tackled several key concerns about initial acceptance of gesture-based interaction. Participants had the opportunity to perform gestures in both a public and a private setting. Although the public setting was relatively controlled given that the experimenter was always present and the participant went through a highly formalised series of actions, it still gave users a real world experience that would not have been possible in the lab. Completing even highly controlled studies in real world settings allows users to experience performance with the real and often surprising reactions from spectators. However, one problem with these highly controlled use studies is that the presence of the experimenter influences the participant. For example, the participant might feel obligated to perform whatever actions required for the study even if they personally find them embarrassing. Participants might be more confident performing these actions because the presence of the experimenter makes clear to spectators that the participant's actions are part of some study. This is a difficult issue to deal with because the experimenter must often be present for ethical reasons and to monitor the progress of the study. Participants were also given multiple chances to experience the gestures, and their opinions changed over time as they gathered more experience with the gestures and performing in public.

3.4 Becoming Multimodal: Reactions to Gesture and Voice Interfaces

The second survey completed as part of understanding acceptance of new interaction techniques extended the first survey in three ways. First, this survey included both gesture and voice-based interaction. Second, this survey organised gesture and voice commands into more fine grained categories in order to look in more detail at some of the aspects of these interactions. Finally, this survey used more specific contexts by presenting combinations of locations and audiences rather than examining them separately. These changes to the survey allow for a greater understanding of the relationship between location and audience and apply this method to a broader range of interaction techniques.

3.4.1 The Survey

This survey adopted a more detailed organisation of categories based on observable features of the interaction techniques for gestures and speech than the previous survey. With respect to the gestures, four categories were used: emblematic (Kendon, 1997), device-based, arbitrary (Nespoulous *et al.*, 1986), and body-based. Emblematic gestures refer to those gestures that have a widely accepted meaning outside of the context of speech, usually within a given cul-

Gesture	Category	Voice	Category
Peace Sign	Emblematic	Say "Close"	Command
Shrugging	Emblematic	Say "Lock"	Command
Device Stroke	Device-Based	Say "I'm Fine"	Speech
Device Flick	Device-Based	Say "That's Nice"	Speech
Hook Finger	Arbitrary	Buzzing	Non-speech
Open Palm	Arbitrary	Doo Doo Doo	Non-speech
Wrist Rotation	Body-Based	Clicking	Non-speech
Head Nodding	Body-Based	Popping	Non-speech

Figure 3.18: Gestures and Voice commands used in the survey organised by category.

ture (Kendon, 1997). For example, a 'thumbs up' gesture can easily be interpreted to mean 'good' without any accompanying speech. Arbitrary gestures are gestures or hand positions that to not have a clear meaning outside of the context of speech (Nespoulous *et al.*, 1986). These may be gestures that have multiple interpretations that could be disambiguated in the context of speech or simple hand positions with meanings that must be learned. The device-based gestures are those that involve directly manipulate a device, such as shaking the device. Body-based gestures are movements of the body that do not directly involve manipulating a device, although external sensors might be used to sense these gestures. Between these categories, there will be some amount of overlap where gestures may belong to multiple categories. In these cases, gestures have been put in the category with the best fit possible. With respect to the voice commands, three categories were used: command, speech, and non-speech. Command inputs include one word commands that related to phone tasks, such as 'call' or 'lock'. Speech inputs included short, commonly said phrases. Non-speech inputs include a variety of vocalisations and sounds, some of which occur normally in everyday life, such as whistling, and some of which do not like buzzing or popping. From each command category, two gestures or voice commands were selected, with the exception of non-speech sounds where four were selected, to make a set of sixteen commands for use in this survey. These are shown in Figure 3.18.

Similar to the previous survey, this study used video prototypes and multiple choice questions. A screenshot of the gesture video and survey question layout can be seen in Figure 3.19. These videos portrayed a male actor in a plain setting where respondents were asked to imagine the contexts where they might perform the given interaction techniques. For each gesture or voice command, the survey webpage displayed the title of the video and a short textual description of the current interaction technique while the video was played on a loop below. For each video, the survey respondents were asked to answer the following six questions:

How acceptable do you feel it would be to perform this command:

- while on the sidewalk or pavement with strangers;

Participant ID: 1093

Device Flicking

Whip-like movement of hand while holding a mobile device.

This is video 1 of 16.

Please rate the statements below based on how acceptable you feel it would be to perform this command in the given scenario.

	Totally Unacceptable	Unacceptable	Slightly Unacceptable	Neutral	Slightly Acceptable	Acceptable	Perfectly Acceptable
Performing this command while on the pavement or sidewalk with strangers?	○	○	○	○	○	○	○
Performing this command while on the pavement or sidewalk with friends?	○	○	○	○	○	○	○
Performing this command while at home with family?	○	○	○	○	○	○	○
Performing this command while at home alone?	○	○	○	○	○	○	○
Performing this command as a passenger on a bus or train with friends?	○	○	○	○	○	○	○
Performing this command as a passenger on a bus or train alone?	○	○	○	○	○	○	○

Comments
Please add any comments you have about the command or your responses here.

Submit Answers

Figure 3.19: Screenshot of the survey video and questions.

- while on the sidewalk or pavement with friends;
- while at home with family;
- while at home alone;
- as a passenger on a bus or train with strangers;
- and as a passenger on a bus or train with friends.

For each question, survey respondents were asked to select a choice from a 7 point Likert-type scale based on levels of acceptability. Respondents could also leave comments and feedback in a text field. These questions were chosen to cover a set of common places where mobile phones are used and look at those settings when there are different spectators present. The full survey can be seen in Appendix A.2 and the video prototypes can be seen in Appendix B.

3.4.2 Results

The survey included responses from 65 respondents, ranging in age from 17 to 72. 57% of respondents were under the age of 30 and 77% under the age of 40. Of the respondents, 55% were male and 45% were female. The majority of participants, at 72%, were currently living in Britain. The survey lasted for a total of two weeks, with participants being recruited through email and social networking websites.

Gesture and Voice: Modality Preferences and Rankings

The results of this survey showed how voice-based commands compared to gesture-based commands with respect to social acceptability in different contexts. This survey also showed how each gesture and voice command category ranked relative to each other when survey responses were grouped by category. Figure 3.20 shows the responses grouped by command category for both gesture and voice categories when performed on the pavement with strangers. This comparison shows how different categories of commands ranked with respect to acceptability.

Figure 3.21 shows pair-wise comparisons between interaction categories using the Wilcoxon Signed Rank test, with the effect sizes shown in Figure 3.22. These results show a clear preference for gesture-based interaction over voice-based interaction. Device-based gestures were significantly more acceptable than every other interaction category, and non-speech interactions were significantly less acceptable than every other category except the speech category. For device and non-speech interactions, the effect size for these significant differences was large for every comparison except the non-significant difference between non-speech and speech. Command-based interaction was the highest rated voice category, being significantly less acceptable than device and body-based gesturing. Speech-based interaction was significantly less acceptable than device, body, and emblematic gestures and non-speech interaction was significantly less acceptable than all gesture interaction categories.

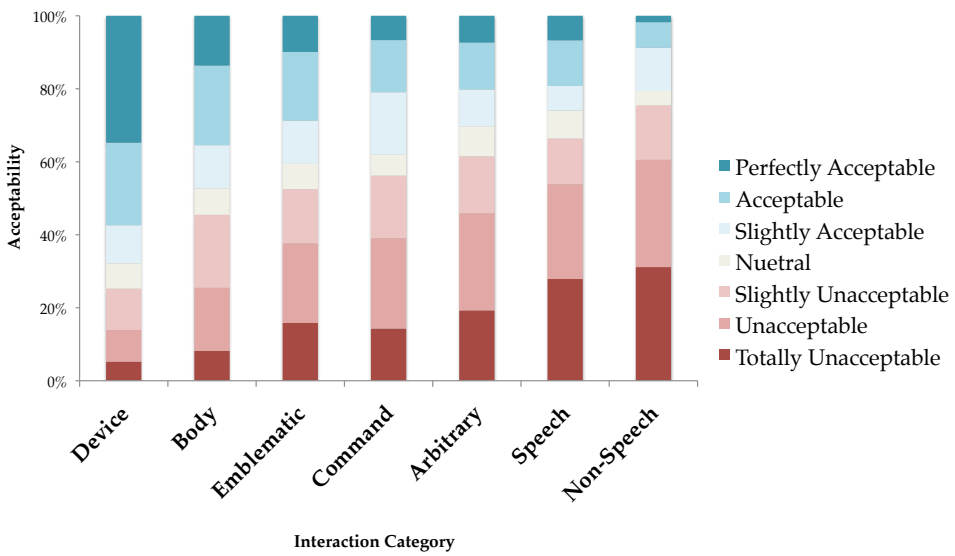

Figure 3.20: Survey gesture and voice commands grouped by category while performed on the pavement in front of strangers.

	Device	Body	Emblematic	Arbitrary	Command	Speech	Non-Speech
Device		**0.0003**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Body			0.003	**< 0.0001**	**< 0.0001**	**< 0.0001**	**< 0.0001**
Emblematic				0.01	0.08	**0.0007**	**< 0.0001**
Arbitrary					0.03	0.07	**0.0003**
Command						0.004	**< 0.0001**
Speech							0.07
Non-Speech							

Figure 3.21: Pairwise comparisons for significant differences between categories while performed on the pavement in front of strangers. Values of 0.002 and below are significant adjusting for Bonferroni's Correction.

	Device	Body	Emblematic	Arbitrary	Command	Speech	Non-Speech
Device		**0.5**	**0.7**	**0.7**	**0.7**	**0.7**	**0.8**
Body			--	**0.6**	**0.6**	**0.6**	**0.8**
Emblematic				--	--	**0.4**	**0.7**
Arbitrary					--	--	**0.5**
Command						--	**0.6**
Speech							--
Non-Speech							

Figure 3.22: Effect sizes for pair-wise comparisons. Values 0.5 and over are considered large effect sizes and values 0.3 and above are considered medium effect sizes. Effect sizes for non-significant comparisons are not reported.

Category	p Value	Effect Size
Home Alone vs. with Family	< 0.0001	0.8
Pavement Alone vs. with Strangers	< 0.0001	0.7
Bus Alone vs. with Strangers	0.0007	0.6

Figure 3.23: Comparison of audience types for different locations with p values and effect sizes. Effect sizes 0.5 and greater are considered large.

Familiar Audiences

The results of this survey show the effects that familiar versus unfamiliar audiences have on the acceptability of multimodal interaction techniques in different locations. Using the non-parametric Wilcoxon Signed Rank Test for correlated samples, the results show that the performance of multimodal interaction was significantly more acceptable in front of familiar audiences as compared to unfamiliar audiences for each location investigated in the study. Figure 3.23 shows the p values and effect sizes for these tests.

When performing in front of friends, even in a highly public location such as a busy pavement, acceptability increased significantly as compared to performing in front of strangers. For example, arbitrary gestures increase in acceptability by 16% when performed in the presence of friends on a busy sidewalk or pavement. These gestures, which have a relatively low acceptance rate of just 30% on the pavement with strangers, do not provide clear visual cues to spectators and therefore are more uncomfortable to perform in front of unfamiliar audiences. Even gestures with relatively high acceptance rates, such as device-based gestures at 68% acceptance on the pavement with strangers, benefited from being performed in front of familiar audiences, increasing in acceptability by 11%. These results demonstrate more clearly how the audience affects social acceptability.

An interesting aspect of these results can be seen in the effect size for acceptability on the bus with friends versus with strangers. Although 0.6 is still considered a large effect size, this category has the smallest effect size of the three locations. One explanation of this is that on a bus or train, the audience is sustained rather than transitory. This means that audience members can continue watching a user for the entire journey, making users more self conscious of their own performances even when with friends. This distinction between sustained and transitory spectators illustrates another aspect of audiences that affects social acceptability and acceptance.

3.4.3 Discussion

The results of this survey show how locations and audience affect social acceptability, looking at specific combinations of locations with both familiar and unfamiliar spectators. Even in highly public settings, the presence of familiar spectators, for example performing an action on the pavement with strangers as compared to being on the pavement with friends, increases

the acceptability of performing multimodal interactions. Although a familiar audience represents one of the factors of location and context that affect social acceptability, familiarity with the location itself may also affect acceptability. For example, a commonly walked route to the grocery store might be considered familiar and therefore a more acceptable location for multimodal interaction than another route that is in an unfamiliar area of town.

This survey also compares the relative acceptability between using gestures or voice-based commands. Overall, gesture-based interactions had higher levels of acceptability than voice-based interactions. This survey does not reveal *why* gestures might be more acceptable than voice-based interactions, but design of interactions with respect to the demonstration of the purpose and meaning of voice-based commands could increase acceptability. For example, the videos used in this study did not make use of a device as part of the voice-based interaction but the acceptability of these interactions might be increased if the actor wore a Bluetooth headset or was portrayed talking into a phone. Considering how voice-based interaction could be demonstrated as part of an interface could significantly improve the acceptability of voice-based interactions in public spaces.

The results show that familiar audiences make performing gestures or voice commands significantly more acceptable. As a way of encouraging interaction, the presence of familiar spectators could become an important consideration for design. For example, interfaces that are designed to be used socially and collaboratively, such as the *Social Gravity* system (Williamson *et al.*, 2010), could have increased social acceptability because users are performing together. These kinds of multimodal systems could be designed such that they support both individual and collaborative interaction, so that familiar audiences are not just spectators but can become performers as well. This kind of interaction, where the user changes roles fluidly between performer and spectator (Leong *et al.*, 2011), might not only provide an engaging experience but also increase the acceptability of new interactions.

3.5 Conclusions

This chapter explores the initial acceptance of multimodal interaction techniques through two surveys and a user study. These studies focus on location and audience as significant factors that influence social acceptability. The first survey evaluated a set of eighteen gestures with multiple choice questions that looked at location and audience independently. This was followed by an on-the-street user study where participants were asked to perform eight gestures in both an indoor private setting and an outdoor public setting over three sessions. Finally, a second survey was completed that looked at gestures and voice-based interactions, building on the first survey by using questions that combined location and audience. The results of these studies demonstrate some of the factors that influence acceptance and propose recommendations for the evaluation of multimodal interaction acceptability.

The first survey demonstrates how respondents ranked the acceptability of performing different types of gestures in different locations or with different audiences. This survey also demonstrates the significantly higher acceptability of device-based gestures, indicating that the survey respondents felt that the visibility of a device made those gestures more acceptable.

Demonstrating the purpose of interaction for spectators not only increases social acceptability but also contributes to the spectator experience. The on-the-street study presents participants' rationales for liking and disliking gestures. Participants in the on-the-street study described the desirable visual aspects of the gestures as playful, cool, or embarrassing aspects of interactions, and how gestures could be hidden as everyday actions. The second survey looked at both gesture and voice-based interaction, with survey questions that focused on the effect audience has on acceptability. The presence of familiar spectators made interaction significantly more acceptable in the locations evaluated in this survey. This indicates that performative interaction could be made more acceptable by interfaces that support collaborative or social interaction. Each of these studies presents some of the initial reactions these users had when evaluating these interaction techniques.

These results provide researchers with concrete tools that can be used to evaluate social acceptability early in the development process. The first survey, which utilised video prototypes and basic multiple choice questions, was validated by the on-the-street user study that involved participants actually performing gestures in different locations. Even though the first survey responses were based on imagined situations, respondents were able to provide ranking data comparable to that from the qualitative on-the-street user study. The ranking data collected in the gesture survey could be used to rule out unacceptable interaction techniques *before* implementation and save valuable time and effort. The on-the-street user study demonstrates two important techniques for completing user evaluations of social acceptability. Multiple exposures to a new interaction technique are important when determining social acceptability. After one good or bad experience, participants' opinions of social acceptability began to stabilise. This study also shows how important realistic experiences are when collecting qualitative data about experience. Participants often described their experiences and opinions based on the public outdoor location, which would not have been possible without completing the study in a realistic setting.

The results of these three studies present two clear trends. Firstly, that users are highly concerned about the visual aspects of performing in public and take this into account when determining social acceptability. Secondly, that users are concerned about the best possible way of using these 'visual aspects' to demonstrate the purpose of their performance and give some meaning to the actions. The next chapter will explore these issues in more detail, examining how users place multimodal interaction techniques into context and how users develop a multimodal performance *in situ*.

Chapter 4

Exploration and Decision Making in Interface Adoption

4.1 Introduction

The *adoption* of technology, for the purpose of this thesis, can be understood as usage going beyond simple *acceptance*, where the purpose of interaction begins to play a role in how individuals evaluate that technology. Where the previous chapter looked at the *acceptance* of multimodal interaction outside of a usage context, this chapter looks at the *adoption* of multimodal interaction techniques by placing these interactions into real world scenarios and allowing users make decisions about how the interactions should work. This decision process, where users are examining how interactions should map to activities and applications or how interactions should be performed and interpreted, is the focus of the studies presented in this chapter.

While much of research revolving around technology adoption looks at adoption on larger social scales such as the technology adoption life cycle and diffusions of innovation, as described by Rogers (2003), this chapter explores adoption on an individual scale by looking at early adoption. With respect to multimodal interfaces, this thesis argues that it is important to evaluate how well an interaction fits with an application. A user of a system should be able to easily make sense of how an interaction works and have a good cognitive model of the relationship between actions and results in a multimodal interface. For example, a loud or disruptive input technique might be a poor mapping to the task of silencing a mobile phone. This 'appropriate fit' should make sense both from a usability perspective and an experience perspective, where usability refers to the objective measures of accuracy or speed and experience refers to the *perceived* usefulness and appropriateness of the interaction. Additionally, because many multimodal interactions require users to adopt highly visual or audible behaviours, the performativity of these actions plays an important role in adoption.

This chapter explores the adoption of multimodal interfaces through two studies. The first study looks specifically at performativity in multimodal interfaces where pairs of participants

were required to create their own multimodal performance *in situ* in different locations. The results of this study show the kinds of performances participants generated, their motivations for performing differently in different locations, and how they imagined their fellow participant performing. The second study looked more specifically at *how* users apply multimodal interaction to different usage scenarios in a focus group setting. This study also looked at age as a factor affecting interface adoption. Together, these studies explore how and why users choose to adopt new multimodal interaction techniques when placing them within a specific application area in real world contexts.

4.2 Exploring Multimodal Performance in Real World Settings

The purpose of this study was to investigate how participants might choose to perform multimodal interactions in real world settings, examine the social acceptability of that performance, and understand more about the user experience of performing within an application context. To address these concerns, participants were required to generate multimodal input *in situ* in public and private locations using a mobile remote awareness application with a partner over repeated trials. However, although the application in this study was based around remote awareness, the purpose of this application was not concerned with the meaning or intention behind communications. This application was designed to support divergent multimodal inputs, create the experience of performing in different settings and participate as a distant audience member for a familiar other's performances.

This application, called MuMo, included a visualisation of a virtual fish tank where each user was represented by a fish in the tank that could be controlled using multimodal input. Users generated input by interacting with a small prop embedded with sensors. MuMo was designed to explore the issues of performance and the usage of props when the user was performing for two different audiences: one audience was the fellow participant watching the performance through the fish tank visualisation and the other was the immediately co-located spectators watching the live performance without necessarily being aware of its purpose or the interface itself. This application used highly flexible input methods, where participants were required to create their own performance style in real world locations using gesture and voice. Using this application, users were free to create a variety of performances to suit their current context and could participate as an audience member by watching the visualisation, where divergent imagined interpretations of the visualisation were possible. The possibility of this kind of extravagant performance (Jones, 2011) creates the opportunity for expression and imagination in real world contexts.

4.2.1 The MuMo Application

In the MuMo application, participants were each represented by a fish in a virtual fish tank, as shown in Figure 4.1. This visualisation could be seen as an animated background on each users' mobile phone and controlled using multimodal input. The application used a server/client architecture where each client updated the server with its current input values

Figure 4.1: Screenshots of MuMo application as an active wallpaper. Left shows fish tank visualisation as wallpaper, right shows visualisation with phone widgets.

and pulled updates from every other user from the server roughly once per second. Thus, users could see the effects of their own actions in the visualisation alongside those of their fellow participants. Participants were told they could use gestures or motions to make their fish swim faster or use audio and voice input to make their fish blow more bubbles. In each case, the fish behaviour was based solely on the magnitude of input, although this was not explained to the participants. For audio input, the louder the sound level the more bubbles the corresponding fish would create. Thus, participants could perform any kind of speech or sound-based action and see the result in the fish tank. Similarly, changes in swimming movements were based on the magnitude of acceleration of the gesture performed. This type of sensing was designed specifically to support both extravagant and subtle input, meaningful and abstract input, or simply environmental input that could be reflected in the fish tank visualisation in real time. This flexible style of input afforded unconstrained interaction in order to encourage participants to generate creative methods of controlling the visualisation. This also allowed for imaginative interpretations for those watching the visualisation since the observed output in the visualisation could be generated in a variety of ways.

The interface was controlled using the SHAKE sensor pack[1] to collect accelerometer data with an added microphone as shown in Figure 4.2. This was then embedded into the various objects or props shown in Figure 4.3. These props were chosen to provide a variety of objects that could facilitate performance or demonstrate interaction in different ways. These included playful objects, an abstract object, an everyday object, and an object that displayed the bare electronics of the sensors. The playful objects included two plush toys and one solid toy in order to allow for enjoyable and playful interactions. The abstract object was a hollow red mould that would simply act to conceal the sensors. The everyday object was a book with a space hollowed out to conceal the sensors in order to disguise the interactive prop. The final prop was a clear glass jar that exposed the bare electronics of the system as a method for demonstrat-

[1]More information: http://code.google.com/p/shake-drivers/

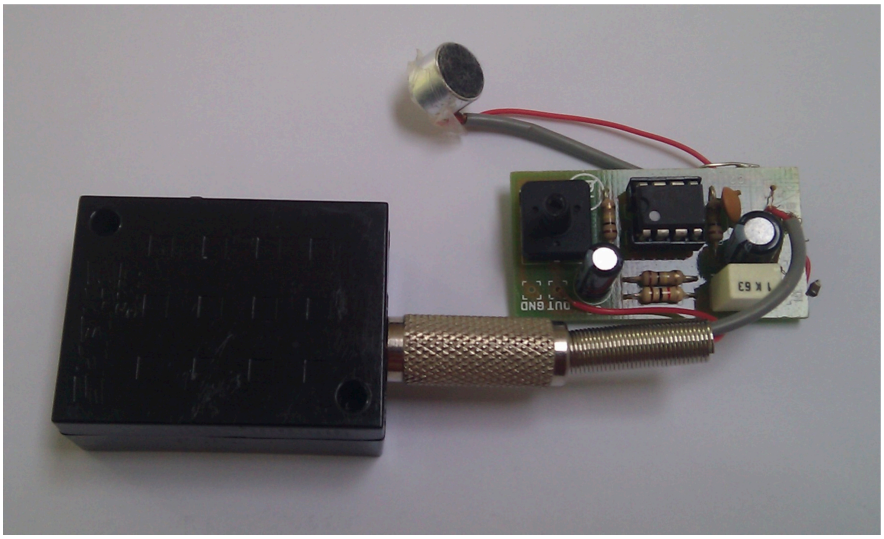

Figure 4.2: The SHAKE sensing device with an added external microphone.

ing the interactive purpose of the prop. These props were selected to provide different visual or cognitive clues for spectators about the performance in order to give performers different methods of exaggerating, disguising, or explaining their performance.

4.2.2 The Study

Participants were recruited in pairs, where each pair completed two usage sessions spaced about one week apart. These sessions were repeated to give participants multiple chances to interact with the system and develop performance preferences based on multiple experiences. Before each session, participants were told only that they could control their fish's swimming behaviour using gestures and the bubbling behaviour using sound and were given a chance to briefly experiment with the system. Then, the session began with the first participant being taken to a public location, a busy pavement, while the second remained in a private indoor location. Once both were ready to begin, the first participant was asked to complete three performance tasks, such as creating more bubbles, while the second was asked to interpret the first participant's actions by watching the visualisation on the phone. After these performance tasks were complete, the first participant was then asked to interpret the other's actions while the second participant completed three performance tasks. The participants would then switch locations and the tasks were repeated. Each task lasted two minutes. This study design allowed participants to perform actions in both the public and the private setting as well imagine how their partner would perform actions in both settings. Once both participants had completed their tasks in each location, they were interviewed together about their experiences.

4.2.3 Results

The study involved eight participants recruited in pairs. The pairs included two couples and two pairs of friends, with four females and four males. The participants ranged in age from

Figure 4.3: Participants could select one of six objects containing an embedded sensor pack to control their fish in the tank.

20 to 28. The results focus on an in depth analysis of a relatively small user group in order to gain a highly personal qualitative insight into the use and experience of this application. These results are based on the observation of the participants, recorded observations provided by the participants and transcripts of the interviews.

Creating Performances

Given that participants were allowed to create open ended performances using gesture and speech, it is not surprising there were a wide variety of styles and actions that resulted in the different locations where this study was completed.

Performative Actions – Even though the sensors were contained solely within the prop, performances were not limited to interactions with that prop and often involved additional interactions purely as an enhancement to the experience and appearance of performing. For example, one participant chose to sing to the prop for voice input in the private indoor setting. Although this was an unnecessarily extravagant interaction, this participant found this performance enjoyable and amusing, especially when his partner imagined this performance. Another participant performed swimming motions with both hands while outdoors. Even though the prop would only sense the movement of one arm in this case, the participant still enjoyed performing with both arms. In these cases, the experience of performance was augmented with either playful or meaningful actions being performed for non-functional purposes.

Hidden/Subtle Actions – Participants found ways of performing input that was subtle or hidden from passersby while still giving their fellow participant highly visible actions on the visualisation. This included actions such as tapping the prop to make noise, fidgeting with the prop in hand, and using environmental noise to create input. For example, one participant

chose to use the music of an outdoor performer as the input for their performance when audio was needed. These types of actions allowed participants to create meaningful input to the system without performing highly visual actions.

Functional Actions – In some cases, participants chose only to perform actions that completed the task without adding any additional performance or play. For example, participants would simply shake or wave the sensor to create gestures or say things like "I'm creating test speech for a system" or "I'm talking into the sensor now to see if something happens." In this case, participants did not try to actively hide or disguise their performance, but instead tried to demonstrate the purpose of the performance clearly by using "test speech" or rigid, purposeful actions.

These different styles of performance were influenced as much by location as personality. For example, one participant performed purely functional actions while outside and highly performative actions inside. Another participant completed highly performative actions both inside and outside. Yet another participant completed hidden or subtle actions both inside and outside. Because the interface supported a variety of actions, participants were able to change their performance style as needed in order to continue participating and feel comfortable about interaction. These decisions varied between participants, depending on personal preferences and personality. These factors represent an interesting influence on social acceptability that needs further exploration.

Imagining Others

Because the MuMo application required participants to create their own input, fellow participants watching the interface could not be sure what kinds of actions their partner was performing given the current output. Participants had to imagine how they thought their partner might be performing based on what they could see in the visualisation and their knowledge about their partner's current social context. This was both a positive and a negative aspect of this application, where some participants found it difficult to attach meaning to the interface while others enjoyed the process of imagining their fellow participant performing highly energetic, silly, or emotional behaviours. These imaginings not only contributed to the spectator experience of this application through the visualisation but also provided motivation for performance.

For those participants that enjoyed imagining their partner performing through the interface, participants allowed and encouraged their partner to imagine highly divergent performances, even when this was not realistic. For example, some participants imagined their partners singing or dancing as input for the visualisation even though their partner was in the outdoor setting. Even though this was relatively unrealistic, participants were able to suspend their disbelief and enjoyed imagining these kinds of actions anyway. These creative imaginings occurred both when pairs of participants used highly visible, performative interactions and when pairs of participants used the most subtle and discreet methods of interaction. For example, one participant imagined her partner "singing a relaxing song" and "jumping with it [turtle] on one leg." These interpretations were recorded even though both participants used extremely subtle actions for input, such as microphone tapping. Participants enjoyed imagin-

ing these playful actions, even if they did not perform these kinds of actions themselves.

Props and Performance

During each of the two sessions, participants could select an object of their choice as their prop. The turtle object was chosen eight times, the dolphin was chosen five times, the book, jar, and owl were chosen once and the red mould was never chosen. When discussing their choices of these objects, participants described how the objects worked and failed as props.

Props as Toys – The most commonly picked objects were the turtle and dolphin plush toys. Participants favoured these props for their playful nature and their ability to relate to the lively and lighthearted application. These props were often used in a playful manner, even though participants knew that these kinds of actions would not provide any additional input to the application. For example, participants would move the fins of the turtle or cover it's eyes as part of their performance even though this did not generate additional effects.

Props as Pairs – Participants often chose their props based on their partner even though they knew the props would not be used together. Choosing props together allowed participants to better understand what kinds of performances their partner might complete and also provided a better connection between partners. For example, one participant stated that "first I wanted to pick the glass jar, but when I saw he picked a toy I wanted to pick a toy as well." Another participant stated that "I picked the dolphin because you picked a toy, so it's two soft toys. Otherwise, I would've picked the book."

Props as Everyday Objects – Although some objects, such as the book, represented common objects one might normally carry around, participants felt less comfortable using these objects when interacting with the application. While using the book as a prop, one participant stated that "when I was inside I sang a song, I just made it up. But when I was outside I tried to talk very quietly. It wasn't as normal as I thought it would be." When discussing other everyday objects that might be used as props, one participant stated that "you might put it [sensors] into an object that you walk around with, like a coffee cup, but you wouldn't talk into a coffee cup." Although these props might disguise or hide sensors effectively, they make poor interactive objects when it comes to performance.

Participants also discussed the benefits of different props with respect to physical attributes like size or texture. For example, when describing why the dolphin was a useful prop, one participant stated that "it's easier to hold than one of the hard objects, nicer to hold." When describing objects that would make the most desirable props, participants stressed the importance of using soft or flexible objects. The ability to manipulate the props and the comfort of holding a soft object made them easier to use. Participants also described the benefits of using different props to conceal the sensors. When describing why a prop would be better than simply holding the sensor pack, one participant stated that "it's bigger, so there's more you can do with it." Participants also described how the prop makes performance more comfortable. For example, one participant stated that "it was much easier to just wave around the turtle than it would've been to wave a bunch of sensors". Other participants would have preferred a more anonymous object. When discussing negative aspects of using props, one participant

stated that "it made me more conscious of it, holding the object. If I just had the sensor in my hand people might not have noticed what I was doing." Because this application clearly had a playful nature, participants often chose props that encouraged this playfulness. However, props that are more abstract or anonymous were still desirable and in a different application area might have been more popular.

4.2.4 Discussion

This study provides some interesting insights into the ways in which these participants created performances in the wild, used props to enhance their interactions and demonstrated their intentions to co-located spectators. By performing through an interface, participants were also performing for their fellow participant watching that performance through the visualisation. Thus, participants in this study were constantly performing for two audiences, or spectator types, and had to balance the needs and expectations of these spectators simultaneously. For example, participants had to balance their desire to generate energetic or amusing input for their fellow participant with their desire to perform socially acceptable interactions in public places. In some cases, this meant that participants chose to limit their performance and the resulting output of the system, limiting the spectator experience for their fellow participant. In other cases, participants found ways of performing that were both comfortable for themselves and created ample output in the application for their fellow participant to enjoy.

Because this application required only basic actions but also supported extravagant ones, participants took full advantage of this flexibility and adapted their performances continually. The types of performances created were highly dependant on the location of the performance, with participants actively making decisions about their adoption of different performances based on their current location. In general, participants were more likely to perform highly visible or noticeable actions in the indoor location as compared to the outdoor location, which is in line with the results discussed in the previous chapter. Additionally, participants often adjusted their performances to match their fellow participant. Because the first session ended with an interview, participants learned what kind of actions their fellow participants had imagined them doing and what actions their fellow participant had actually performed during the first session. This was reflected in the second session where pairs of participants performed actions that were discussed during the first session. This included actions that might be amusing to their fellow participant or actions they thought the other participant might be performing as well. This demonstrates how social influence can affect adoption, even though this example is on a very small scale. For example, usage over time might allow constantly evolving practices and behaviours as the users of the application respond to each other and learn how to interpret the visualisation based on their knowledge of each other. Interpretations that come out of familiarity and extended use of an ambient display have been seen before (Brewer *et al.*, 2007), and certainly this emerging behaviour is an important aspect of these types of applications.

Participants' awareness of their partner watching the visualisation provided motivation for participants to perform amusing actions but also led participants to perform extremely subtle actions and simply allow or encourage their fellow participant to imagine more entertaining

actions. Pairs of participants had varying degrees of enjoyment imagining the performance of their fellow participant, with the two couple pairs being the most imaginative. Even when both participants performed subtle actions in the outdoor settings, both participants enjoyed imagining amusing performances. Although these imaginings provided some motivation to perform amusing actions, these participants were still highly aware of the co-located spectators, or passersby. In some cases, participants modified their performance when outside. For example, one participant used singing input while inside and conversational speech while outside. Both of these actions generated similar output in the visualisation, but participants used these different kinds of actions in order to maintain their comfort, experience, and enjoyment of the application. These adjustments show how considerations for both audiences must be balanced while using this application in public contexts.

This user study focused on how participants explored and adopted performances in real world settings as the interaction unfolded. This kind of explorative *in situ* user study allows for a highly qualitative look at adoption but allows for less control of the experience itself or the exploration of specific interaction techniques. The following user study looks at a specific set of interaction techniques in a focus group setting where participants were asked to imagine how they might use these interaction techniques in real world settings. Although the results of the focus group study are based on *imagined* situations rather than real ones, the previous chapter demonstrated that imagined scenarios are effective for evaluations of social acceptability. The following user study builds on the previous chapter by going beyond simple *acceptance* and asking participants to imagine how they would actually place interactions into a usage context and *adopt* these interactions in their everyday lives.

4.3 Experience Prototyping for Evaluations of Social Acceptability

The following study investigates technology adoption through a focus group study using prototypes. These focus groups explored the acceptance and adoption of a large set of interaction techniques by combining survey-like responses about acceptance with in-depth discussions about adoption. Focus group participants were asked to provide their initial reactions by accepting or rejecting and ranking of the techniques, discussing which they liked or disliked, and describing how they would use the interaction techniques in real world scenarios. The main goal of these focus groups was to explore acceptance and adoption of multimodal interaction techniques using a variety of qualitative and quantitative methods in a more controlled setting than the first user study described in this chapter. As a secondary goal, these focus groups were also completed with different age groups to give the results more breadth and to determine if age plays a significant role in the social acceptability of multimodal interaction techniques.

These focus groups made use of *experience prototypes* (Buchenau and Suri, 2000) in order to give participants a chance to experience the interaction techniques at a relatively low development cost. Experience prototypes refer to "any kind of representation, in any medium, that is designed to understand, explore, or communicate what it might be like to engage with a

product, space, or system" (Buchenau and Suri, 2000). These prototypes allow for the evaluation of large sets of interaction techniques that would not be possible using a fully developed system. Because the implementation of multimodal interaction techniques requires extensive development time and sophisticated sensing and recognition, early evaluations can guide implementation efforts to focus on only those interaction techniques that are both acceptable and usable. In this study, four such prototypes were used in a focus group setting to gain feedback on a set of multimodal interaction techniques using a minimal amount of development cost. Although the aim of this study was not to compare the *effectiveness* of different kinds of prototypes, the results include an analysis of the benefits and detriments of the prototypes used in this study.

4.3.1 Experience Prototypes

Each of the experience prototypes was designed to allow participants to gain enough experience with the interaction techniques to make decisions about how those techniques should be applied and performed. The four prototypes for this study were videos, live demonstration, a Wizard-of-Oz phone prototype with vibrotactile feedback, and a phone-shaped prototype without feedback. Images of each prototype are shown in Figure 4.4.

The video prototypes portrayed a male actor sitting in a plain setting. Each gesture or voice command was performed three times with a short pause in between. These videos were similar to those used in Chapter 3. Demonstration prototypes involved a live demonstration by the focus group leader. Again, this included three repetitions of each gesture or voice command. The Wizard-of-Oz phone prototype had a simple interface with an image of a stoplight. Participants were asked to perform each gesture or voice command. When performed correctly, the on screen stoplight would turn green, the phone would vibrate briefly and play a tone. Two experimenters at a nearby laptop controlled this wirelessly while the participants took turns using two running prototypes. The shape prototype involved a set of phones, which were switched off, and small black clips that were used to represent a clip-on microphone. Participants were asked to clip this on the collar of their shirt and perform each gesture or voice command with the phone in their hand. The phone did not provide any feedback in this case.

4.3.2 The Focus Groups

The goal of these groups was to explore how users place multimodal interaction techniques into context when considering the acceptability and adoption of these new interactions. Additionally, these focus groups investigated the usefulness of experience prototypes for early evaluations of social acceptability with respect to cost, ease of use, and quality of the experience created. The focus groups each examined a set of sixteen gestures and sixteen voice commands, shown in Figure 4.5. These modalities were chosen because of their often highly visible or audible nature, making them interesting interaction techniques from a social acceptability point of view. Although the focus of these groups was looking at adoption by putting these interaction techniques into a context, these groups also looked at acceptance as described in the previous chapter. The first phase of these groups included an evaluation of *acceptance*

Above: Wizard-of-Oz prototype.

Above: Video prototype showing hook finger gesture.

Below: Demonstration prototype showing open palm gesture.

Below: Shape prototype using Nokia N95 and small black clip.

Figure 4.4: Experience prototypes used during focus groups.

through the collection of ranking and rejection data, and questions about *why* interactions were liked or disliked before discussing any context of use. The full focus group schedule can be seen in Appendix D.1

For each focus group, the gesture and voice order was randomised, with half of the groups looking at gestures first and half the groups looking at voice first. Within the voice or gesture category, the order of individual interaction techniques was randomised for each group. Each focus group used one of the four available experience prototypes to familiarise the group with each interaction technique. Then, each participant filled out a short worksheet with rankings and acceptance information, and participated in a semi-structured group interview. The focus group worksheets can be seen in Appendix E.1. The interview topics included discussion of input preferences, locations where these inputs might be used, and the tasks for which these inputs might be used.

Interaction Techniques

For this focus group study, gesture and voice commands were organised into seven categories. The gesture categories were based on the categories used in the previous chapter (categories relating specifically to mobile interaction) including categories from Kendon (1997) and Nespoulous *et al.* (1986). The voice categories introduced here relate specifically to voice-based interaction on mobile phones. With respect to the gestures, four categories were used; emblematic (Kendon, 1997), device-based, arbitrary (Nespoulous *et al.*, 1986), and body-based. Emblematic gestures refer to those gestures that have a widely accepted meaning outside of the context of speech within a given culture, such as a peace sign. The device-based gestures are those that directly manipulate a device, such as a phone tap. Arbitrary gestures are hand positions that do not necessarily have an explicit meaning and may be open to interpretation, where any explicit meaning (such as an associated command in a mobile interface) must be learned. The arbitrary gestures used in this study were chosen based on their previous use in gesture-based systems (Mo and Neumann, 2006). Body-based gestures are movements of the body that do not directly involve manipulating a device, although external sensors might be manipulated for these gestures. With respect to the voice commands, three categories were used: command, speech, and non-speech. Command inputs included one-word commands that related directly to mobile phone tasks, such as "call" or "lock". Speech inputs included short, commonly said phrases that did not directly relate to mobile phone tasks. These inputs were included to evaluate the acceptability of speaking to a mobile phone in a way that is not obviously related to a device or phone tasks. Non-speech inputs included a variety of sounds and noises, some of which occur normally in everyday life, such as whistling (Sporka *et al.*, 2007) or humming (Sporka *et al.*, 2006), and some of which do not, such as buzzing or popping.

Participants

The study was made up of eight focus groups of two to four participants each. For the first six focus groups, which totalled nineteen participants, the participants were selected from local

Gesture	Category	Voice	Category
OK Gesture	Emblematic	Say "Close"	Command
Money Gesture	Emblematic	Say "Open"	Command
Peace Sign	Emblematic	Say "Call"	Command
Shrugging	Emblematic	Say "Lock"	Command
Device Stroke	Device-Based	Say "I'm Fine"	Speech
Device Shaking	Device-Based	Say "Bad Weather	Speech
Device Flick	Device-Based	Say "That's Nice"	Speech
Device Rotation	Device-Based	Say "So Busy"	Speech
Upright Fist	Arbitrary	Humming	Non-Speech
Hook Finger	Arbitrary	Buzzing	Non-Speech
Sideways Fist	Arbitrary	Say "Chh"	Non-Speech
Open Palm	Arbitrary	Doo Doo Doo	Non-Speech
Shoulder Rotation	Body-Based	Say "Psst"	Non-Speech
Wrist Rotation	Body-Based	Whistling	Non-Speech
Foot Tapping	Body-Based	Clicking	Non-Speech
Head Nodding	Body-Based	Popping	Non-Speech

Figure 4.5: A list of all the gesture and voice commands used in the focus group study, organised by category.

university students, ranging in age from 18 to 29. These groups included two groups looking at video prototypes, two groups looking at demonstration prototypes, one group looking at shape prototypes, and one group looking at Wizard-of-Oz prototypes. The study continued with another two focus groups, which totalled six participants, where the participants were recruited from members of the local community ranging in age from 70 to 95. These groups included one group looking at video prototypes and one group looking at demonstration prototypes. Although completing focus groups with participants from different age groups was a secondary goal of this focus group study, this gives the results greater breadth and begins to demonstrate how age might affect perceptions about social acceptability.

Of the twenty-five total participants, 56% of the participants were from the UK, 20% were from Asian countries, 8% were from Europe, and the remaining 16% declined to state. Focus group participants also answered questions about their mobile phone usage habits. Of the first set of focus group participants, aged 18 to 29, 100% of participants used their mobile to make phone calls and send text messages, 58% used the phone Web browser, 53% used their mobile to play games, and 37% used an email client on their mobile phone. Of the second set of focus group participants, aged 70 to 95, 83% used their mobile phone to make phone calls, 33% used the phone to send text messages, and 17% used email, Internet, or the alarm clock on their phone.

4.3.3 Results

The results of these focus groups are based on the worksheets filled out by participants and the transcripts of focus group discussions. Each participant filled out a worksheet with their own rankings for gesture and voice-based commands separately as well as their acceptance or rejection of each of the interaction techniques individually. After ranking each list independently, participants were asked to indicate which of the techniques they would not be willing to perform in a public setting such as a busy pavement. This was followed by a group discussion on preferences and scenarios where multimodal interaction techniques could be applied. The full list of questions can be seen in Appendix D.1 and the worksheets are given in Appendix E.1.

This results section is divided into three parts. Firstly, the results from the participants aged 18 to 29 are discussed. Secondly, the results the focus groups including participants aged 70 to 95 are discussed and compared to the first set of focus groups. Finally, results from both groups are discussed together to identify themes that ran across all the groups.

Focus Groups: Ages 18 to 29

The first six focus groups in this study involved nineteen participants aged 18 to 29 in groups of three to four. Using the non-parametric Friedman Test (Cramer, 1994) to compare ranking data for the gesture and voice categories, these results show that both of these modalities have significant differences in rank between their respective categories ($p < 0.001$). Using a one-tailed Wilcoxon signed-rank test (Cramer, 1994) for pair-wise significance tests, statistical significance was determined for pair-wise comparisons between the categories. These are shown in Figure 4.6. The effect sizes (Corder and Foreman, 2009) for these comparisons are

Comparison	p-value	Comparison	p-value
Device - Body	< 0.02	Command - Non-Speech	< 0.0007
Device - Emblematic	< 0.001	Command - Speech	< 0.0001
Device - Arbitrary	< 0.0005	*Non-Speech - Speech*	*< 0.23*
Emblematic - Aribitrary	< 0.001		
Body - Arbitrary	< 0.009		
Body - Emblematic	*< 0.29*		

Figure 4.6: Pair-wise comparisons for gesture and voice commands using Wilcoxon Signed-Rank test for participants 19 to 29. Italicised items are not statistically significant.

Comparison	Effect Size	Comparison	Effect Size
Device - Body	0.5	Command - Non-Speech	0.7
Device - Emblematic	0.7	Command - Speech	0.8
Device - Arbitrary	0.7	*Non-Speech - Speech*	--
Emblematic - Aribitrary	0.7		
Body - Arbitrary	0.5		
Body - Emblematic	--		

Figure 4.7: Effect sizes for pair-wise comparisons for participants aged 18-29. Effect sizes of 0.5 or higher are considered large effect sizes. Effect sizes for non-significant comparisons are not shown.

shown in Figure 4.7. These results are comparable to the results discussed in the previous chapter, which were based on Likert-style ratings rather than ranks.

These results show that device-based gestures were significantly more acceptable than any other kind of gestures, with arbitrary gestures being the least acceptable, emblematic gestures the second least acceptable, and body-based gestures the second most acceptable. Device-based gestures have higher acceptance rates because they provide spectators with clear cues explaining the performer's actions. Performers appear to feel more confident using these kinds of gestures because they are less likely to be misunderstood by spectators. Figure 4.8 shows the average rankings for each gesture, grouped by gesture category. The rankings ranged from 1 to 16, where 1 was the highest or best rank, and 16 was the lowest or worst rank. During these focus groups, participants discussed why they liked or disliked gestures. The arbitrary gestures were often described as unacceptable because they had unclear or confusing meanings for spectators. This led participants to be unsure how they would be perceived by spectators and uncomfortable performing these commands. For example, the hook finger gesture was described as a looking like a question mark, a beckoning finger, a pirate hook, a gesture for

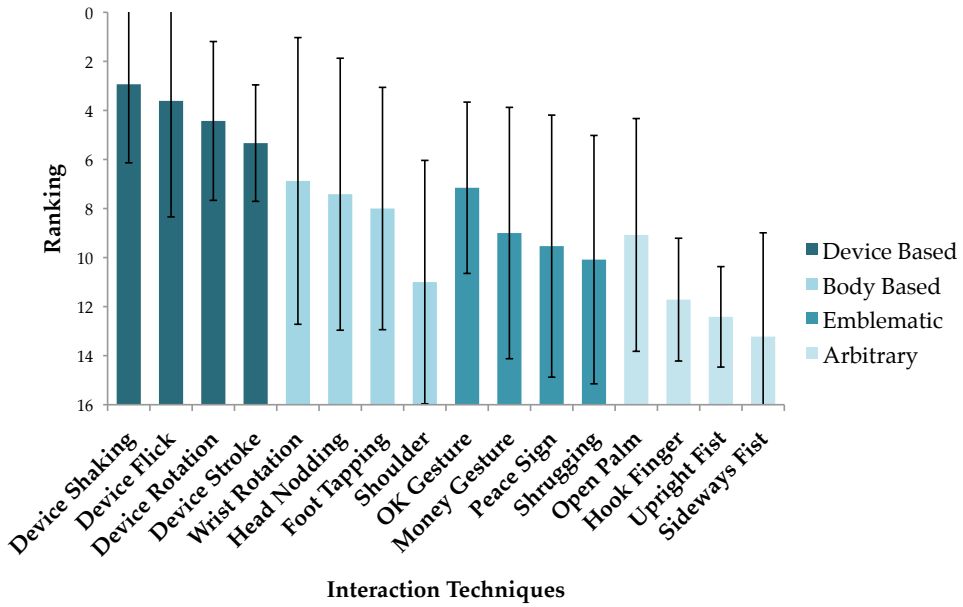

Figure 4.8: Gesture command average rankings for participants aged 18 to 29. Error bars show one standard deviation. Statistical significance for these comparisons is shown in Figure 4.6.

the number four, and a claw. This variety of interpretations led participants to dislike this gesture for its ambiguity. Emblematic gestures were also described as negative for multiple meanings or associations with undesirable meanings. For example, one participant stated that the peace sign gesture "made me think of people who take lots of photos of themselves." Here, pre-associated meanings were just as much of a disadvantage as ambiguous meanings.

It is clear from these results that the imagined interpretations of spectators influence how a performer makes decisions about the social acceptability of multimodal interactions. The ability to demonstrate the purpose of one's actions does not only apply to gestures, but also to speech. Figure 4.9 shows the average rankings for each voice command, grouped by category. Command speech inputs were significantly higher ranked for acceptability than speech and non-speech sounds. Command-based speech relates clearly to mobile phone tasks, which is easier to make sense of for spectators as compared to the speech and non-speech categories. One participant stated that "I don't mind all these easy to relate commands, lock, open, close, I don't even mind if I have to say them aloud." The clear indication that these commands were related to an interface made them more socially acceptable.

Although some categories of gesture or voice commands do lend themselves to a less ambiguous interpretation, the ability to demonstrate an interaction was not limited to command-based speech or device-based gestures. Participants described how unusual gestures could also indicate that actions were part of an interface. For example, one participant stated that "it's [hook finger] not to be mistaken, people won't connect that with a tick, it will be obvious you are steering something." One participant stated that "it is still a bit strange to start talking to yourself in the street. Whereas the things like the doo doo doos, they follow a pattern so someone would pick up that you were doing it for a certain reason. If you've got headphones on you have an excuse." Even a different configuration of the technology could be enough

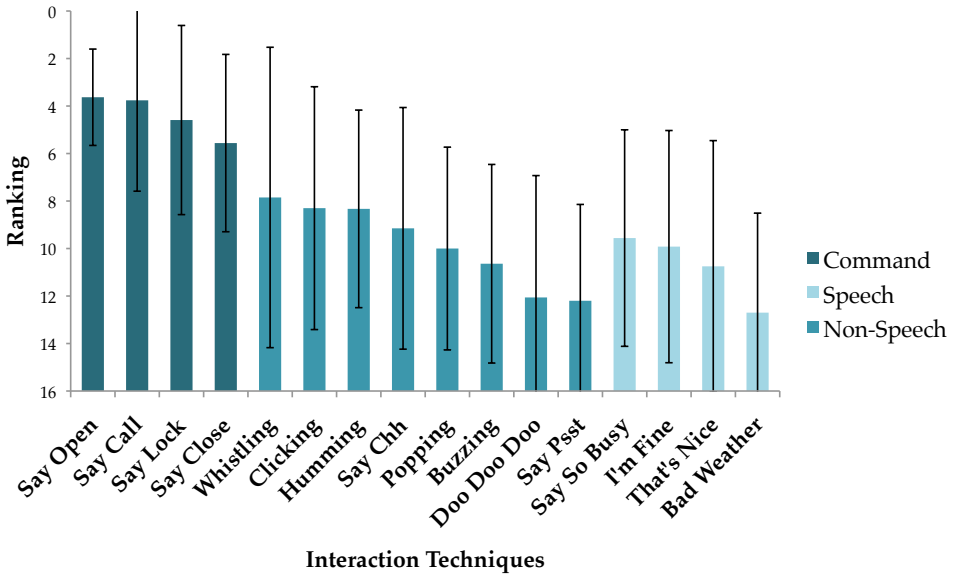

Figure 4.9: Voice command average rankings for participants aged 18 to 29. Error bars show one standard deviation. Statistical significance for these comparisons is shown in Figure 4.6.

to demonstrate an interaction. One participant described how they would be unwilling to perform voice commands with a hidden microphone, but that a visible microphone with a flashing light would be more acceptable. This participant stated that "you want people to notice you are saying it to your phone rather than a hidden mic."

Given the wide variety of gestures and voice commands investigated in this study as compared to previous work, participants described new reasons for liking or disliking gestures or speech. One new reason identified for liking gestures or speech was the *satisfaction* involved in completing a task in this way as compared to traditional interactions. One participant stated that "for cancelling something, shaking is incredibly satisfying." Similarly, participants described other input techniques as fun, silly, and cool. Participants discussed their thoughts on how comfortable it would be to converse with a device. One participant stated that "I'm not sure if I want to be talking to my device as if it's a pet or a creature." While in some cases this may be due to the lack of responses and natural conversation (Crangle, 1997), participants discussed their discomfort with the idea of a phone responding as well. One participant asked "if I say 'bad weather,' what does it [the device] do? Console me? I don't see how that would work with the phone." Participants also described some situations where they disliked intentionally hidden gestures. One participant stated that "if I was saying 'I'm Fine' to my phone I wouldn't want people to think I was on my phone, it's sneaky, like having a fake phone call with my phone."

The focus group setting encouraged participants to consider each other's differing opinions, leading to discussions about personality differences and how social acceptability might change over time. One participant stated that "I think it's one of these things that maybe doesn't seem okay at the moment, but the more people that use it the more OK it would be. I think at the start of using your capacitive things when the iPhone first came out, flicking them, it seemed a

bit strange at the start but it's OK." Talking about the future, participants discussed how they would feel if multimodal inputs became more widely accepted. One participant felt that "I don't want to look like an idiot doing it, but if everyone was doing it you would like an idiot if you weren't doing it." In contrast, another participant stated that "it's hard to say if I would be more comfortable doing it just because other people we doing it." Participants had differing opinions on how greater social change would affect their own adoption of multimodal interfaces.

Focus Groups: Ages 70 to 95

Two focus groups were completed with a total of six participants aged from 70 to 95. The comparison of ranking data for gesture categories between this set of focus groups and those groups aged 18 to 29 is shown in Figure 4.10.

The biggest difference between these two age groups can be seen in the differing rankings for device-based gestures. Using the Mann-Whitney test (Corder and Foreman, 2009) for independent samples of non-parametric ordinal data, these results show that device-based gestures were ranked significantly differently between these user groups ($p < 0.0008$, large effect size of 0.7). The previous chapter demonstrated that device-based gestures were significantly more acceptable than body-based gestures, but those studies did not compare a broad range of ages. In focus groups with participants ranging in age from 18 to 29, the average ranking for device-based gestures was 4.2 out of 16 as compared to 13.3 out of 16 in focus groups whose participants ranged in age from 70 to 95. When describing dislikes, one participant from the older user group stated that "I couldn't imagine stroking the device at all, that's alien to me." Because gestures directly involving the device were often unfamiliar to the older adults, these gestures did not have a clear meaning and were generally disliked.

Another clear distinction between these two groups was the ranking differences with emblematic gestures. The comparison between ranks for emblematic gestures for these two user groups were significantly different, ($p < 0.03$, medium effect size of 0.4). In this case, focus groups with younger participants ranked emblematic gestures at 8.6 as compared to 5.3 in focus groups with older adults. This preference for emblematic gestures is further demonstrated in the rejection rates gathered from the worksheets, as shown in Figure 4.11. Emblematic gestures had a much lower rejection rate amongst the older adults, with a rejection rate of 4% as compared to 29% with the younger adults. Because the emblematic gestures all had easily recognised meanings, their familiarity made them more acceptable amongst the older adults. For example, one participant stated that "the money gesture, I use that all of the time." Familiarity and clear meanings were important factors in determining acceptability for the older adults.

With respect to the voice input, there were significant differences in preferences with respect to the speech-based input that was not directly related to phone commands. A comparison between the ranks of this voice category between the younger adult and old adult group shows that these are significantly different ($p < 0.003$, large effect size of 0.5). The average rankings for the voice command categories are shown in Figure 4.12. These inputs were more highly ranked by the older user groups because they were often described as familiar and common

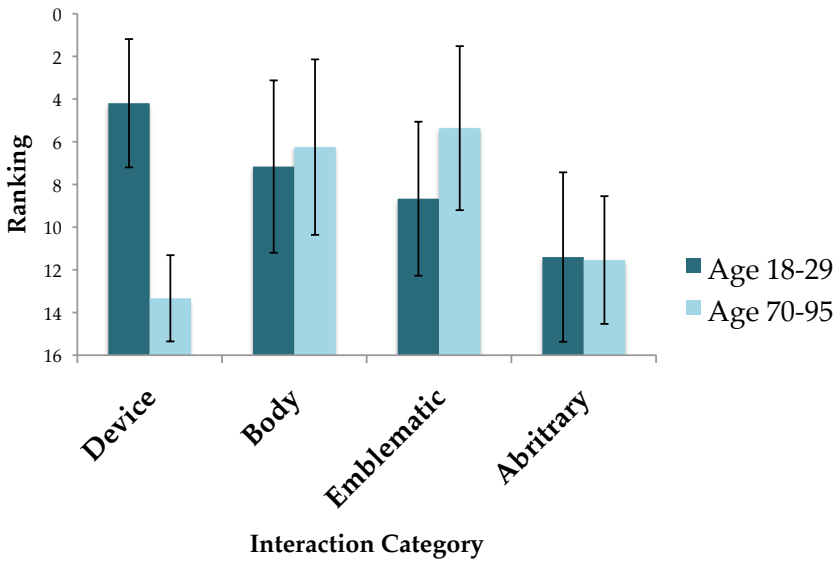

Figure 4.10: Comparison of focus groups by age with average rankings for gesture categories. Error bars show one standard deviation.

phrases in everyday life. One participant stated that "you would say 'That's nice,' that's an expression" The fact that there was a clear and unambiguous meaning associated with these sayings shows again how meaning was more important for the older adults than a clear association to the device or phone tasks.

An issue that repeatedly came up during these groups was the different meanings of gesture or voice commands for each individual. While the relationship of interaction techniques to mobile phones had some importance, the older adults tended to focus more on what these interactions meant personally before they would consider how they might relate to a mobile phone task. When describing voice commands that were liked, one participant described how familiarity was an important factor. This participant stated that "a lot of these things convey nothing to me, Doo doo doo? I've never heard anyone say that." The lack of familiarity or meaning led to situations where gestures or voice commands could not be put into scenarios that made sense. One participant stated that "I don't know the implication of this hook finger. I don't see how I would use many of these." Similarly to the younger user groups, the older user groups disliked interactions with unclear or ambiguous meanings. When discussing why the open palm gesture was disliked, one participant stated that "I felt there could be two meanings. It could mean 'hi' [open palm held at shoulder level], but if someone was being very aggressive to me, I would say stop [open palm held out with arm extended]." The fact that this gesture could mean different and possibly negative things given a slight change in performance made it less acceptable as a gesture to use in public. Multiple meanings were also discussed when these commands were performed in different contexts. One participant stated that "whistling, one might do to entertain the neighbours, but to whistle at someone, to catch their attention, might be rather rude."

Other issues discussed in these groups included gender differences with respect to the acceptability of certain gestures or voice commands. One participant described several gestures

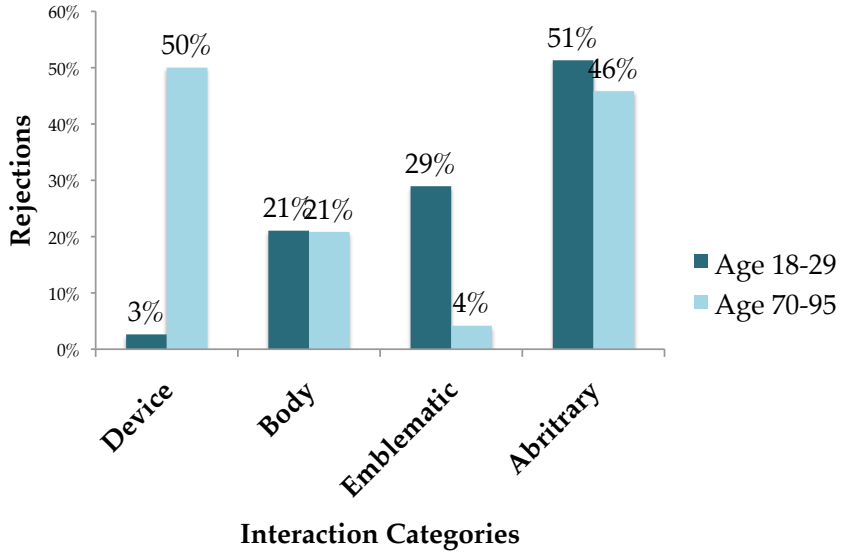

Figure 4.11: Comparison of gesture rejections rates for gesture categories and focus group age ranges.

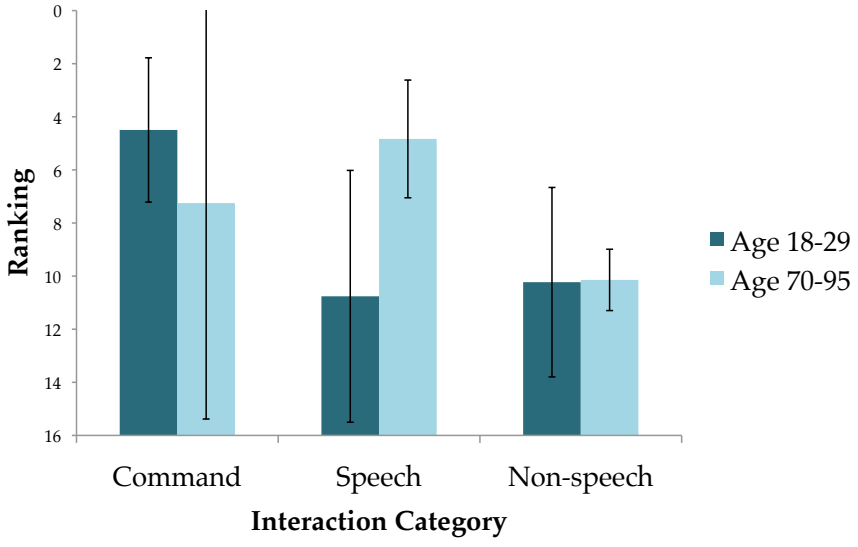

Figure 4.12: Comparison of focus groups with average ranking for voice categories.

as simply unladylike. She said "what's the saying? Cackling hens and something whistling women. No, it's not a good idea. At least that's what I was taught." In some cases, participants said they would not be embarrassed to do anything, or that if gestures or speech became popular with younger generations they would also be unembarrassed to adopt the behaviours. One participant stated that "if everyone else was doing it [popping, whistling, psst, doo doo doo], the young people were doing it ... I mean, I was never going to get a mobile phone. I was never going to get an iPod, I've got both." These results show again that participants were often considering how the greater social acceptance of a new interaction technique might influence their own behaviours.

Concerns About Usability in the Wild

In all focus groups, participants brought up specific concerns about how gesture or voice input might work in the wild and discussed some of their anxieties about performing these interactions in different failure situations. Although some of these are common failure scenarios, such as false positives or false negatives, the roles that performativity and spectators play in these interactions brings out a variety of failure scenarios that involve the audience and develop in highly noticeable ways.

False Positive Recognitions – The greatest concern participants discussed was the possibility of false positive recognitions given that many of the inputs could be accidentally performed during everyday activities. One participant stated that "the gestures were things you would do without thinking, like with your shoulders, you might accidentally activate the device." If gestures were not likely to be performed accidentally, participants felt more comfortable about how the system might work. One participant stated that "it's [shoulder rotation] a very unnatural gesture, it's not something you would normally do, you can keep it as a steering gesture. This is a gesture that is sufficiently easy to do but not a normal part of your behaviour that you could do it by mistake." Participants were often concerned with how the resulting actions from their mobile interface might embarrass them if accidentally activated. For example, one participant stated that "it's a bit dangerous. If you had the phone in your pocket, a lot of these things you might say quite frequently. It might be unlocking and taking pictures and wasting batteries."

Audience False Positive – The second most commonly discussed concern was the possibility that spectators might mistake one of the inputs as an action directed towards them. One participant stated that "if you say something like 'Psst' somebody might think you are trying to talk to them." In particular, this was seen as a problem for inputs that did not involve a device. One participant stated that "it makes it even more ridiculous if you don't even have a phone in your hand, like playing with my iPod, I can shake it and it makes sense because I'm holding it. But if I'm walking down the street and suddenly [demonstrates a shaking motion] people might think I'm casting spells on them." These unwanted reactions from spectators could lead to highly embarrassing situations, where some interactions made participants particularly worried about this failure scenario. For example, an interaction like the 'Psst' command were more likely to be misinterpreted as being directed towards a spectator than an interaction like foot tapping.

Distance of Device – For scenarios where the device was not directly manipulated, participants described a feeling of disconnect from the interaction and uncertainty about successful execution. One participant pointed out the importance of having a "clear connection with the action and the device." Because some of these interaction techniques do not directly manipulate a device, participants were concerned about usability and feeling in control. One participant stated that "I think for the ones where you are actually holding the phone, you have some control over it." Another participant was worried that being distant from the device would make it more difficult to know if commands were interpreted correctly, which highlights the importance of feedback and the challenges of effective feedback when the device is distant.

Failure to Recognise Inputs – In a scenario where intended inputs were not correctly recognised, participants were concerned about the necessity to repeat movements or voice commands until a successful recognition was achieved, leading to frantic and embarrassing behaviour. While demonstrating an erratic shrugging gesture, one participant stated "if it happens that when you try to do it and there's no execution, you keep on shrugging." The false negative failure scenario is a common concern in most interfaces, but develops in an interestingly different manner when the interaction is performative. In this situation, the performance of a failed interaction becomes even more noticeable and embarrassing. Repeated actions might feel unnatural or uncomfortable, and could cause embarrassment because repetition is often used as a social signal. For example, tapping a glass loudly with a few repetitions is a good signal for getting attention at a dinner party. Because repeated actions are highly noticeable, participants imagined that failed recognitions would be highly embarrassing and attract unwanted attention from spectators.

Experience Prototypes

One of the secondary goals of this study was to evaluate different types of low-cost prototyping methods with respect to their usefulness and cost. Although differences in 'effectiveness' between these prototypes cannot be evaluated given the relatively small number of times each prototype was used, differences in cost, control over experience, and flexibility can be analysed based on the experiences brought out in this study.

The cheapest prototype to create was the demonstration prototype. This prototype only required a demonstrator who had practiced and was familiar with each technique. The benefits of this prototype include the ability to either require participants to practice each command, request specifically that they only watch, or observe how participants respond to demonstration and which commands they choose to practice naturally. The ability to request or discourage repetition gives the focus group leader more control over the experience created by the prototype. The detriments of this prototype include the need for a trained demonstrator and the possible lack of consistency in performance, especially if different individuals act as the demonstrator. Also, the appearance and cultural background of the demonstrator may have an effect on how the demonstrations are perceived. For example, an individual with a different accent from the focus group participants might be difficult for them to understand. Gender roles may play a significant role in some cultures and not others.

The second cheapest prototype to create was the shape prototype. This prototype only re-

quired a set of mobile phones and objects that approximated peripheral devices such as external microphones. The benefits of this prototype include the ability for each participant to try the interaction techniques with a tangible object. Especially for gestures that involve the device, this allows participants to experience how it feels to perform the command with a realistic object in their hand. For voice commands, the shape prototype allows participants to explore a variety of configurations easily and cheaply. For example, configurations including a hidden microphone in the collar and talking to the phone held to the ear could be easily experienced using the shape prototype. One of the detriments of this prototype is the lack of consistency between individual participants' performances. Because the shape doesn't provide feedback, it is difficult to ensure each participant performs the commands consistently. Also, this prototype may be uncomfortable for participants who are shy or nervous about performing commands in front of others.

The second most expensive prototype to create was the video prototype. This prototype required an actor to perform each command as well as video capture and editing equipment. This prototype also required a way of projecting the videos to be viewed during the focus group. The benefits of this prototype include a perfectly consistent performance since each participant will see the same videos. Like the demonstration prototype, the experimenter can also control the experience by requesting users to respond in a specific way to videos or by recording their automatic responses. One of the detriments of this prototype is the need for a projector or other way of showing videos to a group. Like the demonstration prototype, the choice of actor may also affect how individuals perceive performances.

The most expensive prototype to create was the Wizard-of-Oz prototype. This prototype requires a set of phones, laptop or other controlling equipment, development time, and a trained 'wizard' to run the system. Depending on the desired level of sophistication for the prototype, the development time can vary significantly. The benefits of this prototype include the ability to provide a hands-on experience that provides feedback to the participants and thus ensures a consistent performance for each participant. The detriments of this prototype include the fact that not all participants were able to use the prototype at once, which led to a slow and relatively segmented experience. Also, participants who are shy or embarrassed performing these commands in front of others might be made uncomfortable when using this prototype since it necessarily requires performance in front of the group.

Placing Interactions into Context

When placing interactions into context, participants used three main approaches to understanding and pairing interactions with tasks. Figure 4.13 shows the frequencies for three tags that were used to code the focus group transcripts, where each dot on the figure represents an occurrence of that tag within each focus group. These tags and frequencies demonstrate the three approaches that participants used in understanding the interaction techniques in each group and placing them into a usage context. One approach focused on placing interactions into a usage context based primarily on personal meaning. In this case, the *meaning* tag was used whenever interaction techniques were connected with some kind of personal or social meaning. This included cultural meanings, the meanings of familiar actions, or the lack of

a meaning. For example, one might think of the "peace sign" gesture as funny pose for a picture and want to associate that gesture with the camera functionality. Another approach focused on the device itself, and interactions were chosen to fit best with existing ideas about how devices should be used. The *device* tag was used when any command was discussed in relation to a device, in both positive and negative ways. For example, one might feel that foot tapping was not a useful technique because its relationship to the device was unclear. Finally, another approach looked at functionality first, then interactions were chosen that made the functionality more easily accessible and usable. The *usage* tag was used whenever commands were discussed within a specific usage scenario or when these scenarios were unknown or unimaginable. For example, if email is the primary use of one's mobile phone, one might want to create a 'gesture shortcut' to that application using a simple gesture. For all of these tags, examples of quotes that represent that tag are shown in Figure 4.13. These three tags represent the three approaches that were used by participants in these focus groups to apply these interactions to usage scenarios.

These three approaches to placing interactions into a usage context represent three important aspects of usability for multimodal interaction. Semantic meaning plays an important role in how users might make sense of these interaction techniques and perform them in public places. Participants discussed how the meanings of different interactions would make them feel more or less comfortable performing these actions in public places. The functionality or goal of the interaction is also important when users determine how interactions could make accessing functionality easier than using traditional buttons and menus. Participants described how single interactions could replace a series of actions in a traditional interface. However, making interaction easier was not the only way that participants considered functionality and multimodal interfaces. Participants also discussed how certain actions could be more satisfying than simply pushing a button. For example, participants liked the idea of tapping or hitting the phone in order to turn off or snooze an alarm. Considering functionality with respect to ease of use and satisfaction was important to these participants. Finally, the device itself plays an important part in the experience of using the system and the ways in which the interactions can be performed. Participants, particularly those from the young adult user groups, discussed how manipulating a device directly gives more confidence in performance and helps the interactions feel more physical when manipulating virtual elements.

There were differences in how different age groups evaluated these interaction techniques. For focus groups with younger adults, the frequencies of these different tags demonstrate that discussions were generally evenly distributed, with some groups focusing more on one tag than others. For example, one group that used the demonstration prototype focused on discussions about the device while the other demonstration prototype groups hardly talked about the device at all. The two older user groups, however, had a clear focus on the meaning of these interactions with much less concern for the device itself. These differences in approach led to very different preferences for interactions, with the older user groups having a clear preference for interactions with clearly defined meanings. Another reason these three approaches might be distributed across different focus groups could be the social dynamics of the groups themselves. Stewart *et al.* describe a variety of factors that influence the dynamics of a focus group, including individual personalities, age, and appearance (Stewart *et al.*, 2007). These factors are most likely the cause of variance between groups rather than the prototype used. How-

ever, the groups clearly demonstrate the importance of these three approaches for matching interaction techniques with functionality.

4.3.4 Discussion

There were some important differences that were observed between the younger adult and older adult groups. This includes a preference for emblematic and speech-based inputs rather than device and command-based inputs. This was largely due to the fact that the older adults focused more on personal meaning and understanding rather than a clear connection or application to the phone when examining these interaction techniques. This led to notably different preferences for interaction techniques and different methods of applying those techniques to usage scenarios. For example, older user groups discussed how the meaning of gestures might map to actions on a mobile phone, where younger user groups discussed how gestures might make commonly used functionality easier to access. These different approaches demonstrate the importance of interaction design in creating a metaphor between action and consequence that makes sense for a variety of users and their differing goals for interaction.

These focus groups also highlighted the anxieties participants have about using these multimodal interactions as part of their daily lives with respect to possible failures. Because these interactions can be highly visible or audible, the ways in which participants imagined these interactions failing revolved around the performance of the interaction and the reactions of spectators rather than the failure of some functionality. These failure scenarios highlight an interesting issue with the adoption of these interaction techniques because those techniques that are the most comfortable to perform are not always the same as those that participants feel most confident performing with respect to usability and accuracy. For example, the shoulder tap gestures was poorly ranked, but would make a good gesture in a real world interface because it would not usually be performed outside of the interface and it can be reliably recognised (Strachan *et al.*, 2007). The relationship between confidence in successful performance and the appearance of that performance is an interesting issue. For example, are users willing to lose some reliability in recognition in order to have higher comfort while performing? If an interaction requires one to shake a phone, in public users might be willing to lose some reliability to perform the action less noticeably, but while at home feel happy to perform the gesture with more animation.

An important aspect of multimodal interaction design concerns how those interactions can be demonstrated as part of an interface. Participants discussed the importance of avoiding confusion about why they were gesturing or speaking when such actions might be misunderstood by spectators, leading to an embarrassing or confusing situation for the user. While some interactions can easily be demonstrated as part of an interaction by holding or displaying a device, other methods, such as making peripheral devices more visible, were also identified as successful ways to demonstrate interaction. A possible area of interest in future studies of social acceptability could investigate the different ways that hands-free or device-free demonstration of interaction can be achieved.

Participants also discussed a variety of cultural differences and how they thought acceptance

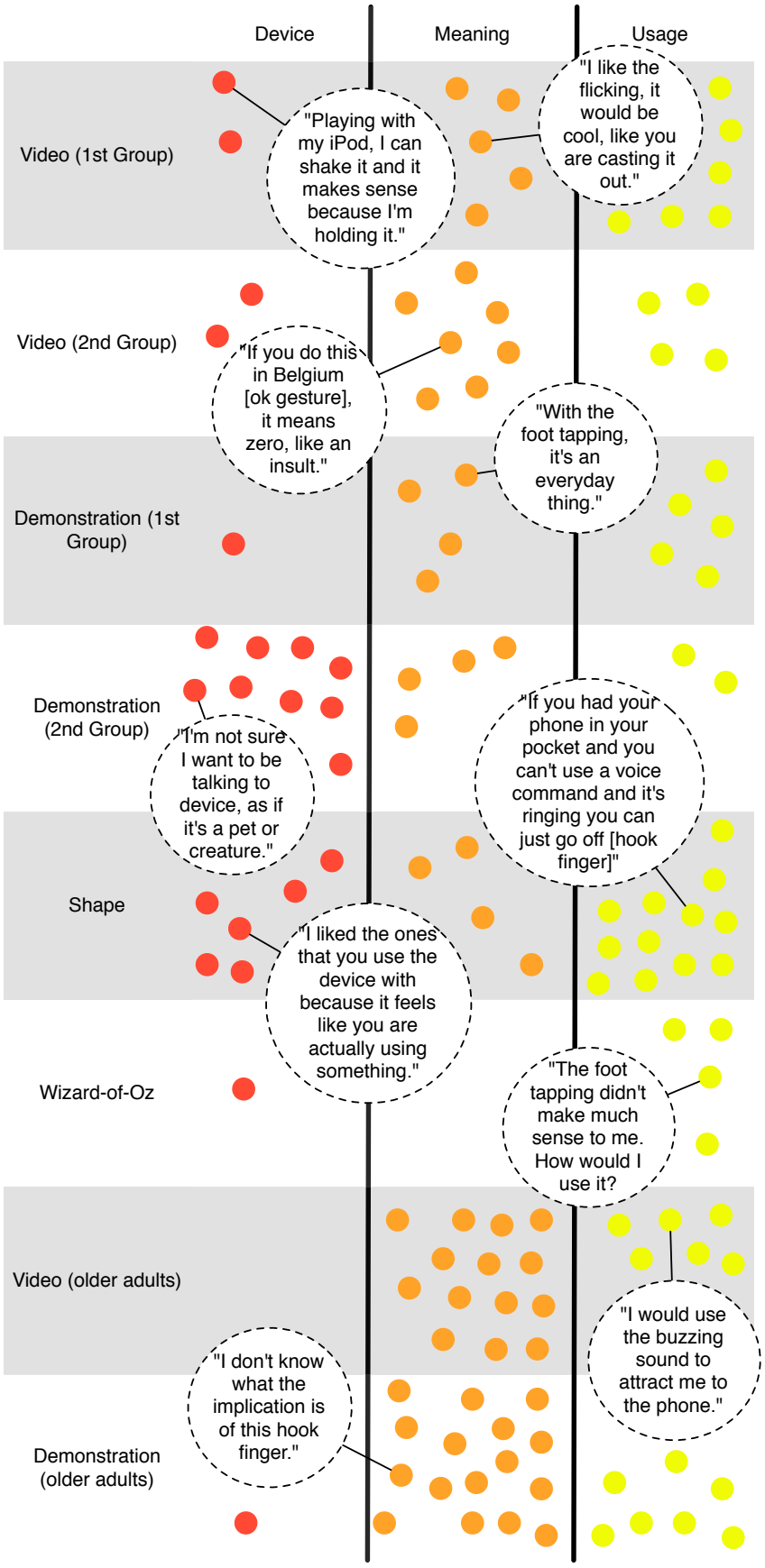

Figure 4.13: Selection of tag frequencies with example quotes for each focus group. Each red, orange, and yellow dot represents an occurrence of the device, meaning, and usage tag, respectively. Each quote represents an example of the type of tag it is linked with on the transcript.

of gesture and voice inputs might change over time. Cultural differences with gesture indicated that many of the movements could be impolite or even offensive in some cultural settings. For example, the OK gesture was also described as a gesture for meaning zero, worthless or filthy in some cultures. Participants also discussed the differences in the appropriate level of noise making in public places in different cultures. One participant described how whistling in public can be considered improper in some cultures. These kinds of differences were brought up by many of the focus group participants.

4.4 Conclusions

The two studies presented in this chapter explore two important aspects of multimodal interaction adoption by looking at a set of multimodal interaction techniques within specific application areas and usage contexts. The first study described looks specifically at performativity in interface adoption by requiring participants to create multimodal input in real world settings over repeated trials. The results show how performances were created based on location, spectators, and props. The second study looked at how participants might apply multimodal interaction techniques to different application areas through a focus group study. Each focus group utilised an *experience prototype* in order to give participants some experiences with a wide variety of interaction techniques at a relatively low cost. The results of these focus groups demonstrate three approaches that participants used when placing interactions into context and highlighted the anxieties participants had about using multimodal interactions in public.

The results of the first study develop an in depth and personal analysis of participants interacting and performing with a remote awareness application. Participants took advantage of the wide variety of possible performances, and created input that gave their fellow participant interesting output to watch while maintaining their own comfort. In some cases, experience and enjoyability were more important than subtlety, where participants performed highly amusing actions purely for the sake of creating a satisfying experience. The choice of prop used in the application had an impact on performance, not only by making participants feel more connected to their fellow participant but also by changing some of the visual aspects of performance to make it fun, practical, or disguised. The places where performance can take place and the objects that might be used as props play an important role in the overall experience of using technology in public. These issues play a key role in the design of mobile interactions where reliability and accuracy are just as important as experience and appeal.

The results of the second study described in this chapter demonstrate some important aspects of the experience of using multimodal interactions as imagined by the study participants. The ability to demonstrate interaction to spectators, whether through a clear connection to a device or by observable aspects of the interaction, plays a major role in acceptability. Participants also discussed how they would go about placing multimodal interactions into real world contexts, using three approaches across the eight focus groups. These three approaches include focuses on meaning, functionality, and the device itself. This study demonstrated the differing approach to understanding and accepting multimodal techniques between different generations

of users. For the older adults, familiarity and personal meaning were more important than clear connection to a device when evaluating social acceptability. The variety of experience prototypes used in this study also provided insight into prototype design with respect to the cost of the prototype, the ability to control the experience, and the consistency provided.

By examining how interactions might actually be adopted and used as part of a mobile interface, this chapter builds on the previous chapter in two ways. Firstly, it looks at how interactions are paired with functionality and motivated by goals within an interface by examining how users might create multimodal input and how users place multimodal input into context. Secondly, this chapter explores the ways that participants imagine multimodal interaction actually unfolding in real world settings. In these focus groups, participants imagined highly performative or embarrassing failures caused by themselves, the applications, and spectators alike. Going beyond simple acceptance to adoption also meant that social factors began playing a more significant role. Participants discussed how cultural differences may influence adoption, how smaller social groups could influence adoption, and how usage over time might affect an individual's usage and adoption habits. These issues form the basis of the next chapter, which looks specifically at interface *appropriation* over time in different cultural groups.

Chapter 5

Motivation and Presentation of Self in Interface Appropriation

5.1 Introduction

The surprising and interesting ways in which users adapt new interactions to fit into their everyday lives is an important aspect of technology in society that has been studied in a variety of settings. This *appropriation* refers to the process by which users makes sense of and come to be comfortable with new interactions to the point where they begin using them in their own way. In the words of Dix, this is the process of making technology the "users' own, not simply what the designer gave them." (Dix, 2007) Evaluations of appropriation have a strong tradition in computer supported cooperative work (CSCW), such as Orlikowski's ethnographic work on Lotus Notes (Orlikowski, 1992) or Grudin's work on workplace calendars (Grudin and Palen, 1995). However, studies of appropriation are not limited to CSCW or groupware applications, but also allow for the effective evaluation of mobile technologies (Sambasivan *et al.*, 2009), tangible interfaces (Brewer *et al.*, 2007), and performative experiences (Iacucci *et al.*, 2002). These evaluations are essential for understanding how technology might be used realistically in everyday life and how users might adapt interactions to make them more acceptable.

This chapter explores appropriation by focusing on the adoption and adaptation of *performance* in everyday life. This kind of appropriation occurs when users adopt performances, add personal meanings to interactions, and successfully integrate performance into their everyday lives. In order to successfully deploy a multimodal mobile application for an extended user study, this chapter looks at the entire design process of a multimodal system from initial explorations of acceptance, to in depth discussions about adoption and use and through to evaluation of the final product in real world settings. In order to thoroughly test this design process, the studies presented in this chapter were completed concurrently in two very different cultural settings to better understand the how cultural factors affect social acceptability and to test the effectiveness of the methodology. Cultural factors are likely to influence social acceptability because multimodal interactions, which often involve highly noticeable gestures

or speech, might be interpreted differently, allude to cultural references, or hold different personal meanings in different cultural settings. Gestures with arbitrary meanings in one setting may be heavily significant in another. Speech commands which are highly disruptive in one setting may be part of everyday life in a different setting. The design process followed in this chapter began by identifying possible interaction techniques for evaluations and any cultural differences that might influence preferences. This was followed by the evaluation of user reactions about acceptance for a large set of possible interaction techniques that explores multimodal adoption scenarios. Finally, a complete mobile multimodal system was evaluated in the wild in order to study adoption and appropriation.

This chapter presents three studies that build on each other to examine usage and appropriation in the wild. This builds on the previous chapters by demonstrating how prototype evaluations of social acceptability can be incorporated into the design of a complete multimodal mobile system. These studies were completed concurrently with participants from Glasgow, United Kingdom and Bangalore, India. These different cultural settings were chosen in order to compare a typical Western user group with an emerging market user group. These groups offer an interesting comparison because both will be familiar with mobile technology but individuals from these cultures will use mobile technology in very different settings with different histories of mobile technology in everyday life. The first study in this chapter looked at existing mobile phone etiquette, as in the existing commonly accepted mobile phone practices in a given social or cultural context, and usage habits through a survey study in order to create a basis for comparison between Glasgow and Bangalore. The second study explored acceptance and adoption through focus groups that utilised video prototypes and a combination of quantitative questions about technology acceptance and qualitative questions about technology adoption. The final study was an extended 'in the wild' user study where participants used a multimodal application for six days in their everyday lives. The multimodal application used in this study was designed based on the results of the focus groups and surveys. This study demonstrates some of the ways that participants incorporated multimodal interaction into their everyday lives and became comfortable *performing* in public.

5.2 Understanding What Users are Doing Now: Etiquette and Habit

In order to understand how users are currently using their mobile phones, with respect to which functionalities are most used and what etiquette influences mobile phone usage in public spaces, a survey study was completed to explore these issues in different cultural settings. Surveys looking at mobile phone etiquette in different cultural settings have been completed in previous work (Campbell, 2007). The survey presented in this chapter consisted of multiple and short answer questions that were collected in Bangalore and Glasgow over the Internet and on paper. This first step in the design process served as a baseline of comparison for this cross-cultural study and identified areas of inquiry where user responses were unexpected or highly divergent. The results of this survey highlight the differences in mobile phone habits between respondents in Glasgow and Bangalore and provide a starting point for the following focus group evaluation of multimodal interaction techniques.

5.2.1 The Respondents

The respondents for this survey were carefully chosen from similar individuals in different cultural settings with respect to familiarity with technology, education, and work environment in order to focus on the cultural aspects. The survey focused on university students in full time education and adults working in office-style professions, such as human resources, financial services, public service, and information technology, described as 'working adults' throughout this chapter. The respondents were recruited through email, social networks, and university networks. Surveys were completed both on the Internet and on paper.

5.2.2 The Survey

The survey was designed to explore existing mobile phone practices, with questions relating to the respondents' mobile phones, their usage of different functionalities, and acceptability of using their mobile phone in different locations. The full list of survey questions can be seen in Appendix A.3. The survey was divided into three main sections: personal information, information about the respondent's mobile phone and information about how respondents used their mobile phones in a variety of locations. In the first section, respondents were asked to fill out basic demographic information such as their age and occupation. In the second section, respondents were asked to answer questions about their mobile phone, such as what kind of phone and service plan they had, how long they have had their phone, and what functionalities their phone had. The final section used multiple choice questions to determine how often different mobile phone functionalities were used with a four point scale from 'daily' to 'never.' Additionally, this section asked questions about where different functionalities were used, where respondents could indicate whether or not they used the given functionality in that location.

5.2.3 Results

The results of this survey are based on a total of 147 responses, comprised of 61 respondents from Glasgow and 86 respondents from Bangalore. In Glasgow, 41% of the respondents were students, ranging in age from 17 to 29 with an average age of 20. The remaining 59% were working adults between the ages of 21 and 46 with an average age of 31. The survey respondents in Glasgow were 37% percent female, 63% male. In Bangalore, 44% of the respondents were students, ranging in age from 20 to 28 with and average age of 23. The remaining 56% were working adults between the ages of 21 and 48 with an average age of 29. The survey respondents in Bangalore were 44% female, 56% male.

Respondents were asked to provide basic data about their current mobile phone practices. Respondents from Bangalore kept their phones for an average of 23 months, sent an average of 16 text messages per day and made an average of 13 phone calls per day. 66% of the survey respondents from Bangalore used a phone with Internet capabilities. Respondents from Glasgow kept their phones for an average of 11 months, sent an average of 13 text messages per day, and made an average of 2 phone calls per day. 90% of respondents from Glasgow

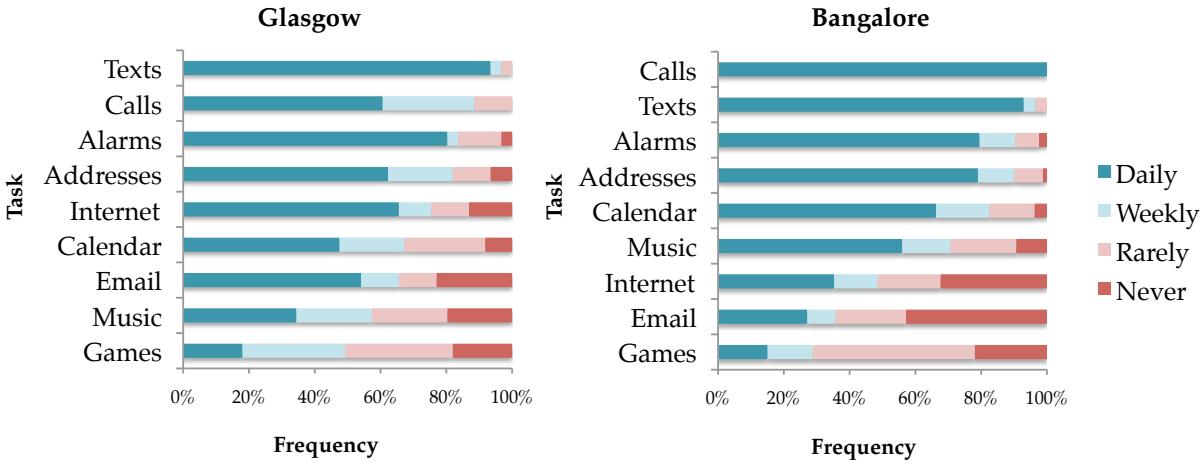

Figure 5.1: Frequency of use for common mobile phone functionalities for respondents in Glasgow and Bangalore. Graphs are presented with 'daily' on the left to 'never' on the right.

used a phone with Internet capabilities. The most prominent difference between respondents in Glasgow and Bangalore can be seen in the average number of phone calls per day. Using a two-tailed Mann-Whitney test for non-parametric data (Corder and Foreman, 2009), the results show that respondents in Bangalore made significantly more phones calls per day than respondents in Glasgow ($p < 0.0001$, large effect size $r = 0.7$). This clear difference between the number of phone calls made per day for respondents in these different cultural settings demonstrates a significant difference in existing practices. The following sections examine these differences in more detail by looking at the frequency of use for a variety of mobile phone tasks and the acceptability of using these in different locations.

Mobile Phone Functionality and Frequency of Use

The survey asked questions about how often common mobile phone functionality was used, where respondents could answer with daily, weekly, rarely, or never. Figure 5.1 shows the frequency of use for common mobile phone functions for respondents in Bangalore and Glasgow.

These results further illustrate the differing habits for respondents in Glasgow as compared to Bangalore with respect to mobile phone calls and text messages. 100% of the respondents in Bangalore reported using phone calls daily, as compared to just 61% in Glasgow. Using the Mann-Whitney test, these results show that respondents in Bangalore used phone calls significantly more often than respondents in Glasgow ($p < 0.0001$, medium effect size $r = 0.3$). This results confirms that respondents in Bangalore not only use phone calls significantly more often than respondents in Glasgow, but also complete significantly more phone calls per day as reported above. The most frequently used functionality by respondents in Glasgow was text messaging, with 93% of respondents using this functionality daily. This demonstrates a clear preference for text messaging over using phone calls in Glasgow, where the opposite is true for respondents from Bangalore. These differing preferences could have significant impacts on the kinds of multimodal interaction techniques that would be accepted in these different settings.

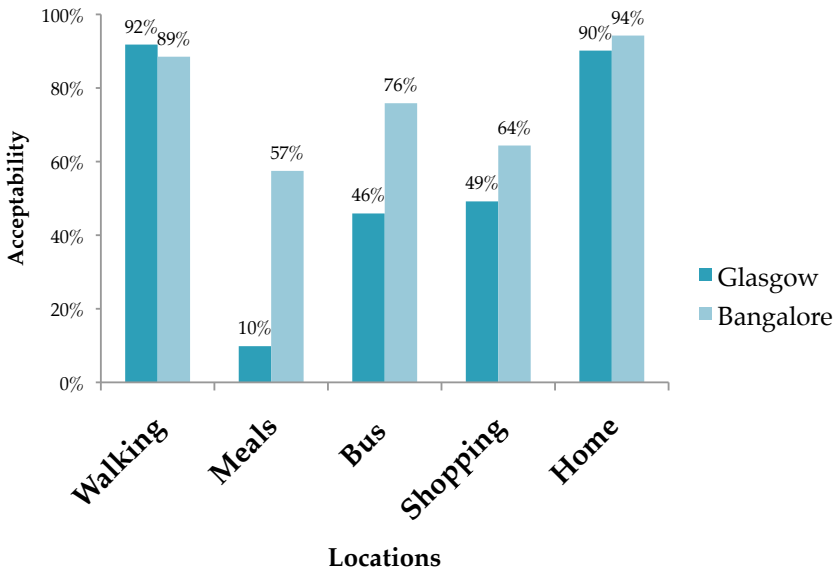

Figure 5.2: Acceptability of phone call usage in different locations for respondents within different cultural settings.

For example, speech-based interaction for browsing a contact list or initiating a call might be highly acceptable in Bangalore and highly unacceptable in Glasgow. This could be due to a variety of factors, such as the lack of phone call usage in Glasgow or the unacceptability of speaking in public. Although this survey cannot describe *why* these differences were observed, the identification of these differences can help to focus a more qualitative study exploring these differences.

Where Mobile Phones are Used

This survey asked respondents whether or not they felt that common mobile phone functionalities were acceptable in a variety of locations, where 'accept' or 'reject' were the only possible responses. Figure 5.2 shows the percentage of acceptance for mobile phone calls in these locations, and Figure 5.3 shows the acceptance for text messages.

These results show differences in the acceptability of phone calls and text messages in different locations, where different functionalities were preferred in different cultural settings. For example, respondents in Bangalore were much more likely to use a phone call while walking than a text message, as compared to respondents in Glasgow, who were just as likely to use calls or text messages in this setting. Although respondents from Glasgow and Bangalore were equally likely to use text messages during meals, there was a more pronounced difference in the acceptability of making a phone call in this setting, with only 10% of respondents from Glasgow finding this acceptable as compared to 57% in Bangalore. A similar comparison can be seen in the bus setting, where respondents were similarly likely to use text messages, but where respondents from Bangalore were more likely to use phone calls as compared to respondents in Glasgow.

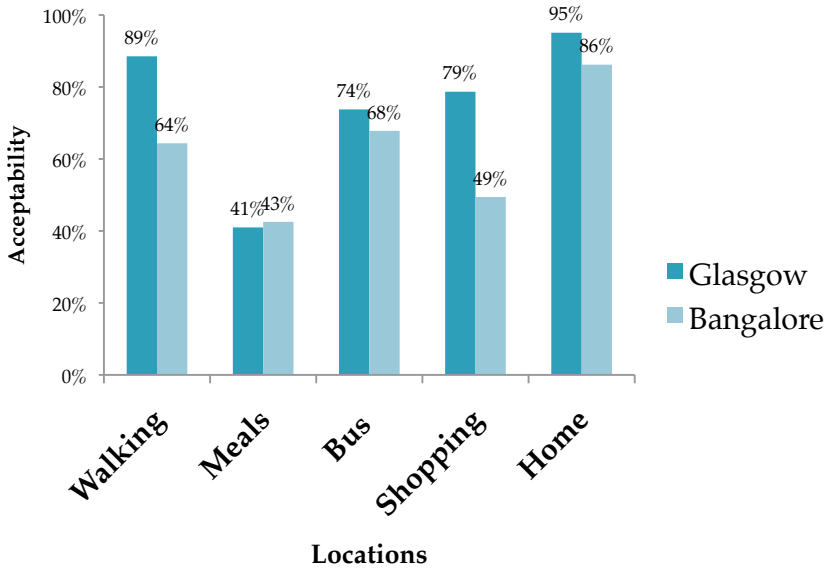

Figure 5.3: Acceptability of text message usage in different locations for respondents within different cultural settings.

5.2.4 Discussion

These results demonstrate some of the differing mobile phone preferences and practices for respondents from two different cultural settings. Respondents from Bangalore made more phone calls per day, and used phones calls more frequently in their everyday lives. In contrast, respondents from Glasgow made significantly fewer phone calls, but sent a similar number of text messages as respondents in Bangalore and used text messages more frequently than phone calls. These differences in existing mobile phone usage indicate variations in mobile phone etiquette and preferred mode of communication in these different settings that could affect preferences and acceptance of multimodal interaction techniques. For example, if respondents in Bangalore are more likely to use mobile phone calls in public settings than they might also be more likely to accept voice-based interaction in public settings as well. This could be especially true if the interaction is designed to mimic the appearance of a mobile phone call. In contrast, participants in Glasgow used phone calls significantly less often in public locations. If voice-based interaction was deployed in Glasgow, it might be more acceptable to make the interactions less noticeable rather than disguising them as mobile phone calls. For example, users might be more willing to whisper to their phone than have an audible conversation. These are the kinds of issues that can be uncovered with an initial survey and explored in more detail in a follow-up qualitative study.

Respondents from Glasgow favoured using text messages in a variety of locations, including during meals, while on the bus, and while shopping. On the other hand, respondents from Bangalore preferred using phone calls in these locations. The differing preferences for these functionalities could be caused by a variety of factors. For example, typical mobile phone plans in these different cultural settings may favour text messages over phone calls. With 30% of respondents from Glasgow having unlimited text message plans, as compared to just 6% in Ban-

galore, it is not as surprising that respondents in Glasgow favoured text messaging. However, respondents from Glasgow also had an average of 447 free minutes per month. In contrast, 23% of respondents from Bangalore had phone plans that charged per minute, with an average cost of 0.5 rupees (roughly 0.01 in US $) per minute. These different allowances don't fully explain the observed differences between respondents, and interactions with mobile phones usually follow some rules of social etiquette. This etiquette may limit the appropriateness of making noise or require that one pays attention to the current situation. For example, individuals in a lecture theatre are expected to remain quiet and pay attention during a class. Rules of etiquette may be formalised in areas where phone use is strictly forbidden, such as in airplanes or courtrooms. These rules of etiquette, both implicit and explicit, vary between different cultural settings and influence the amount of mobile interaction that is acceptable in those settings. Additionally, factors such as personal and social values around communication, considerations about privacy or the perceived disruption for spectators might also play a role in the differences observed in this survey. For example, Geert Hofstede's cultural dimensions[1] describe cultural differences in terms of five general categories that describe society (Hofstede, 2011). The ways that these general dimensions could correlate to mobile phone practices has been explored in a variety of studies, such as those focusing on mobile user experience (Walsh *et al.*, 2010) and those focusing on mobile phone adoption (Swar *et al.*, 2009).

Studying existing practices and etiquette with respect to mobile phone usage across cultural settings provides a solid basis for looking at how new interactions might be accepted. These results demonstrate some of the preferences and mobile phone etiquette differences for respondents in Glasgow and Bangalore, but they do not explain why these preferences exist or how they will affect the *adoption* of multimodal interaction techniques. The following section presents a focus group study that builds on this quantitative data to explore qualitatively how individuals make decisions about the acceptability of new interaction techniques, and how they place them into a usability context. This next step in the design process, where participants are asked to imagine a usability context, which includes a setting, task, and defined series of inputs and outputs, include evaluations of *acceptance* and *adoption.*

5.3 Putting Interaction into Context: Evaluating Acceptance and Adoption

To begin looking at the acceptance and adoption of new multimodal interaction techniques in these different cultural settings, a focus group study was completed that combined multimodal acceptance surveys with group discussions about adoption scenarios. Similar to the focus groups described in the previous chapter, this study *used experience prototypes* (Buchenau and Suri, 2000) in order to familiarise participants with a diverse set of possible interaction techniques. The first goal of these focus groups was to identify differences and similarities in the acceptability of different interaction techniques in different cultural settings and understand the motivations for accepting or rejecting interaction techniques. The second goal was to better understand participants' approaches to placing interactions into context and imagining

[1] More information: http://www.geert-hofstede.com/

Figure 5.4: Screenshots from video prototypes for the Glasgow (right) and Bangalore videos (left). These portray the ”crazy” gesture (top) and the device scanning gesture (bottom).

how interaction could be successfully integrated into everyday life. The results demonstrate how preferences stem from the same underlying motivations, and how these motivations influence the ways that participants placed multimodal interactions into context.

5.3.1 The Prototypes

Video prototypes were used in order to familiarise participants with the interaction techniques in these focus groups. A culturally appropriate actor was used for each set of focus groups in order to make the videos more familiar for participants and limit bias. Example screenshots can be seen in Figure 5.4. Each video lasted from three to ten seconds and portrayed the given interaction technique three times. Each video was recorded in a plain setting showing a neutral coloured wall and a small desk, similar to the video prototypes used in Chapters 3 and 4.

5.3.2 The Interaction Techniques

A total of fifty interaction techniques in five categories were examined during each focus group, as shown in Figure 5.5. These five categories were device-based gestures, body-based gestures, emblematic gestures (Kendon, 1997), voice commands, and audio feedback. These represent a subset of the categories described in detail in Chapter 4 with the addition of au-

dio feedback. Device-based gestures included those that involve directly manipulating a device, such a shaking and squeezing a phone. Body-based gestures included any actions that move a part of the body without directly manipulating a device, such as wrist rotation or clapping. Emblematic gestures included any gesture that has a standalone meaning outside of the context of speech, such as a thumb up or shrugging gesture. Voice commands included any speech-based interaction directed towards a device, such as queries or commands for the device. Audio feedback included any speech or non-speech sounds generated by a device, such as chimes or synthesised speech.

5.3.3 The Focus Groups

Each focus group consisted of three distinct phases. The full focus group schedule can be seen in Appendix D.2 and the worksheets filled out by participants can be seen in Appendix E.2. The first phase involved viewing the interaction techniques and becoming more familiar with them. Each interaction technique was shown in a random order and participants were asked to rate its acceptability for use on the pavement using a five point Likert scale. This location was used as a baseline for acceptance based on the results presented in Chapter 3, where pavements were shown to be one of the least acceptable locations for interaction. After all the techniques had been demonstrated, participants were asked to rank each technique and indicate whether they would accept or reject each one in their daily lives. The second phase of the focus group involved talking about initial preferences and reactions to the techniques outside of a specific usage context. The final phase involved evaluating the appropriateness and usability of the interaction techniques within a specific set of usage contexts, where participants were asked to describe how the interactions would work in real world scenarios.

5.3.4 Results

Each focus group consisted of two to five participants organised into groups by gender and occupation, either 'students' or 'working adults.' A total of four groups were completed in Bangalore and four in Glasgow with a total of thirty participants. Each focus group lasted approximately ninety minutes. These results are based on the worksheets filled out during the focus groups and transcripts of the discussions.

Different Preferences: Differing Concerns for Different Cultural Settings

When looking at initial acceptance, participants in Bangalore and Glasgow had different preferences and concerns when accepting or rejecting these multimodal interaction techniques. Comparisons between the average rejection rates between participants in Glasgow and Bangalore show some interesting differences. Using a two-tailed Mann-Whitney test (Corder and Foreman, 2009) to compare non-parametric data, the rejection rates are significantly higher from participants in Bangalore ($p < 0.03$, medium effect size $r = 0.4$). These rejection rates are shown in Figure 5.6, where the average rejection rate from participants in Bangalore is 10% higher than for participants in Glasgow.

Interaction	Category
Wrist Rotation	Body Based
Snapping	Body Based
One Foot Tapping	Body Based
Phone Shaking	Device Based
Phone Squeezing	Device Based
Phone Twisting	Device Based
Turning Phone Over	Device Based
Phone Flick	Device Based
Phone Tilting	Device Based
Tapping Phone	Device Based
Stroke Back of Phone with 2 Hands	Device Based
Hold Phone to Ear	Device Based
Spinning Phone	Device Based
Hold Phone on Shoulder	Device Based
Pointing Phone at Object	Device Based
Scanning with Phone	Device Based
Hold Phone on Hip	Device Based
Stroke Back of Phone with 1 Hand	Device Based
Say "Set Alarm for 8 AM"	Voice Command
Say "Where is my Phone?"	Voice Command
Say "Do I have messages?"	Voice Command
Say "Next"	Voice Command
Say "Delete"	Voice Command
Say "Call Sri/Mary"	Voice Command
Say "Reply"	Voice Command

Interaction	Category
Pointing at Someone	Body Based
Head Nodding (Side to Side)	Body Based
Writing on your own hand	Body Based
Clapping	Body Based
Swipe	Body Based
Head Nodding (Up and Down)	Body Based
Two Feet Tapping	Body Based
Point to an Object	Body Based
Gaze with Eyes	Body Based
Finger Chording on Leg	Body Based
Pointing at Someone	Body Based
Belt Tapping	Body Based
Gaze with Head	Body Based
Gesture "Come Here"	Emblematic
Gesture "Blah Blah Blah"	Emblematic
Gesture "Crazy"	Emblematic
Gesture "Telephone"	Emblematic
Shrugging	Emblematic
Thumbs Up	Emblematic
"Vrishti/Martin sent you a photo"	Audio Feedback
"You have a missed call from Sri/John"	Audio Feedback
"Sanika/David sent you a text"	Audio Feedback
Listen to Musical Chime	Audio Feedback
Listen to Alarm Chime	Audio Feedback
Listen to Bell Chime	Audio Feedback

Figure 5.5: Set of interaction techniques used in focus groups.

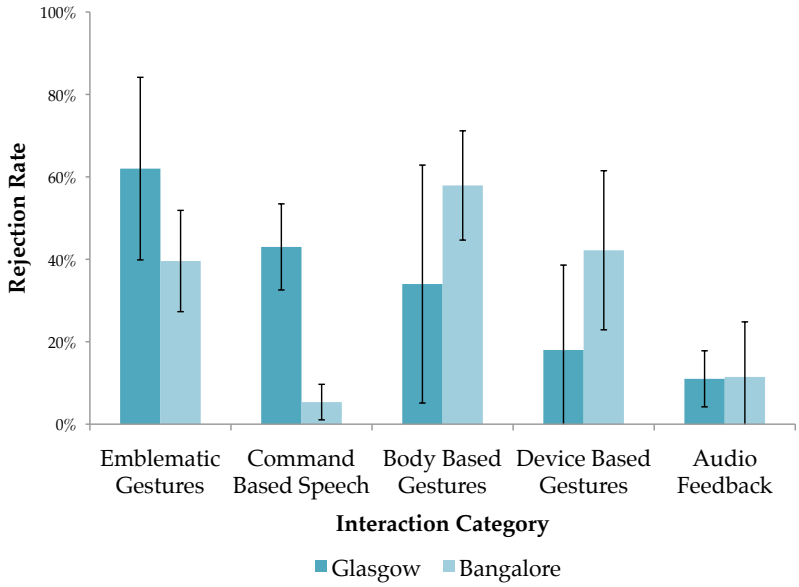

Figure 5.6: Comparison of rejection rates by category for users in Glasgow and Bangalore.

While overall rejection rates were higher for participants in Bangalore, the results show that there are some interesting differences between rejection rates for different interaction categories. One category where participants in Bangalore had significantly lower rejection rates was command-based speech. Using a two-tailed Mann-Whitney test (Corder and Foreman, 2009), the results show that participants in Bangalore had significantly lower rejection rates for command-based speech ($p < 0.02$, large effect size $r = 0.5$). Additionally, participants in Bangalore had significantly lower rejection rates for emblematic gestures ($p < 0.02$, large effect size $r = 0.5$). These quantitative differences demonstrate that participants from different cultures preferred different interaction categories. The following qualitative analysis looks at *why* these differences were observed.

During the focus groups, participants discussed why they accepted and rejected interaction techniques, describing their own motivations for liking or disliking interactions and how they felt about performing these in front of spectators. These discussions yielded a set of motivations for interaction technique acceptance or adoption, where some of these motivations were expressed in different ways in these cultural settings, leading to different preferences. These expressions represent the most prevalent approaches to the given motivation, where the majority of participants from the given cultural setting shared this approach. One of the most common concerns for participants was attracting unwanted attention from spectators. However, the way that this motivation was expressed in different cultural settings led to interesting differences in preferences.

- **Motivation (M1):** I want to perform comfortable and acceptable actions that do not attract a lot of attention.
- **Expression of M1 in Bangalore –** I like actions that do not disturb others.
- **Expression of M1 in Glasgow –** I like actions that are subtle or surreptitious.

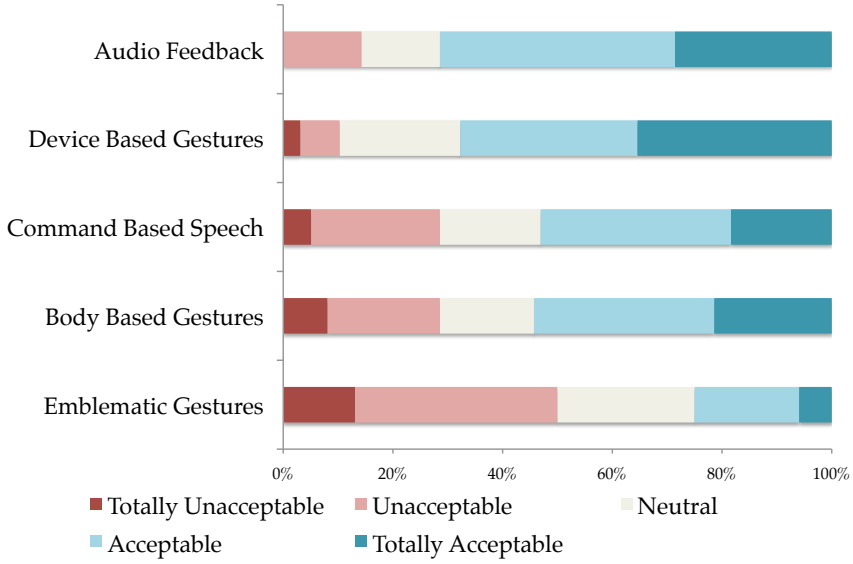

Figure 5.7: Likert-style responses for participants in Glasgow for the acceptability of interaction techniques, grouped by category.

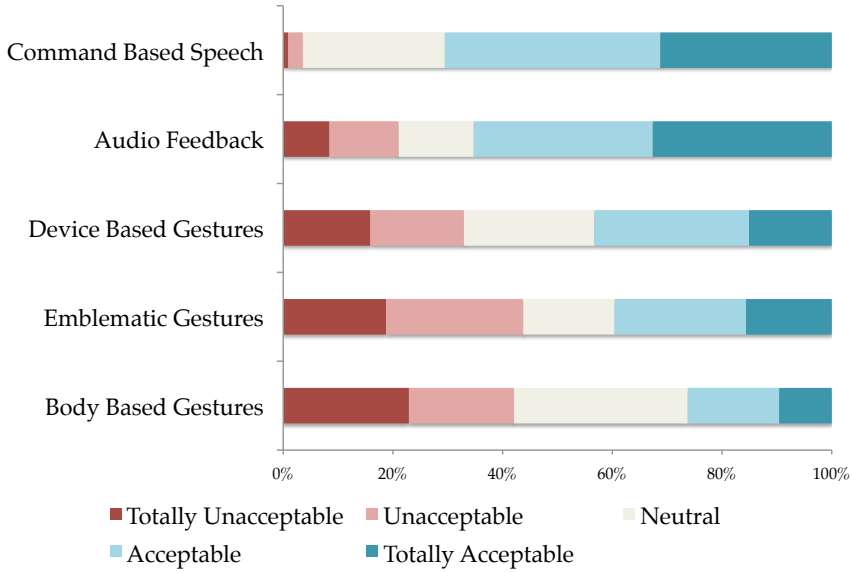

Figure 5.8: Likert-style responses for participants in Bangalore for the acceptability of interaction techniques, grouped by category.

The desire to interact with a multimodal interface without attracting unwanted attention has been demonstrated in the previous studies described in this thesis, and was an equally important concern for participants in this cross-cultural study. This motivation, expressed from two different perspectives, represents a common concern that manifests itself in different ways. When discussing which interaction techniques might be acceptable, one participant from Glasgow stated that "it would be the gazing with the eyes, the device twisting, and the belt tapping. The things you could do really surreptitiously without anybody noticing." When describing unacceptable interactions, one participant in Glasgow summarised these unacceptable actions as "the big movements." Another participant in Glasgow stated that "I don't like the swiping. It just looks weird." Participants in Glasgow more often considered the visibility or audibility of their actions when determining acceptable or unacceptable interaction. In contrast, participants in Bangalore described acceptable interactions from a different perspective. When discussing acceptable interaction techniques, one participant in Bangalore stated that voice-based interactions were acceptable because "there's no disturbance from these one liners. It doesn't disturb the public." When describing actions which were undesirable, these differing perspectives led to different ideas about what was unacceptable. For example, one participant in Bangalore stated that "foot tapping is always irritating. There may be hundreds of people around you but actually one person will be next to you. It might not bother the 100th person, but it will bother the 2nd person next to you." The concern that actions could be disruptive for spectators was the primary concern when determining both acceptable and unacceptable interactions for participants in Bangalore. Although the underlying motivation is the same, actions that do not disturb others and actions that are subtle or surreptitious are not always the same. Hence, the same motivation leads the very different preferences when satisfied from different perspectives.

Another aspect of acceptable interactions was their possible acceptability and usability in everyday life. Participants discussed how they would use multimodal interactions in different contexts, and again approached this problem from different perspectives that yielded different preferences.

- **Motivation (M2) –** I want to use actions that are acceptable in my everyday life.
- **Expression of M2 in Bangalore –** I need to use different actions for different locations.
- **Expression of M2 in Glasgow –** I need to use different actions based on who is around and what I am doing.

While both user groups considered social norms and standards when determining acceptability, the way these issues were considered and organised reflected the different cultural settings and different expressions of the same motivation. For example, participants from Glasgow were more concerned about issues of social influence, the opinions spectators might form about them, and whether or not others might be aware of the purpose of their actions. For example, one participant from Glasgow stated that "if you were with friends or family and there is this application and it's based on hand gestures and snapping or pointing, it wouldn't really matter if it was part of the game and everyone was in on it. Everyone knows what's going on compared to a busy street where it's full of strangers and no one knows what you are

doing." This participant looked at different contexts with respect to who might be there and how actions could be explained to spectators. In contrast, participants in Bangalore focused on the affordances that different locations had that supported or hindered multimodal interaction. For example, one participant from Bangalore stated that "at home we don't have to be silent, we can talk at any volume and the ring tone is okay at home." Another participant from Bangalore stated that "I do accept that [voice commands] in a meeting. Because [you might receive] important messages, that is what I feel and I do accept that in a meeting." Participants in Bangalore were also more concerned with the appropriateness of communicating personal information in different contexts. For example, several of the interaction techniques communicated which personal contact was sending a message or calling. One participant from Bangalore stated that "something might be private. Something which you don't want to disclose in front of your family or in from of your office people. 'You have a missed call from so and so, you got a text from someone,' these things don't go well." This motivation demonstrates again that participants in these different cultural settings had similar motivations that were addressed in very different ways.

There were also similarities in the ways that participants determined social acceptability. For example, participants in both user groups considered factors such as social norms and standards, mobile phone etiquette, the direction of gestures with respect to spectators, and how interactions might be useful or productive. For example, when discussing voice commands one participant from Bangalore stated that "just saying it is very easy, especially when you're short of time." Similarly, another participant from Glasgow described how wrist rotation might be used, stating that "it has got a familiarity angle. I could see that being quite natural, quite an easy one, possibly easier than hitting a button." These results demonstrate some of the interesting similarities and differences participants described when discussing how acceptable multimodal interactions might be in everyday life.

Placing Interactions into Context: Cultural Approaches

A major portion of these focus groups was dedicated to discussing how inputs and outputs could be paired together into a usability context for real world usage. Participants discussed a variety of application areas, as listed in Appendix D.2. For each application area, participants discussed which interaction techniques they would use to control the given application and how the feedback would work with the interaction. One of the application areas that participants discussed was how they would go about checking the time on their phone. Figure 5.9 shows how the most commonly discussed interaction techniques played out and how these interactions were described.

The ways that participants discussed this scenario provide interesting insights into how they made sense of these multimodal techniques and how these interactions might be used in daily life. For example, participants in Bangalore were more likely to choose different *input* techniques in different contexts, where participants from Glasgow were more likely to choose different *output* techniques in different contexts. For example, one participant from Glasgow discussed how feedback should switch from private audio to visual based on how the phone is set up. This participant stated that "if the headphones are plugged in ... you've set it to actu-

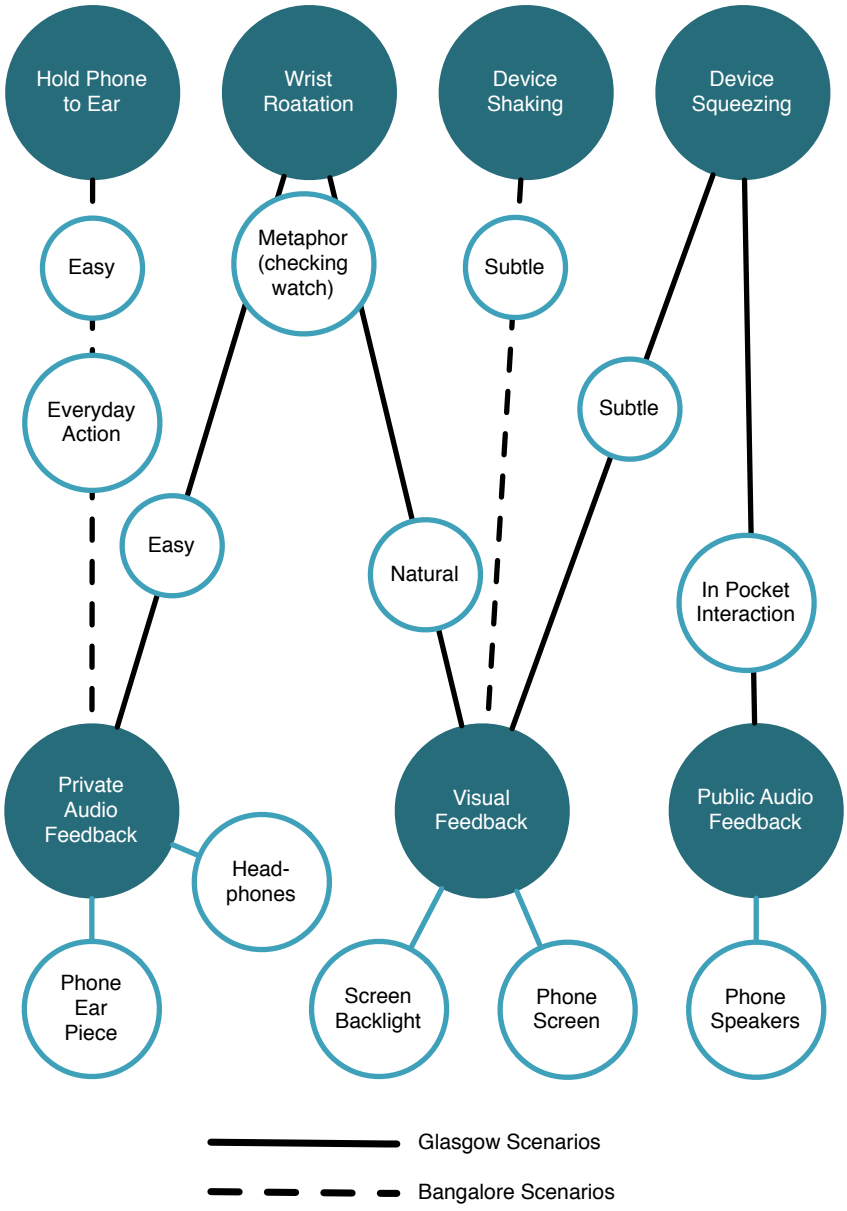

Figure 5.9: Scenarios discussed for pairing input with output for checking the time on a mobile phone. The top row of dark circles represents the chosen input techniques and the bottom row of dark circles represents the chosen output or feedback. White circles describe how the interaction techniques were described.

ally say the time. So you could have the phone tell you the time into your headphones or if you don't have the headphones connected then just light up the screen and show the time." The consistency of input rather than output was common in scenarios discussed by participants in Glasgow, where participants preferred to have one consistent command to use in different settings. In contrast, participants in Bangalore favoured different input and output techniques for different contexts. One participant from Bangalore stated that in an "outdoor public place you can hold the phone to your ear so that it tells you the time." The same participant discussed using this functionality in a workplace, and stated that "when you're at work ... you have your phone handy, just squeeze it and the time is displayed." While both of these output styles are unobtrusive and do not make sound, different input techniques were favoured in different locations.

When discussing this scenario, participants from different cultural settings showed a variety of similarities, considering many of the same factors when determining how this interaction should work. For example, participants from both Glasgow and Bangalore considered the location and proximity of the device, discussing how the interaction might work relative to the device. For example, participants from both settings described how feedback should change if the device was in the hand versus in the pocket. Participants also discussed how multimodal interaction could make mobile phone tasks easier either by shortening a series of steps into a single action or making commonly used functionalities easier to access. Participants also discussed a variety of ways that gestures could be used for continuous or discrete control, with some gestures described as appropriate for both of these interaction styles. These similarities in approach demonstrate some of the key concerns that can be addressed by designers in order to give an interface wider appeal for diverse user groups.

5.3.5 Discussion

The results of these focus groups demonstrate some clear differences between the preferences of participants in Glasgow compared to Bangalore for multimodal interaction techniques. Although participants in Bangalore had significantly higher rejection rates for multimodal interaction techniques overall, participants from Bangalore had significantly lower rejections rates for command-based speech and emblematic gestures. When comparing these results to the cross-cultural survey results, there are several similarities that explain some of these differing preferences. For example, respondents from Bangalore made significantly more phone calls than respondents from Glasgow, and used phone calls in public more often than respondents from Glasgow. Given this existing mobile phone usage, it is not surprising the participants from Bangalore found voice-based commands more acceptable than participants in Glasgow. If having mobile phone calls and making noise in public is more acceptable for participants in Bangalore, using voice-based commands should also be relatively acceptable.

Participants often made decisions about acceptability based on the same motivation, but expressed differently in different cultural settings. In both cultural settings, participants described the desire not to attract unwanted attention to themselves while interacting with a multimodal interaction. For participants in Bangalore, this was expressed as a desire not to disrupt others. In Glasgow, this was expressed as a desire to perform subtle or less noticeable

interactions. Although there is some overlap between these expressions, preferences based on these were relatively different. Participants in Bangalore described how voice commands would be less disruptive for others than unusual or unexpected gestures. In contrast, participants in Glasgow discussed how actions could be too small for others to notice or hidden as everyday actions.

When placing interactions into a usability context, participants from different cultural settings demonstrated some differences in their approaches. Participants in Bangalore tended to discuss scenarios where different input techniques should be used in different scenarios, where participants in Glasgow favoured consistency of input and different output for different scenarios. Differing preferences aside, participants in Glasgow and Bangalore showed more similarities than differences when discussing how multimodal interactions should be placed into context. Participants in both cultural settings considered the location and proximity of the device, the ways that multimodal interactions could make mobile phone tasks easier, and how interactions could be used in discrete and continuous control contexts. The similarities suggest that once preferences are established, the applications of those interactions to a usage scenarios will share many of the same motivations and rationales.

The results of these focus groups identify interesting preferences that are incorporated into the final phase of this design. The main input techniques evaluated in these focus groups included emblematic gesturing, device-based gesturing, body-based gesturing, and voice-based commands. The use of audio feedback was widely accepted and therefore should not present any significant issues if incorporated in an application. The following section presents an in the wild user study with an application design based on the results of the survey and focus group studies.

5.4 Looking at Appropriation: Usage in the Wild

When looking at the appropriation of multimodal interfaces, evaluations should investigate how users choose to perform actions, how they feel their actions are perceived by others, and when and where they choose to interact. Often, it is not possible or effective to evaluate these kinds of issues in a lab setting and user studies must be brought out into real world contexts. However, evaluations of multimodal interactions 'in the wild' present challenges, especially for mobile interfaces that will be used while 'on-the-go' in a variety of contexts. Traditional ethnographic methods of observation are typically not reasonable or practical if interaction occurs naturally and sporadically throughout the day. There are also technical challenges since such a deployment would be subject to a huge variety of contexts, noise levels, and other situational background disturbances that could affect sensor accuracy, usability and experience. For example, recognition accuracy rates for gesture-based interactions can often be affected by unexpected circumstances, such as sensor noise from walking or riding public transport. Additionally, multimodal feedback may be difficult to perceive in different settings where noise or distractions limit focus. These challenges make longitudinal or 'in the wild' studies difficult for multimodal mobile interfaces.

The design of the user study presented in this chapter approached these issues by carefully

creating a multimodal interface for practical use in a variety of contexts. This study combined traditional usability metrics, instrumented usability techniques (Crossan *et al.*, 2005), and qualitative methods to gain the best possible view into the usage and experience of the application. The user study revolved around a gesture controlled mobile game application that allowed for different modes of gesture control for continuous and discrete interaction. This application, called Mo!Games, was designed to incorporate gesture-based input with audio, visual, and vibrotactile feedback. Of the two games included in the application, one used continuous gesture control and one used discrete gesture control. Both games also support different modes of gesturing, including gesturing with the wrist, the head, and the device itself. This study was completed concurrently with participants in Glasgow and Bangalore. The results of this study explore participants' perceived usability of the interactions, their preferences and appropriations of the different gesture modes, the experience of gesturing as part of a mobile interface in everyday life, and the cultural differences between these user groups.

5.4.1 The Mo!Games Application

The Mo!Games application was designed for use in everyday life for an 'in the wild' study. Each game lasted roughly one minute, allowing participants to easily pick up the application for short playing sessions throughout the day. It included an interface for leaving voice notes, where participants could leave feedback, comments, or relate stories during the course of the study. Additionally, the application included help screens with information about how to use each gesture mode, how to position the external sensor used in the study, and how to troubleshoot issues with the external sensor and the application itself. The application also included an 'achievements' page where participants could keep track of their daily usage and gain achievements based on how many games they had played, how well they had been doing, and other simple aspects of the games. This was included to encourage participants to play the games and be aware of how long they had played thus far.

The first game in the application was a trajectory/toss game, where users had to toss marshmallows onto a target, shown in Figure 5.10. In the Marshmallow Toss game, the angle of the toss is controlled using gestures with continuous visual feedback. The tilt or rotation corresponds to the angle of the launch, and a marshmallow is tossed after the user taps the touchscreen. Vibrotactile and audio feedback is provided when the marshmallow successfully hits the target, with a different vibration pattern when the target is missed. The goal of this game is to gain as many successful hits within one minute, with achievements for actions like most successful hits, best hit rate, and fastest time to successfully hit five marshmallows. The target moved randomly around the shooting area after each toss.

The second game was a 'Simon Says' style game where each game began by presenting a list of icons corresponding to gestures that must be memorised within 30 seconds, including rotate left, rotate right, and flick downwards, shown in Figure 5.11. Once 30 seconds had passed, the list disappeared and the user had to perform the gesture list correctly from memory within 30 seconds. After three successful games, the list size increased by one. If users made a mistake, they moved back one level. The interface provided visual, audio and vibrotactile feedback when users were repeating the gesture list based on the correctness of their actions. The goal of

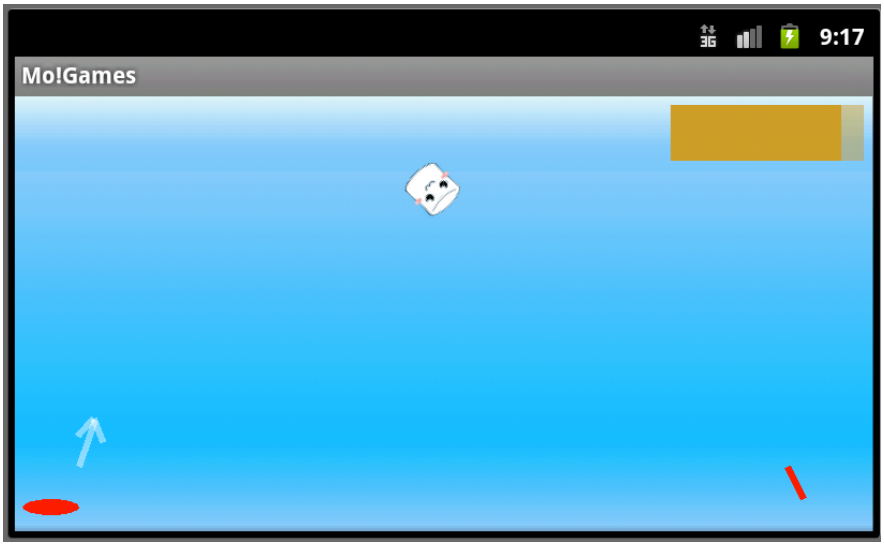

Figure 5.10: Screenshot of the beginning of a Marshmallow Toss game. Marshmallows are targeted using the white arrow on the left side of the screen, and tossed at that angle when the screen is tapped. The target on the right side of the screen is a red stick, which moves after every toss. The orange bar in the top right is a timer for the game.

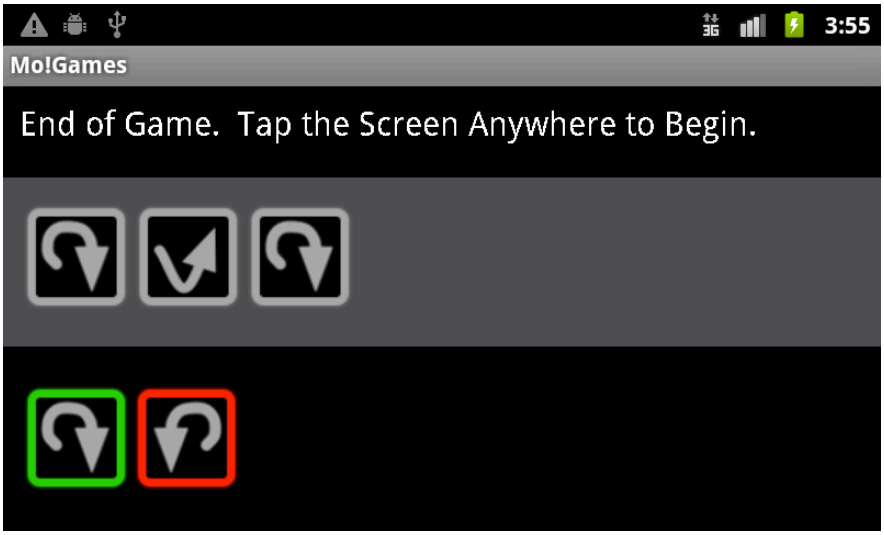

Figure 5.11: Screenshot from the end of a Simon Says game. The grey list of icons in the top row are presented during the first half of the game. These must be memorised and repeated during the second half once the list is no longer displayed. As the gestures are repeated, the corresponding icon is displayed in the bottom row with a red or green border to indicate correctness. At the end of the game, as shown in this screenshot, both lists are displayed together.

this game was to successfully complete as many rounds as possible without making mistakes, with achievements such as passing different levels and best success rates.

Both of these games could be completed using one of three gesturing styles, including gestures with the wrist, the head, and the device. For the Marshmallow Toss, the angle of the trajectory was continuously controlled by the angle of the wrist, head, or the device itself. For the Simon Says game, rotate left, rotate right, or flick could be performed with the wrist, head, or device. Thus, each game can be completed using the same types of movements performed with different parts of the body. The application was deployed as an Android application on a Google Nexus One[2] paired with a SHAKE sensor pack[3]. The SHAKE is a Bluetooth enabled sensor that includes an accelerometer, gyroscope, and magnetometer. The external sensor pack was only used for the head and wrist conditions with the internal Nexus One sensor used for the device condition.

5.4.2 Gesture Design

Gestures were chosen as the primary input technique for both practical and social reasons. Because this application was used independently by participants in their everyday lives, gesture control provided a relatively robust input technique that was less likely to be affected by ambient noise, users' regional accents, and other environmental factors that would make successful voice input difficult in a mobile application. Additionally, mobile speech recognition would require extensive processing power or a constant Internet connection, both of which would have been impractical given the locations and application used in this study. Significant delays in processing would have inhibited game play and the availability of reliable data connections was poor. Thus, gestures were chosen as the primary input technique since these could be efficiently recognised on the device in real time, and voice input was limited to the voice note functionality. However, participants were asked about their comfort in using the voice note functionality and asked about how they would feel using voice commands for this application during interviews at the end of the study.

The gestures used in this study provided both continuous and discrete control using similar actions. These gestures were also designed such that it would be possible to perform the same movements with different parts of the body. In the Marshmallow Toss game, tilt left and tilt right were used to continuously control the angle of the trajectory using visual feedback for control. In the Simon Says game, discrete gesture control was used when repeating back patterns of gestures.

The selection of the these three gestures was based on the acceptance of these gestures in the previous focus group study and their ability to be performed with different parts of the body. Based on the results of the focus group, device-based gestures were the most acceptable input technique for participants from Glasgow and the second most acceptable input technique for participants from Bangalore. Therefore, device rotation left and right and device flick were supported in this application. These three gestures were also translated to the wrist and the head. With respect to the acceptability rankings for body-based gestures, wrist rotations were

[2]More Information: http://www.google.com/phone/detail/nexus-one
[3]More Information: http://code.google.com/p/shake-drivers/

the fifth most acceptable body-based gestures for participants in Bangalore with a rejection rate of 50%, and the seventh most acceptable for participants in Glasgow with a rejection rate of 29%. With respect to head nodding side to side, this was the second worst body-based interaction with a rejection rate of 75% for participants from Bangalore, and the fourth worst body-based gesture with a rejection rate of 43% for participants from Glasgow. These selections provide a good diversity of relatively acceptable and unacceptable interactions.

Recognition Design

Because this application was used while mobile in real world settings without an experimenter present, it was important to keep the gesture recognition simple and robust. The gesture recognition was designed using accelerometer-based sensing that calculated the roll and the current energy or total movement of the sensor. These values were automatically calculated by the SHAKE drivers. Using a thresholding technique combined with a simple state machine, rotation left and right were recognised by rotating a sufficient amount in the appropriate direction and returning to a central position. The flick gesture was recognised by calculating the current change in energy and testing against a threshold. Small adjustments to these thresholds were made for the head mode, so that slightly less movement was required for successful recognition.

5.4.3 The Study

The focus of this user study was to compare the adoption and possible appropriation of three multimodal conditions in two cultural settings. Participants were asked to use the Mo!Games application for six days, completing three gesture conditions of two days each. Participants were not required to complete a certain amount of usage, but were encouraged to use the system in as many different locations as possible. As a guideline, participants were asked to use the system for roughly twenty minutes per day. The order of the gesture conditions was counterbalanced between the participants and was controlled and changed automatically by the game application. Throughout the course of the study, participants were asked to tag their current location from a list of possible locations consisting of: home, work, pavement or sidewalk, on public transport, on private transport, in a shop, in a restaurant, or other with an optional text input field. These location categories were used in order to provide useful information about the participants' current context without compromising privacy or security. For this reason, the application did not log GPS or any other personal information that would identify participants' current physical location. Additionally, participants were asked to provide feedback about their experiences by recording voice notes. All activities within the application were logged, including any click or touch input, the path taken through the game screens, and all gesture interactions. Logging of the gesture interactions included successful and unsuccessful actions as well as the progress of the current game. At the end of the study, participants were interviewed about their experiences with the game, their preferences for different gesture modes, and discussed the incorporation of additional interaction techniques, such as voice control, as possible additions to the mobile interface. The full list of interview questions can be seen in Appendix C.3.

Because of the inherent challenges with completing an in the wild investigation, the design of this study was based loosely around another successful longitudinal evaluation of multimodal feedback completed by Hoggan and Brewster (2010). Important aspects of that study were incorporated into the one presented here, such as giving participants specific equipment to use over the period of roughly a week while going through a set of multimodal conditions each lasting two days. The work completed by Hoggan and Brewster also demonstrated the success of asking users to tag their current location whenever using the system, and allowing them to provide feedback throughout the course of the study using voice notes recorded within the application.

5.4.4 Results

The results of this user study are based on the logs generated by participants, voice notes and other feedback provided by participants, and semi-structured interviews completed at the end of the study. This includes the results for a total of twenty users, with ten each from Glasgow and Bangalore. The participants ranged in age from 20 to 45, 7 were female and 13 were male. 10 of the participants were students, and 10 worked in technical office jobs. The participants were recruited through email and social networks.

These results are based on over 35.5 participant hours playing games, distributed throughout the participants' everyday lives. Participants spent an average of 106 minutes interacting with the application over the span of six days, spending similar amounts of time in each gesture mode. On average, participants interacted with the application for 40 minutes in the device mode, 37 minutes in the wrist mode, and 34 minutes in the head mode. This section will begin by discussing the cross-cultural comparisons between these two user groups, then go on to discuss overall results from both groups.

Cultural Differences in Input Technique Preferences

Participants from different cultural settings had different preferences for gestures modes. Figure 5.12 shows the number of participants that favoured each gesture mode, or had no clear preferences, describing different preferences for different games. Pair-wise tests using the non-parametric Mann-Whitney test for independent samples (Corder and Foreman, 2009) indicate there are no significant differences between the numbers of users preferring each mode in Glasgow and Bangalore. Overall, device, head, and wrist mode were ranked in a similar order for participants from these two cultural groups. However, the order of those ranks was significant. The non-parametric Friedman test (Cramer, 1994) for ordinal data shows a significant difference ($p = 0.0017$) between the ranks for different gesture modes for all participants. Pair-wise tests using the non-parametric Wilcoxon Signed-Rank (Corder and Foreman, 2009) test show significant differences between the ranks of gesture modes for all participants. Wrist mode was ranked significantly higher than head mode ($p = 0.05$, medium effect size $r = 0.4$). Device mode was rank significantly higher than wrist mode ($p = 0.04$, medium effect size $r = 0.4$). Device mode was also ranked significantly higher than head mode ($p = 0.0004$, large effect size $r = 0.8$).

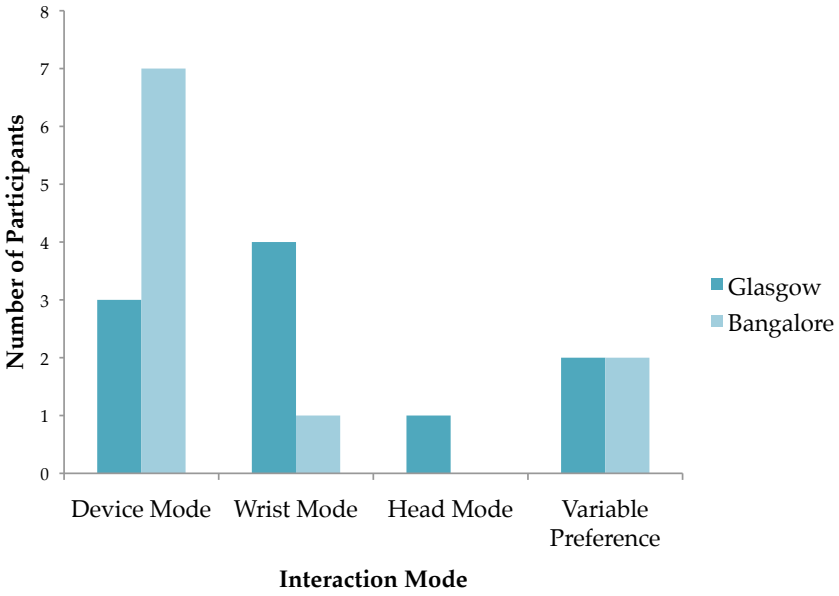

Figure 5.12: Number of participants that preferred different gesture modes, or described variable preferences based in the game or location, for participants in Glasgow and Bangalore.

Another area where differing preferences were observed was voice control. Although voice control in this application was limited to the voice note functionality, participants were interviewed about their comfort using the voice notes, and asked to imagine how they might use voice control as part of the game in place of the touch screen control. For example, the Marshmallow Toss used a touchscreen tap to launch the marshmallows, but this could potentially be replaced with a voice command in a possible version of the application. When discussing their experiences with the voice note functionality, half of the participants from Glasgow stated they were uncomfortable using voice, as compared to just one third of the participants in Bangalore. When discussing the possibility of using voice-based commands to interact with the application, the differing preferences became even more pronounced. Six out of ten participants from Glasgow stated that they simply would not use a voice command with their mobile phone, as compared to just one participant from Bangalore. Half of the participants from Bangalore said they would be happy to use voice controls, with a further four participants stating that they would consider using voice in certain situations. These results are comparable to the previous focus group presented in this chapter, where participants from Bangalore rejected just 5% of command-based speech interaction as compared to 43% for participants from Glasgow.

Participants in Glasgow were often worried that the voice commands would be inappropriate or embarrassing in the places where they most often interacted with the application. One participant from Glasgow stated that "I would prefer the tap, just because in the context where I played the game I felt it would have been intrusive to be speaking to something as opposed to just making funny motions." Another participant described how voice commands required more explanation for spectators, stating that "I probably wouldn't use it [a voice command]. Even in the house, I would feel a bit strange. I do use voice control on the phone occasionally, but I always find myself explaining myself first." Participants from Glasgow often thought that the initial embarrassment of performing the voice command would make the application

much less enjoyable. When discussing voice commands, another participant from Glasgow stated "I don't think I would've played as often. It would've been a chore." Another participant stated that "I would probably have only played it when I was by myself [if using voice commands]. I don't think I would've enjoyed that very much." In many cases, participants simply did not consider how voice commands might be used because they did not find them acceptable to perform in the first place.

In contrast, participants in Bangalore were more interested in how voice commands could increase the usability of a mobile interaction and found voice commands generally acceptable in public places. For example, one participant from Bangalore stated that "one of the problems was holding the device and tapping at the same time. Saying 'shoot' would solve some of that problem. Socially, at home and office I wouldn't be bothered about speaking. If I'm with friends, and they see me playing this game, it's not a problem at all." Another participant discussed how voice commands could increase the usability of certain scenarios. This participant stated that "personally, I would like to have both the controls, based on the situation. For shooting in head mode, I would prefer giving a voice command. Otherwise, I am comfortable tapping, like in device mode. I would like the control to switch it for this mode or that mode. I would like to have both."

Perceptions about Accuracy and Usability

During the post-study interview, participants described how the accuracy of the gesture modes varied. Even though the successful hit rates were not significantly different, having similar means and standard deviations as shown in Figure 5.13, participants had strong opinions about which modes had the best control and accuracy. The majority of participants felt that the device mode was the most accurate, closely followed by the wrist mode. Head mode was often described as the least accurate and usable mode. Given that the recognition techniques were nearly identical for each gesture mode, these different perceptions of accuracy and usability between gesture modes are based on a variety of experiential factors. For example, the flexibility and dexterity of the head and wrist are very different. However, participants discussed a variety of issues that went beyond simple physical constraints.

Perceived Size of Movement – Participants described how some of the gesture modes did not seem to respond as quickly as others, where the perceived size of the movement had a strong influence on how well the system seemed to respond. Even though the recognition techniques required roughly the same amount of displacement for each of the three gesture mode, participants often felt that some modes required more movement than others. For example, when describing the head gestures, one participant stated that "it felt like it was the least accurate, it felt like you had to make big movements, it didn't feel good to use and I didn't seem to pick it up as easily." Differences in the physical constraints of the different modes and how the gestures were performed made some modes appear to require more movement. For example, the head gestures often seemed to require more movement because head gestures used a larger proportion of neck flexibility than wrist gesture required of wrist flexibility. In other cases, participants performed extraneous movements while completing gestures and mistakenly thought that those interactions required larger movements. Previous work has

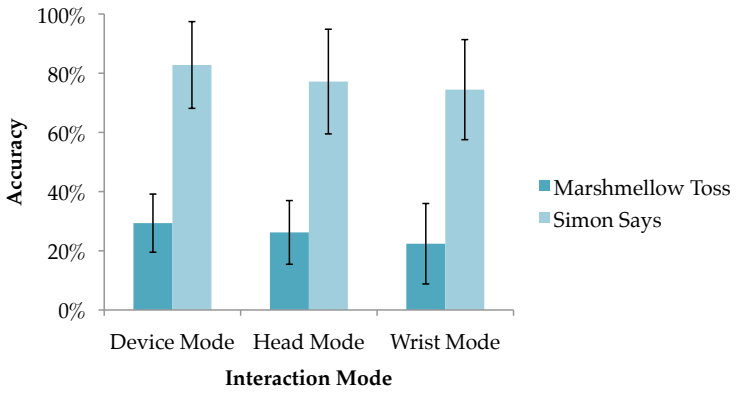

Figure 5.13: Successful hit rates for each gesture mode. Error bars show one standard deviation.

investigated how users develop misconceptions about the actions required for successful interactions (Freeman *et al.*, 2009), where these results demonstrate how these misconceptions can have significant effects on the acceptability of interaction.

Isomorphism Errors – Certain gesture conditions seemed more prone to isomorphism errors (Williamson, 2006), where there was a disconnect between a user's perception of what a system is recognising and what a system is actually recognising. Particularly with the wrist condition, participants had difficulty identifying and repeating successful performances of the tilt left and tilt right gestures. For example, one participant stated that "initially I could not get the hang of this, though I could move it around easily the exact direction was difficult so I could not score."

Lack of Familiarity – Participants found actions which interacted with the device in an unusual or unfamiliar way to be less accurate and more difficult to use. One participant stated that "I'm not used to using head or wrist to activate anything with the device. So with the device, I found it easier. I found device mode easy and comfortable." The strangeness and unfamiliarity of an interaction gave participants much less confidence when performing interactions and led to perceptions of decreased accuracy.

In some cases, participants recognised that the 'most accurate' mode was not their favourite. For example, one participant stated "My best score was in the wrist mode, but actually that's my least preferred mode." Another participant stated that "I felt the most control with the head mode, but I think it would be the one I would be least likely to use." These situations highlight the need to explore how desirability and usability affect actual usage behaviours.

Gesture Mode Preferences

When looking at the overall gesture mode preferences for both user groups, the device mode was the most commonly favoured with ten of twenty participants ranking this as the best. This is not surprising, as device based gestures have been shown to be the most acceptable types of gestures in Chapters 3 and 4. However, those participants that preferred other gesture

modes provided interesting insights into the reasons why body-based control might be more desirable. Five of the twenty participants favoured wrist mode, one favoured head mode, and four had variable preferences depending on the game or the context. During the interviews, participants described different aspects of the body-based interactions that made them more desirable than the device-based ones.

Better Control – Participants described how body-based gestures gave them a better sense of control and greater confidence while performing interactions. One participant stated that "I found that in wrist mode, it was responding much better. Also, when you are in the device mode, you have a fear that the device might fall while you are doing the gestures. That's not a problem with the wrist mode, once it [the sensor] is tight on your wrist you can move your wrist around as much as you want." This sense of confidence and control allowed participants to explore the experience of gesturing more thoroughly, leading to a more enjoyable experience overall.

Hiding Interaction – In some cases, participants appropriated body-based actions by disguising or hiding their performance. Although the application was designed to support performative actions, the ability to perform more subtle interactions or hide interactions altogether made body-based interactions more comfortable in public places. One participant described how the body-based gesture modes could be performed with subtlety. This participant stated that "for the marshmallow game, I could just tilt my head slightly and tap, and no one could know what I was doing." Another participant described how using the device-based gestures to play the game was more obviously playful and thus inappropriate for a work place than the body-based modes, which could be performed surreptitiously. Although the device-based interactions made it clear that actions were directed towards an interface, the obvious use of an interface was not always desirable.

Novel Experience – Some users preferred body-based interactions simply because this was an enjoyable and new way to interact with a mobile phone. One participant stated that "I quite like the [wrist] gesture because it was novel, and I thought the device mode was less interesting." Participants enjoyed experimenting with new interaction techniques simply for the sake of getting a new experience. Although this effect might wear off over time, other factors might encourage sustained use. For example, an interaction might be initially accepted for its novelty then continue to be used because it increased ease of use and allowance for better control. This may be a way for an initially less socially acceptable gesture to become accepted into normal life.

Some participants described how their preference for a gesture mode was based on the game they were playing, with body-based interactions favoured for continuous interaction and device based interaction favoured for discrete interaction. One participant stated that "it felt odd doing the movements so jerkily, at least with the marshmallow game it was a smooth moving of the head as opposed to a twitch." For discrete control, body-based interaction often had a much more rigid and abrupt appearance, making these interactions less acceptable and comfortable than their continuous counterparts.

Gesturing in Everyday Life

Because this application was used in participants' everyday lives, they had the opportunity to experience performance in a variety of settings. Figure 5.14 shows some of the most commonly tagged locations by participants throughout the study. Although the majority of games were played either at work or home, some were also played on the pavement, on public transport, and at restaurants. In these scenarios, participants provided detailed feedback about the effect of spectators and their feelings about how their performance appeared.

Participants described how spectators could make them feel uncomfortable performing to the point that they would stop altogether. One participant stated that "I was sitting on the steps in the mall and some people who were passing by stopped and looked at me. And then I didn't continue that game. I stopped it, and took off my cap and put it in my bag. So obviously I'm not comfortable playing it in head mode in public." Even the anticipated effect of unwanted attention from spectators was enough to prevent interaction in public places. One participant stated that "I made a very conscious decision that the only one I would ever play outside of work or home would be the device mode." Another participant also described how only certain modes were acceptable in public locations: "I played the marshmallow game even in the restaurant, but I would not have felt comfortable playing it in the head mode or the wrist mode. Since I was in the device mode, it was more comfortable to be playing it in public places." For some participants, these initial barriers to interaction meant that they were not able or willing to experience the application in certain settings. This deployment specifically included relatively unacceptable interaction techniques, so rejection of the application in certain locations is not surprising. However, these 'missed opportunities' for interaction would be detrimental to a large scale deployment if they were unexpected. This demonstrates the importance of completing early evaluations of social acceptability to identify which interaction techniques are simply unacceptable for usage in everyday life.

Another factor that participants discussed was the greater social acceptance of interactions, with consideration to how this kind of acceptance might change over time. For example, one participant described how the ubiquity of smart phones meant that moving a device in a strange way is not as unacceptable as it used to be. This participant stated that "I was quite aware of moving my head around in an unnatural way, as where no one would say anything if you are doing that with a device." Other participants described how interactions would become more familiar and more acceptable as more people began to use them. When describing the use of voice commands, one participant stated that "telling the command, right now I don't feel comfortable. Maybe with time, when everybody is speaking to their device it might change."

Interaction Design

Participants described aspects of the application that made interaction feel less fluid and more difficult to control. Because the application required visual attention for successful play, participants often held the device differently for different gesture modes. For example, participants described how they were forced to hold the device at eye level while in head mode because

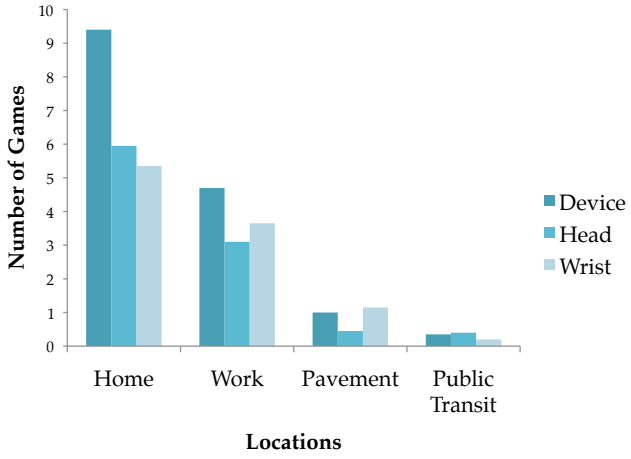

Figure 5.14: Average number of games played by locations for each gesture mode.

they felt that the interaction did not work if their head was not upright.

Separation of Action and Attention – One issue with using the body-based gestures was that the input was performed away from the device, leading to a sense of disconnect or confusion. For example, one participant stated that "the device that holds the interface is also the thing you are interacting with, so it becomes a single thing [in device mode]. In other modes, it's an additional sensor that you are using on your wrist or your head, so your attention is split. So here [in device mode], the concentration is more so it gives you a more engaged feeling but when it is split, it takes away some of the joy." Another participant stated that "One hand is doing the gesture part, and one hand is assigned to holding, and tapping. There it becomes difficult." This disconnect between action and the focus of attention led to a less satisfying experience for some participants.

Symmetry of Interaction – Participants discussed the lack of symmetry when using the wrist mode because each hand was set to a different task and the range of comfortable wrist rotations favoured rotating to the right. One participant stated that "with the wrist mode, I found it was difficult to use both wrists. I had to keep my hand like this and again I need to use one to tap." When gesturing was asymmetric across the body, participants found this more difficult. This issue was mainly described with respect to the wrist gesture mode, because gesturing was performed with one wrist and interaction with the device had to be completed with the other hand.

Relationship Between Action and Device – In this application, interaction included tilting the device and tilting the head. Both of these actions could change the users' perspective on the visual feedback provided by the application. For example, one participant described how device tilting gestures made it difficult to properly view the visual feedback on the screen. This participant stated that "because you have to tilt it (the device), it makes it difficult to target." Another participant stated that "if you use your head, you are using the phone in a natural position, whereas with the wrist you are holding the phone with your wrist at the side, it looks unusual." The orientation of the device with respect to interaction played a significant role in the enjoyability and usability of the application.

5.4.5 Discussion

The purpose of this study was to give participants in different cultural settings the opportunity to interact with a multimodal application in their everyday lives. Although the user study lasted only six days, participants were able to gain a wide variety of experiences, form strong opinions about the performances and the application itself, and found ways of performing interactions that were acceptable in their daily lives. This occurred in the variety of ways that participants took a simple performance and adapted it to make it work personally and socially for them. For example, participants discussed how head-based interaction could be easily hidden in everyday life by performing subtle actions and appearing to use their mobile phone traditionally. Alternatively, participants described how they included friends and family in their game play in order to explain their actions to others and make their performances more acceptable and even enjoyable in front of these spectators. Although the game was designed for individual use, many participants described how they used the game socially to show off novel technology as well as make performance a better experience. This balance between the ability to dynamically hide or perform interactions allowed for more possibilities for interaction in everyday life. When participants were alone, subtle and hidden actions worked best. However, the same interface could be used in a performative manner when with a group of friends.

In contrast, appropriation failed in cases where usage was deemed unacceptable or where issues with the interface made interaction too awkward. In some cases, participants were simply unwilling to experiment with different gesture modes in public places. Head-based interaction was intentionally included in this study as a relatively unacceptable interaction technique based on the previous cross-cultural focus group study, and many participants were simply unwilling to perform these gestures in public. However, participants described ways that even the least acceptable interactions could be incorporated into everyday life. By exploiting some of the benefits of unusual or traditionally unacceptable interactions, users of a multimodal system could be gradually encouraged to adopt these new interactions using the 'foot in the door' technique (Freedman and Fraser, 1996). This technique, named for the successful practice of door-to-door salesmen, describes how compliance can be much more easily achieved through incremental steps rather than large changes. With respect to a multimodal interface, the implementation of a single compelling interaction using a relatively unacceptable technique might be enough to 'get the foot in the door' for future applications of that technique. Chapter 3 demonstrated that even one bad or good experience can be enough to set a user's opinion, and introducing new and unusual interactions gradually, for example by exploiting the qualities of body-based interactions described by participants in this study, could help prevent users from simply rejecting new interactions.

The interaction preferences developed differently for participants in different cultural settings mainly due to initial preferences for multimodal interaction techniques. For example, the possibility of using voice-based control as part of a mobile interface was much more acceptable for participants in Bangalore than in Glasgow. Participants in Bangalore discussed how voice control would be acceptable, would save time, and would be an interesting a novel interaction technique. In contrast, the majority of participants from Glasgow stated they simply would not use voice-based control for fear of embarrassment or acting inappropriately. Between these

two user groups, voice versus gesture-based control proved the most different with respect to participant preferences. However, once different preferences were discussed, participants from both cultural settings described the same types of issues with embarrassment in everyday life, issues with the interaction design of the application, and methods for making interaction work in daily life. This indicates that it is important to choose the right initial set of techniques for different cultural settings, which will then be adopted and used in very similar ways. This could be enacted by completing an initial survey looking at acceptance of possible interaction techniques in different cultural settings and selecting culturally appropriate sets of techniques before implementation, as in this chapter.

5.5 Conclusion

This chapter explored the appropriation of multimodal interfaces by demonstrating a complete multimodal design process from early brainstorming and prototyping to a full deployment in the wild. This design process also demonstrated the successful application of the entire methodology presented by this thesis from beginning to end, evaluating acceptance, adoption, and appropriation in different cultural settings. By completing this process in these different settings, this chapter demonstrates the effectiveness and general applicability of this method. The first study in this design process was a survey study to identify respondents' existing mobile phone practices and preferences. This survey provided a baseline from which the subsequent focus groups were designed. These focus groups explored a variety of input and output techniques in order to identify participants' preferences and differences with respect to the acceptance and adoption of multimodal interaction techniques in different cultural settings. The final study was completed in the wild with a multimodal application designed based on the results of the focus group. This study demonstrates how these multimodal techniques were adopted, appropriated, or rejected in real world situations.

This chapter successfully demonstrates the methodology proposed by this thesis for evaluating social acceptability and user experience. The application, designed using a combination of qualitative and quantitative data collected through a survey and a focus group, resulted in an acceptable and usable interface. Participants' preferences for different gesture modes in the Mo!Games application matched those found in the focus group study. Additionally, participants' preferences and perceptions about voice control were also predicted by the focus group study, where participants from Bangalore found voice-based control significantly more acceptable than participants in Glasgow. This demonstrates the validity and usefulness of early evaluations of social acceptability, where users can make informed judgements about the acceptability of using different interactions in their everyday lives using simple prototypes with survey responses and group discussions. These kinds of evaluations are useful for guiding design because they can be completed early on in the design process with a divergent set of possible interaction techniques and easily can be incorporated into design.

The results of the first survey study served mainly to provide a baseline for comparison since this entire evaluation was completed in two highly different cultural settings. The results of this study identified differences between favoured mobile phone functionalities and different

etiquette influencing acceptable use in different cultural settings. For example, the significant differences between the usage of phone calls in public places between Glasgow and Bangalore identified an interesting line of inquiry that was followed up in the subsequent studies. In the focus group study, the divergent set of possible interaction techniques included a variety of input and output techniques that required making noise in public places similar to the kind of noise a phone call makes. Because this study began without restrictions on what kind of application should be developed, this survey served as a simple method to gain some inspiration and a better understanding of how different locations might affect multimodal interaction.

The focus group study looked at the acceptance and adoption of a divergent set of multimodal interaction techniques by combining survey-style responses about acceptance with focus group discussions about adoption. These focus groups demonstrate the different factors influencing technology usage in public as seen in different cultural settings. Overall, participants from Bangalore had significantly higher rejection rates than participants in Glasgow. However, some interaction categories had significantly higher acceptance rates in Bangalore than in Glasgow. Participants in Bangalore had significantly higher acceptance rates for voice-based interaction, a finding that was incorporated into the design of the Mo!Games application. Although participants had similar motives for accepting multimodal interactions, the ways in which these were expressed resulted in very different preferences. The desire to use technology in a comfortable and satisfying way meant different things in these different settings.

The results of this 'in the wild' user study demonstrate how participants might actually use and appropriate multimodal interactions in their everyday lives. Over the span of six days, with two days per gesture mode, participants were able to develop strong preferences for different interaction techniques, opinions about the accuracy and usability of the application, and discuss how possible additions to the application might work in their daily lives. Participants adapted performance in order to make interaction acceptable in different contexts. In some cases, performance was hidden in public places or shared with familiar spectators in order to successfully incorporate interaction into public places. Culture played a significant role in the initial preferences for different interaction techniques but had a much smaller effect on how participants actually went about adopting and appropriating the techniques in their daily lives. Gesture modes played a major role in how the interface was used, with many participants simply refusing to use certain modes in public places. The selection of acceptable interaction techniques, or the exploitation of the benefits of unusual techniques, is an important factor to consider when designing interfaces for real world usage.

Chapter 6

Conclusions

This thesis explores the user experience of multimodal interaction techniques by completing evaluations at increasing levels of depth and complexity. Using a novel approach to user experience, this thesis explores multimodal interaction as *performative* interaction, organising the results based on individuals' subjective experiences, their perceptions about spectators, and their memories about past experiences.

Reviewing the thesis statement and contributions presented in Chapter 1, three main themes emerge:

- **Evaluations of User Experience –** The results of surveys and user studies that identify how users make decisions about how they will use multimodal interaction techniques.
- **Factors of Social Acceptability –** The independent factors that influence individuals' decisions about the social acceptability of multimodal interaction.
- **Methodology –** The application and verification of the approach to performative user experience proposed by this thesis.

6.1 Evaluating User Experience, Performative Interaction, and Social Acceptability

The gesture survey study using video prototypes, presented in Section 3.2, demonstrated that device-based gestures were significantly more acceptable than body-based ones. Because these interactions display the device and demonstrate the purpose of the interaction as part of a mobile interface, these actions are less likely to be misunderstood by spectators and give performers greater confidence in their interactions. The ability of interactions to demonstrate the purpose of the action as part an interaction plays a significant role in the acceptability of that interaction.

The on-the-street user study, described in Section 3.3, demonstrated some of the reasons why

users accept or reject multimodal interaction techniques. One of the most commonly discussed reasons for enjoying gesture-based interaction was that these actions appeared or felt similar to everyday actions. For example, participants described how foot tapping was highly acceptable because this is an action which already occurs in everyday life. Participants also described how actions with a mobile phone could be metaphorically similar to everyday actions. For example, one participant described how shaking the phone was acceptable because it was similar to shaking a bottle of juice. The ability for interactions to mimic everyday life should be considered when designing new interactions for mobile phones to increase possible acceptance.

One of the most commonly discussed reasons for disliking gesture-based interaction was that some of the actions were highly noticeable or attention seeking. Gestures which were likely to attract unwanted attention or to be misunderstood by spectators were generally unacceptable. For example, participants described how actions that did not have a clear purpose, such as shoulder tapping, were likely to attract the attention of spectators. This rationale for disliking gestures brings up the point that performance is not always desirable, where users also want to hide their interactions.

The *in situ* user study of a remote awareness application, described in Section 4.2, demonstrated how users fluidly switched between highly performative interactions to more subtle or hidden interactions based on location. Because the MuMo application supported extravagant and subtle interactions equally, participants employed both of these styles of performance during the course of the study unprompted. This highlights the importance of incorporating *flexibility* into recognition classes so that users can perform the same gestures but adjust the appearance and size of the performance based on their current context. Designers should specifically seek out interactions that afford tailoring and customisation to support this kind of flexibility. In some situations, users might enjoy or desire more performative experiences and this should be supported just as much as smaller more subtle interactions.

The focus group study that explored gesture and voice-based interaction, described in Section 4.3, looked at a wider set of interaction techniques than the previous studies. These focus groups evaluated the acceptance of gesture and voice-based interactions organised into seven categories. The results of this study demonstrated that arbitrary gestures (Nespoulous *et al.*, 1986), or those gestures that do not have an immediately clear meaning outside of the context of speech, are significantly less acceptable than device, body, and emblematic gesture (Kendon, 1997) interaction. Because arbitrary gestures lack a clear meaning, they are more likely to be misunderstood by spectators and users found these interactions relatively unacceptable in public places. Similarly, non-speech interactions, which might not have a clear meaning for spectators, were significantly less acceptable than command-based speech interactions. Multimodal input that was likely to attract unwanted attention or unwanted reactions from spectators were relatively unacceptable as compared to interactions with more legible intentions or meanings. These focus groups also demonstrated three approaches that users employed when placing interactions into a usability context: relationship to the device, personal meaning, and relationship to functionality. Participants used these approaches to make sense of how the interactions would work and imagine how they could incorporate these interaction into their everyday lives. Although participants did not always consider each of these concerns, de-

signs of multimodal interfaces should specifically incorporate ideas relating to each approach in order to appeal to a wider audience of users and gain greater acceptance.

The longitudinal study of the Mo!Games Application, described in Section 5.4, demonstrated how users might actually incorporate multimodal interactions into their everyday lives. This study demonstrated how user perceptions about the accuracy of different gesture modes greatly influenced their experiences, even though accuracy rates were not significantly different between modes. These perceptions of low accuracy are often due to issues such as larger perceived size of movement, isomorphism errors (Williamson, 2006), and lack of familiarity. This study also demonstrated some of the reasons that participants described for enjoying relatively unacceptable interactions. These include a greater sense of control, the ability to hide interactions, or simply the novel experience of interactions. These results give designers methods of encouraging users to adopt new interactions that might otherwise be deemed unacceptable.

6.2 Factors of Social Acceptability

The gesture survey using video prototypes, described in Section 3.2, demonstrated that location and audience plays a significant role in gesture interaction acceptance. The most acceptable location was home and the least acceptable location was driving, closely followed by the pavement. With respect to audience, the most acceptable was the lack of any audience (alone) and the least acceptable were strangers. Designers must take this into account when choosing interactions for an interface, particularly for interfaces that will be used in public locations or in front of strangers. These situations had significantly lower acceptance rates than others such as home, workplace, friends or family. The gesture and voice survey, presented in Section 3.4, further explored the issue of audience by looking at how different audiences in the same locations affected acceptance. These results showed that the presence of familiar others significantly increased the acceptability of performing gesture or voice-based interactions, with large effect sizes in all the locations examined in this survey. This indicates that the development of social or collaborative interfaces, where users must interact together, could exploit the presence of familiar audiences to encourage more performative and expressive interactions.

The focus group study using a variety of experience prototypes, presented in Section 4.3, demonstrated that users from different age groups had significantly different preferences based on their initial reactions to multimodal interaction techniques. Notably, participants from the older user groups (70-95) found emblematic gestures significantly more acceptable and device-based gestures significantly less acceptable than participants from the younger user groups (18-30). These differing preferences are explained by the differing approaches used by participants when placing these interactions into context. Participants from the older user groups often focused on personal meanings before placing interactions into a usability context. Because emblematic gestures have easily identifiable meanings, these were more acceptable for participants in the older user groups and more easily made sense of as part of a mobile interface. In contrast, participants from the younger user groups often focused on the device itself or how the interaction might be used in a scenario before considering meaning. This explains the preference for device-based interactions because these younger user groups often

considered interaction to be focused around the device itself.

The effect of culture on social acceptability is also explored in this thesis in a series of studies concurrently in Glasgow and Bangalore. The cross-cultural focus group study, presented in Section 5.3, demonstrates how the differing initial preferences for multimodal interaction techniques are driven by the same underlying motivations. For example, the motivation to interact without attracting unwanted attention was present in both cultures, but expressed differently and led to different preferences. In Glasgow, this motivation was expressed as the desire to perform subtle or hidden actions such as gazing with the eyes or belt tapping. In Bangalore, this motivation was expressed as the desire not to disturb others with preferences for voice-based interactions. Participants in Bangalore had significantly lower rejection rates for voice-based interactions because these actions were not considered disruptive to others in that cultural context. Participants in Glasgow had significantly higher rejection rates for emblematic gestures because these would be difficult to perform surreptitiously due to their highly iconic, recognisable nature. However, even though participants had differing initial preferences, they showed more similarities than differences when placing interactions into context. Participants from both cultural settings considered how interactions should work relative to the device, how commonly used functionalities could be made more accessible, and how interactions could be continuous or discrete in different applications. The longitudinal deployment of the Mo!Games application, described in Section 5.4, confirms this finding by demonstrating that although users had differing initial preferences, their perceptions about how the interactions worked in everyday life and their methods of incorporating those interactions were very similar. These results highlight the value of completing early evaluations of social acceptability to determine initial preferences for users in different social or cultural groups to increase the acceptance of those interaction techniques.

6.3 Approach to User Experience and Social Acceptability

This thesis proposes a performative perspective on user experience with theoretical roots in phenomenology, a mixed-methods approach to user studies, and analysis techniques based on dramaturgical metaphors. The effectiveness of this methodology is demonstrated through a variety of studies the build on each other with increasing depth and complexity. The survey study that used video prototypes for the evaluation of social acceptability, described in Section 3.2 was verified by the on-the-street study, described in Section 3.3. Even though respondents in the gesture survey did not have the opportunity to perform the gestures in real world settings, the rankings from that survey were comparable to those found in the on-the-street study. The cross-cultural focus group study presented in Section 5.3 was verified by the longitudinal study of the Mo!Games application described in Section 5.4. Again, even though participants in the focus groups were asked to imagine the experience of performing multimodal interaction techniques in public places, the rankings and acceptances from the focus group study were comparable to those found in the longitudinal study. The completion of low-cost evaluations is a valuable preparation for more sophisticated user studies and deployments.

This thesis demonstrates how evaluations of varying costs can be completed at different stages

of development depending on the type of data needed and the control over the experience desired. The use of video prototypes is demonstrated throughout this thesis as a technique that allows for low cost evaluations of large sets of interaction techniques. Even though users simply watch the interaction techniques being performed and are asked to imagine how they would feel performing these in different settings, they are able to provide accurate responses about the social acceptability of these interaction and participate in detailed discussions about how these might work in real world settings. Video prototypes were used in both quantitative studies (Sections 3.2, 3.4) and qualitative studies (Sections 3.3), where the combination of both approaches (Sections 4.3, 5.3) makes the most of these prototypes.

This thesis also demonstrates how relatively higher cost user studies can be completed at different stages of development. The on-the-street user study, presented in Section 3.3, looked at acceptance over three usage sessions with a prototype system developed to control the experience by prompting users and timing gesture repetitions. This user study presented an in depth qualitative analysis of the early stages of interaction acceptance. The MuMo application, presented in Section 4.2, allowed users to perform interactions with a fully interactive application at the early stages of adoption. Although this was a relatively controlled experience, users had the opportunity to interact with the application in real world settings. The cross-cultural user study of the Mo!Games application, presented in Section 5.4, was completed entirely in the wild, where experimenter control over the experience was relatively limited. This represents one of the highest cost studies presented in this thesis with respect to development and deployment time required. However, this study revealed a variety of results that could not have been explored in the previous studies. For example, participants were able to develop their preferences over time and accuracy, and experience could be evaluated on the move without an experimenter present. Additionally, participants were able to develop perceptions that incorporated both technical and social influences. Each of these evaluations demonstrates how in depth user studies can be completed at different stages of use and with prototype or deployment systems.

This thesis also demonstrates how the results of user experience studies can be reincorporated into design at different stages of development and use. For example, the gestures used in the on-the-street study, presented in Section 3.3, were selected based on the results of the previous gesture survey presented in Section 3.2. This ensured that the on-the-street study looked at gestures with a wide distribution of acceptability in order to create very different experiences for users. Even though these studies examined the early stages of acceptance, results could be reincorporated into the design of follow-up studies to better target specific questions about social acceptability. The series of studies demonstrated in Chapter 5 show how the results of prototype evaluations of social acceptability can lead to the development of a multimodal interface for use in everyday life. The results of the cross-cultural focus group study, presented in Section 5.3, are used to guide the design of the deployment application. This allowed for the informed selection of interaction techniques for deployment and the intentional creation of both acceptable and unacceptable interactions. Future evaluations of multimodal interactions could use this approach to select only the most socially acceptable interaction techniques for development and deployment to save valuable design and implementation resources and lead to more acceptable interfaces overall.

6.4 Summary

This thesis lays out a well-grounded framework for completing evaluations of user experience across a variety of contexts. These are verified by comparing results across evaluations. Low cost quantitative survey studies using video prototypes are effective for gathering initial results about the social acceptability of multimodal interactions. Focus group studies, where participants imagine how interactions should work in real world settings, can effectively predict how users will incorporate multimodal interaction techniques into their daily lives in a longitudinal setting. Because surveys and focus groups have a comparatively lower cost to complete than traditional user studies with working systems, these methods are a valuable tool that make feasible the evaluation of large sets of interaction techniques at different stages of the design process.

The techniques presented can be extended and refined for future evaluations of user experience for performative interactions. Although the focus has been on mobile interactions, these same methods would be effective for evaluating how new interaction techniques would be accepted in larger interactive installations. This methodology would also be effective for evaluating different aspects of installations that could affect usage. For example, the experience of performing in front of camera-based systems could lead to different preferences and behaviours than performance using inertial systems. Although these issues were not explored in this thesis, this methodology could also be used to uncover user perceptions about how the failures of different technologies could lead to embarrassment, or how different technologies create a different sense of control. Another way this approach could be extended is by incorporating personality (John *et al.*, 2008) as an analytical tool to give more depth to the results.

The evaluations presented in this thesis make a significant contribution to the understanding of how users make decisions about performative interactions in real world settings. This thesis also explores how some of the factors of social acceptability such as location, audience, age, and culture affect how users make decisions about the social acceptability of multimodal interactions. As technology changes and advances to create new ways of interacting, the evaluation of user experience from a performative perspective will help to incorporate these new interactions successfully into everyday life.

Appendix A

Surveys

A.1 Gesture Survey

The following gestures were included (in random order in the Gesture Survey.

- Device Whip – Whip-like movement of the device in the hand.
- Device Shake – Shaking the device in the hand, side to side.
- Pocket Tap – Tapping the device (in the pocket) with the hand.
- Device Squeezing – Squeezing the device in the hand.
- Table Tap – Tapping the device (on a table) with the hand.
- Shoulder Tap – Tapping the shoulder with the device.
- Device Rhythm – Shaking the device in the hand up and down with an even rhythm.
- Shoulder Rotation – Rotation of the shoulder forwards and backwards.
- Hand Proximity – Measuring a space between the hands.
- Nose Tap – Tapping the nose with the hand.
- Head Nodding – Nodding the head side to side.
- Shifting Balance – Shifting balance side to side while sitting.
- Wrist Rotation – Rotation of the wrist side to side.
- Clapping – Clapping the hands.
- Foot Tapping – Tapping one foot while sitting or standing.
- Belt Tapping – Tapping the belt with the hand.
- Finger Proximity – Measuring a space between the fingers.

- Arm Squeezing – Squeezing of the forearm with the opposite hand.

For each gesture respondents were asked to answer the following questions, where respondents could select as many of the possible options as desired.

In which locations would you use this gesture?

☐ Home

☐ Pavement or Sidewalk

☐ While Driving

☐ As a Passenger on a Bus or Train

☐ Pub or Restaurant

☐ Workplace

Who would you perform this gesture in front of?

☐ Alone

☐ Partner

☐ Friends

☐ Strangers

☐ Family

A.2 Gesture and Voice Survey

The following gestures were included (in random order) in the Gesture and Voice Survey.

- Device Flicking – Whip-like movement of hand while holding a mobile device.
- Buzzing – Making a buzzing sound.
- Head Nodding – Nodding of the head forward.
- Clicking – Making a clicking sound using the inside of the cheek.
- Hook Finger – Holding the index finger in a hook shape.
- Say "Close" – Saying the word "Close" once.
- Open Palm – Open palm held forward at shoulder level.
- Doo Doo Doo (Ascending Pitch) – Three "Doo" sounds of any ascending pitches.
- Peace Sign – Index and middle finger held up together and in 'V' shape.

- Say “I’m Fine” – Saying the words “I’m Fine” once.
- Shrugging – Lifting of the shoulders.
- Say “Lock” – Saying the word “Lock” once.
- Device Stroking – Stroking the back of a mobile device.
- Popping – Popping sound made with the lips.
- Wrist Rotation – Rotation of the wrist side to side.
- Say “That’s Nice” – Saying the words “That’s Nice” once.

For each gesture participants were asked to answer the following questions, where each question included a seven point Likert scale response (Totally Unacceptable, Unacceptable, Slightly Unacceptable, Neutral, Slightly Acceptable, Acceptable, Totally Acceptable).

How acceptable do you feel it would be to perform this command:

- while on the sidewalk or pavement with strangers;
- while on the sidewalk or pavement with friends;
- while at home with family;
- while at home alone;
- as a passenger on a bus or train with strangers;
- and as a passenger on a bus or train with friends.

Personal

Age: _____ **Gender:** ☐ **Male** ☐ **Female**

Occupation: _____

Living Status *(tick all that apply)*:

☐ Alone ☐ With Partner ☐ With Parents
☐ With Children ☐ With Flatmates ☐ Other: _____

In which city/town/village do you currently live?

_____

What languages do you speak fluently?

_____

Mobile Phone

What kind of mobile phone do you have?
Make: _____ Model: _____

How long have you owned this mobile?
Years: _____ Months: _____

Which mobile devices do you usually carry with you?

☐ Mobile Phone ☐ Additional Mobile Phone
☐ PDA ☐ MP3 Player
☐ Radio ☐ Camera / Video Camera (Not on Phone)
☐ GPS Device ☐ Other: _____

What kind of mobile phone service do you have?

☐ Pre-paid (No Data) ☐ Post-paid (No Data)
☐ Pre-paid (With Data) ☐ Post-paid (With Data)

What allowances does your service include?

_____

How many of the following items do you do per day?
Text Messages: _____ Picture Messages: _____ Phone Calls: _____
How many of your text/picture messages are forwards? **%**_____

Phone Usage

Which capabilities does your mobile have? *(tick all that apply)*

☐ Phone Calls ☐ Address Book ☐ Text Messages
☐ Picture Messages ☐ Email ☐ Games
☐ Alarm Clock ☐ Internet ☐ Music
☐ Calendar ☐ GPS ☐ Other: _____

How often do you use these capabilities? *(tick one per row)*

	Daily	Weekly	Rarely	Never
Phone Calls	☐	☐	☐	☐
Address Book	☐	☐	☐	☐
Text Messages	☐	☐	☐	☐
Picture Messages	☐	☐	☐	☐
Email	☐	☐	☐	☐
Games	☐	☐	☐	☐
Alarm Clock	☐	☐	☐	☐
Internet	☐	☐	☐	☐
Music	☐	☐	☐	☐
Calendar	☐	☐	☐	☐
GPS	☐	☐	☐	☐

When do you use these capabilities? *(tick all that apply)*:

	While Walking	During Meals	While on a Bus	While Shopping	While at Home
Phone Calls	☐	☐	☐	☐	☐
Address Book	☐	☐	☐	☐	☐
Text Messages	☐	☐	☐	☐	☐
Picture Messages	☐	☐	☐	☐	☐
Email	☐	☐	☐	☐	☐
Games	☐	☐	☐	☐	☐
Alarm Clock	☐	☐	☐	☐	☐
Internet	☐	☐	☐	☐	☐
Music	☐	☐	☐	☐	☐
Calendar	☐	☐	☐	☐	☐
GPS	☐	☐	☐	☐	☐

Appendix B

Video Prototypes

All of the video prototypes can be found at:

http://www.juliericowilliamson.com/thesis

Appendix C

Interview Questions

C.1 On-the-Street User Study

C.1.1 First Interview

1. In your opinion, how would you define gestures?

2. How do the gestures used today fit with your definition?

3. How did you feel in the lab setting versus the pavement setting?

4. What if you had been part of a group of people?

5. Which gestures did you feel most comfortable performing?

6. Why? How?

7. Which gestures did you feel least comfortable performing?

8. Why? How?

9. Which gestures would you most likely use with a mobile interface?

10. Any others not evaluated today?

11. How did you feel using the SHAKE sensor pack?

C.1.2 Second Interview

1. Which gestures did you feel most comfortable performing?

2. Which gestures did you feel the least comfortable performing?

3. How did you feel about performing these gestures for the second time?

4. What was different?

5. How satisfied did you feel about your performance as compared to the first trial? Any differences?

6. Why?

7. Do you think that using these gestures daily would make a difference in your ideas about them?

8. After we talked about how these gestures might be applied to an application last week, were your ideas about these gestures affected?

C.1.3 Third Interview

1. Which gestures were you the most comfortable using?

2. Which gestures were you the least comfortable using?

3. How did you feel performing these gestures for the third time?

4. Over the course of this study, how did you learn to perform these gestures?

5. What did you use mostly? The videos? Your memory?

6. Given that you were allowed to perform the gestures with your own interpretation, how did you choose the way you would perform gestures?

7. How do you think your performances varied from session to session?

8. Why?

9. If these gestures were implemented for a mobile interface, how would your experience with these gestures influence your usage of them in the future?

C.2 MuMo User Study

C.2.1 First Interview

1. Why did you choose the objects you chose?

2. Did they help or inhibit control over the visualisation?

3. How easy was it to control the visualisation?

4. What kinds of actions did you use to send bubbles?

5. What kinds of actions did you use to send movements?

6. What kinds of actions did you use to send happy thoughts?

7. What kinds of actions did you use to send Busy thoughts?

8. Where there any situations where you felt uncomfortable doing these actions?

9. How do you think this would fit into your daily life?

10. What kind of interpretations did you think of while watching the fish tank?

11. How did you imagine the other person controlling the visualisation?

12. So, what were your overall experiences?

13. What were you favourite aspects of this experiment?

14. What were your least favourite aspects of this experiment?

15. Did you do anything differently from the first session to the second session?

16. When you were performing different actions, whom were you mostly thinking about?

C.2.2 Second Interview

1. What did you imagine the other person did to send bubbles?

2. What kinds of actions did you use to send bubbles?

3. What do you imagine your partner performing when sending bubbles?

4. What kinds of actions did you use to send swimming movements?

5. What do you imagine your partner performing when sending swimming movements?

6. What kinds of actions did you use to send happy thoughts?

7. What do you imagine the other

8. What kinds of actions did you use to send Busy thoughts?

9. What were you favourite aspects of this experiment?

10. What were your least favourite aspects of this experiment?

11. Did you do anything differently from the first session to the second session?

12. How would you compare you experiences this time and last time?

13. When you were performing different actions, who were you mostly thinking about?

14. How did you go about completing the packet? How did you get on with it?

C.3 Mo!Games User Study

1. What was your favourite mode? Why?

2. What was your least favourite mode? Why?

3. Which do you think was the most usable? Why?

4. Which do you think was the least usable? Why?

5. Which locations did you use this system in?

6. How did the different modes work in these locations? What was your experience?

7. Which location was the most comfortable?

8. Which location was the least comfortable?

9. How many voice notes did you leave? How comfortable were you using the voice capabilities?

10. How would you have used this system if it included voice/touch control instead of touch/voice control?

11. How did you feel about using the external sensor pack?

Appendix D

Focus Group Schedules

D.1 Adoption Focus Groups

Each session will comprise of three basic sections:

Prototype Trial (30 Mins)

Try out prototypes. Depending on time, this might be just a subset of the prototypes. Any Refreshments will be served up at this time. I think having at least some drinks will be beneficial.

Social Acceptance Discussion (30 Mins)

Questions revolving around social acceptance, including asking participants to imagine different situations, different audiences, and identifying why they like inputs and what issues they would imagine.

Take a few minutes to identify at least a top and bottom 5 for both the gestures and voice inputs. Which inputs did you like the most? Why? Which of these did you like the least? Why? Of these, which ones do you imagine using in your daily life? Brainstorm some locations where phones are used. How do you use your mobile phones in these places? What makes these locations different to you? Thinking of your top 5 and bottom 5, how would you feel about using these inputs in these locations?

Adoption/Task Discussion (30 Mins)

Brainstorming of tasks, and possibly the presentation of a set of tasks and discussing which inputs they would use, and why. Perhaps provide a sheet of some kind. Brainstorm a list of tasks you commonly use on a mobile phone. Which input techniques do you think fit best with these tasks? Why?

How do you imagine these working? How would they be incorporated in your daily life? Phone usage?

For a selected list of tasks (or a predetermined one) discuss usage scenarios (incorporating the different usage contexts) for 2-3 input techniques. Have paper and markers available if people want to draw things out.

D.2 Appropriation Focus Groups

Explain worksheets and take 5-10 minutes to complete.

First, rank the interactions from 1 to 50, where 1 is the best and 50 is the worst. Focus on your top 10 and bottom 10, and fill in the rest time permitting.

The bottom half of this worksheet lists all of the techniques. Cross out any of the techniques that you would simply not find acceptable in a public setting such as a busy sidewalk.

Now we'll do some quick activities to get you a little more familiar with these interaction techniques. As a group, I would like you to organise these cards into groups. You can organise them in any way you like, in as many groups as you like. Once one categorisation has been completed, or if real difficulties occur:

- Acceptable/Unacceptable

- What Functionality You Would Use Them For

- What They Mean to You

Interview

Which of these interaction techniques did you find the most acceptable? Why?

Which of these interaction techniques did you find the least acceptable? Why?

Thinking of your top 5 and bottom 5, how would you feel about using these why walking on the street?

How about while in a meeting or class?

How about while at home?

How about while walking down the sidewalk?

How about at meals?

Scenario Activity

Discuss the following usage scenarios, where each participant will assign 1 or more interaction techniques to the task, and act out how it should work with some verbal descriptions.

One example of this should be performed by the group leader to get things moving. Think about the scenarios as a task, an action to complete that task, and a result or some feedback

that it has been completed. For example, if I'm busy and I want to send a set reply to a text message, I might snap my fingers or swipe in the air. Then the phone could vibrate or make a chime to let me know this has been completed. If I was in a meeting, I might choose to squeeze the phone instead. If I was at home cooking, I might just say a command to the phone if my hands were busy like "I'm busy, I'll call you back later."

For each scenario, go through the following structure:

Select 2 or 3 interaction techniques that you think might work for this task.

Why have you selected the interaction techniques?

How would you choose to perform the interaction technique or change it to better suit the purpose? Why?

How do you expect it to work? If you do the action, what will the phone do? How would the scenario change based on different usage scenarios and locations? Home? Work? Outdoor public place? Indoor public place? In Transit? Why?

Checking Time

Flashlight/Torch

Alarm Setting

Check Messages/Check for Forwards

Looking Through Photos or Text Messages

Silencing Phone

Music Player

Appendix E

Focus Group Worksheets

E.1 Adoption Focus Groups

Participant ID:

This sheet, which will be collected after this session, can be used to keep notes and record your thoughts. You may also write any thoughts you wish to be kept private between you and the experimenter here.

General Questions

How old are you? _____

Where did you grow up? _____

What do you study/what field do you work in? _____

What mobile phone (if any) do you own? _____

Which of the following functionalities do you used on your mobile phone?

___ Phone Calls ___ Email ___ Games ___ Alarm Clock

___ Texting ___ Internet ___ Calendar/Planner ___ Music

Ranking – Please do not complete this section until asked to

Take a few minutes to think about the inputs that you liked and disliked. For **both** lists below, please number your favourites (1 to 5) and your least favourites (13 to 16). Think about **why** you like or dislike the input techniques. Then, cross out all of the inputs that you would not use in an everyday public setting, such as a busy pavement.

Gestures

_____ Device Rotation (Joystick)
_____ Foot Tapping
_____ Shrugging
_____ Sideways Fist
_____ Head Nodding
_____ Upright Fist
_____ Money Gesture
_____ Device Stroke
_____ Wrist Rotation
_____ Device Flick
_____ Shoulder Rotation
_____ Device Shaking
_____ Hook Finger
_____ Peace Sign
_____ OK Gesture
_____ Open Palm

Voice

_____ Say "Lock"
_____ Say "I'm Fine"
_____ Whistling
_____ Say "Psst"
_____ Say "Open"
_____ Say "Bad Weather"
_____ Buzzing
_____ Humming
_____ Clicking
_____ Popping
_____ Say "Chh"
_____ Say "So Busy"
_____ Say "That's Nice"
_____ Doo Doo Doo (Ascending Pitch)
_____ Say "Call"
_____ Say "Close"

E.2 Appropriation Focus Groups

Multimodal Interaction Focus Group

Participant ID:

How acceptable do you find the following interaction techniques to use in public places? Please rate each of the interaction techniques as you watch the videos.

Interaction Technique	Totally Unacceptable	Unacceptable	Neutral	Acceptable	Totally Acceptable
Phone Shaking	○	○	○	○	○
Wrist Rotation	○	○	○	○	○
Say "Set Alarm for 8 AM"	○	○	○	○	○
Listen to Bell Chime	○	○	○	○	○
Thumbs Up	○	○	○	○	○
Device Squeezing	○	○	○	○	○
Head Nodding Side to Side	○	○	○	○	○
Say "Where is my phone?"	○	○	○	○	○
Gesture "Come here"	○	○	○	○	○
Phone Twisting	○	○	○	○	○
Pointing at Someone	○	○	○	○	○
One Foot Tapping	○	○	○	○	○
Say "Do I have messages?"	○	○	○	○	○
Gesture "Blah Blah Blah"	○	○	○	○	○
Turning Phone Over	○	○	○	○	○
Listen "You have a missed call from John"	○	○	○	○	○
Snapping	○	○	○	○	○
Say "Next"	○	○	○	○	○
Phone Flicking	○	○	○	○	○
Writing on your own hand	○	○	○	○	○
Phone Tilting	○	○	○	○	○
Head Nodding Up and Down	○	○	○	○	○
Listen "Martin sent you a photo"	○	○	○	○	○
Gesture "Crazy"	○	○	○	○	○
Phone Tapping in Hand	○	○	○	○	○
Clapping	○	○	○	○	○
Say "Delete"	○	○	○	○	○
Phone Stroke with Two Hands	○	○	○	○	○
Swipe	○	○	○	○	○
Gesture "Telephone"	○	○	○	○	○
Hold Phone to Ear	○	○	○	○	○
Listen "Mary sent you a text"	○	○	○	○	○
Finger Positions on Leg	○	○	○	○	○
Shrugging	○	○	○	○	○
Pointing to the Side	○	○	○	○	○
Spinning Phone	○	○	○	○	○
Listen to Musical Chime	○	○	○	○	○
Hold Phone on Shoulder	○	○	○	○	○
Pointing Phone at Someone	○	○	○	○	○
Gaze with Eyes	○	○	○	○	○
Say "Call Marina"	○	○	○	○	○
Pointing Phone to the Side	○	○	○	○	○
Listen to Alarm Chime	○	○	○	○	○
Scanning with Phone	○	○	○	○	○
Hold Phone on Hip	○	○	○	○	○
Gaze with Head	○	○	○	○	○
Belt Tapping	○	○	○	○	○
Stroke with One Hand	○	○	○	○	○
Say "Reply"	○	○	○	○	○
Two Feet Tapping	○	○	○	○	○

Multimodal Interaction Focus Group

Participant ID:

Please rank these interaction techniques from 1 to 50, where 1 is the best and 50 is the worst. Rank each technique based on how acceptable it would be to use in a public place.

_____ Phone Shaking
_____ Wrist Rotation
_____ Say "Set Alarm for 8 AM"
_____ Listen to Bell Chime
_____ Thumbs Up
_____ Device Squeezing
_____ Head Nodding Side to Side
_____ Say "Where is my phone?"
_____ Gesture "Come here"
_____ Phone Twisting
_____ Pointing at Someone
_____ One Foot Tapping
_____ Say "Do I have messages?"
_____ Gesture "Blah Blah Blah"
_____ Turning Phone Over
_____ Listen "You have a missed call from John"
_____ Snapping
_____ Say "Next"
_____ Phone Flicking
_____ Writing on your own hand
_____ Phone Tilting
_____ Head Nodding Up and Down
_____ Listen "Martin sent you a photo"
_____ Gesture "Crazy"
_____ Phone Tapping in Hand

_____ Clapping
_____ Say "Delete"
_____ Phone Stroke with Two Hands
_____ Swipe
_____ Gesture "Telephone"
_____ Hold Phone to Ear
_____ Listen "Mary sent you a text"
_____ Finger Positions on Leg
_____ Shrugging
_____ Pointing to the Side
_____ Spinning Phone
_____ Listen to Musical Chime
_____ Hold Phone on Shoulder
_____ Pointing Phone at Someone
_____ Gaze with Eyes
_____ Say "Call Marina"
_____ Pointing Phone to the Side
_____ Listen to Alarm Chime
_____ Scanning with Phone
_____ Hold Phone on Hip
_____ Gaze with Head
_____ Belt Tapping
_____ Stroke with One Hand
_____ Say "Reply"
_____ Two Feet Tapping

For each of the interaction techniques, think about how you would feel performing them in a public space. Cross through every interaction technique that you would simply not accept for use in your everyday life.

Phone Shaking
Wrist Rotation
Say "Set Alarm for 8 AM"
Listen to Bell Chime
Thumbs Up
Device Squeezing
Head Nodding Side to Side
Say "Where is my phone?"
Gesture "Come here"
Phone Twisting
Pointing at Someone
One Foot Tapping
Say "Do I have messages?"
Gesture "Blah Blah Blah"
Turning Phone Over
Listen "You have a missed call from John"
Snapping
Say "Next"
Phone Flicking
Writing on your own hand
Phone Tilting
Head Nodding Up and Down
Listen "Martin sent you a photo"
Gesture "Crazy"
Phone Tapping in Hand

Clapping
Say "Delete"
Phone Stroke with Two Hands
Swipe
Gesture "Telephone"
Hold Phone to Ear
Listen "Mary sent you a text"
Finger Positions on Leg
Shrugging
Pointing to the Side
Spinning Phone
Listen to Musical Chime
Hold Phone on Shoulder
Pointing Phone at Someone
Gaze with Eyes
Say "Call Marina"
Pointing Phone to the Side
Listen to Alarm Chime
Scanning with Phone
Hold Phone on Hip
Gaze with Head
Belt Tapping
Stroke with One Hand
Say "Reply"
Two Feet Tapping

Multimodal Interaction Focus Group

Participant ID:

As a group, we will go through the following scenarios of use. Use this page to keep your own notes.

Checking Time

Flashlight/Torch

Alarm Setting

Check Messages/Check for Forwards

Looking Through Photos or Text Messages

Silencing Phone

Music Player

Bibliography

ISO DIS 9241-210:2010. Ergonomics of human system interactions - part 210: Human centered design for interactive systems. International Standards Organization, Switzerland 2010.

Thoman Attig. *Phenomenology in Practice and Theory*, chapter Existential Phenomenology and Applied Philosophy, pages 161–176. Martinus Nijhoff Publishers, 1985.

Steve Benford. Performing musical interaction: Lessons from the study of extended theatrical performances. *Comput. Music J.*, 34:49–61, December 2010.

Steve Benford, Andy Crabtree, Stuart Reeves, Jennifer Sheridan, Alan Dix, Martin Flintham, and Adam Drozd. The frame of the game: Blurring the boundary between fiction and reality in mobile experiences. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, CHI '06, pages 427–436, New York, NY, USA, 2006. ACM.

D. Benson and J.A. Hughes. *The perspective of ethnomethodology*. Books on demand. Longman, 1983.

Richard A. Bolt. 'Put-that-there': Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '80, pages 262–270, New York, NY, USA, 1980. ACM.

Johanna Brewer, Amanda Williams, and Paul Dourish. A handle on what's going on: combining tangible interfaces and ambient displays for collaborative groups. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 3–10, New York, NY, USA, 2007. ACM.

Lorna M. Brown and John Williamson. Shake2talk: multimodal messaging for interpersonal communication. In *Proceedings of the 2nd international conference on Haptic and audio interaction design*, HAID'07, pages 44–55, Berlin, Heidelberg, 2007. Springer-Verlag.

Marion Buchenau and Jane Fulton Suri. Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, DIS '00, pages 424–433, New York, NY, USA, 2000. ACM.

Scott W. Campbell. Perceptions of mobile phone use in public settings: A cross-cultural comparison. *International Journal of Communication*, 1(1):738–757, 2007.

Justine Cassell. *Computer Vision in Human-Machine Interaction*, chapter A Framework for Gesture Generation and Interpretation, pages 191–215. Cambridge University Press, 1998.

W.J. Conover. *Practical nonparametric statistics*. Wiley, 1971.

G.W. Corder and D.I. Foreman. *Nonparametric statistics for non-statisticians: a step-by-step approach*. Wiley, 2009.

D. Cramer. *Introducing statistics for social research: step-by-step calculations and computer techniques using SPSS*. Routledge, 1994.

Colleen Crangle. Conversational interfaces to robots. *Robotica*, 15:117–127, January 1997.

Michael Cross. Bridge. *http://www.michaelcross.eu/morebridge2.html*, September 2011.

Andrew Crossan, Roderick Murray-Smith, Stephen Brewster, James Kelly, and Bojan Musizza. Gait phase effects in mobile interaction. In *CHI '05 extended abstracts on Human factors in computing systems*, CHI EA '05, pages 1312–1315, New York, NY, USA, 2005. ACM.

Andrew Crossan, John Williamson, Stephen Brewster, and Rod Murray-Smith. Wrist rotation for interaction in mobile contexts. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*, MobileHCI '08, pages 435–438, New York, NY, USA, 2008. ACM.

Peter Dalsgaard and Lone Koefoed Hansen. Performing perception—staging aesthetics of interaction. *ACM Trans. Comput.-Hum. Interact.*, 15:13:1–13:33, December 2008.

Fred D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 09 1989.

Alan Dix. Designing for appropriation. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 2*, BCS-HCI '07, pages 27–30, Swinton, UK, UK, 2007. British Computer Society.

P. Dourish. *Where the action is: the foundations of embodied interaction*. Bradford Books. MIT Press, 2004.

J.L. Freedman and S.C. Fraser. Compliance withouth pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4:195–202, 1996.

Dustin Freeman, Hrvoje Benko, Meredith Ringel Morris, and Daniel Wigdor. Shadowguides: visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 165–172, New York, NY, USA, 2009. ACM.

Colombine Gardair, Patrick G.T. Healey, and Martin Welton. Performing places. In *Proceedings of the 8th ACM conference on Creativity and cognition*, C&C '11, pages 51–60, New York, NY, USA, 2011. ACM.

A. Giddens. *The constitution of society: outline of the theory of structuration*. University of California Press, 1986.

E. Goffman. *Behavior in public places: notes on the social organization of gatherings*. Free press paperback. Free Press, 1966.

E. Goffman. *Frame analysis: an essay on the organization of experience*. Northeastern University Press, 1986.

E. Goffman. *The presentation of self in everyday life*. Penguin psychology. Penguin, 1990.

J. Grudin and L. Palen. Why groupware succeeds: Discretion or mandate? In *Proceedings of the European Conference on Computer Supperted Cooperative Work.*, pages 263–278. Kluwer Academic Publishers, 1995.

Geert Hofstede. Geert Hofstede's Cultural Dimensions. *http://www.geert-hofstede.com/*, 2011.

Eve Hoggan and Stephen A. Brewster. Crosstrainer: testing the use of multimodal interfaces in situ. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 333–342, New York, NY, USA, 2010. ACM.

Giulio Iacucci, Carlo Iacucci, and Kari Kuutti. Imagining and experiencing in design, the role of performances. In *Proceedings of the second Nordic conference on Human-computer interaction*, NordiCHI '02, pages 167–176, New York, NY, USA, 2002. ACM.

Carlo Jacucci, Giulio Jacucci, Ina Wagner, and Thomas Psik. A manifesto for the performative development of ubiquitous media. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*, CC '05, pages 19–28, New York, NY, USA, 2005. ACM.

Giulio Jacucci. *Interaction as Performance. Cases of configuring physical interfaces in mixed media*. PhD thesis, University of Oulu, 2004.

Giulio Jacucci, Ann Morrison, Gabriela T. Richard, Jari Kleimola, Peter Peltonen, Lorenza Parisi, and Toni Laitinen. Worlds of information: designing for engagement at a public multi-touch display. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 2267–2276, New York, NY, USA, 2010. ACM.

O.P. John, R.W. Robins, and L.A. Pervin. *Handbook of personality: theory and research*. Guilford Press, 2008.

Matt Jones. Journeying toward extravagant, expressive, place-based computing. *interactions*, 18:26–31, January 2011.

A. Kendon. Current issues in the study of gesture. *Journal for the Anthropological Study of Human Movement*, 5(3):101–134, 1986.

Adam Kendon. Gesture. *Annual Review of Anthropology*, 26:109–128, January 1997.

David B. Koons and Carlton J. Sparrell. Iconic: speech and depictive gestures at the human-machine interface. In *Conference companion on Human factors in computing systems*, CHI '94, pages 453–454, New York, NY, USA, 1994. ACM.

Richard L. Lanigan. Is Erving Goffman a phenomenologist? *Critical Studies in Mass Communication*, 5(4):335–345, 1988.

Effie Law, Virpi Roto, Arnold P.O.S. Vermeeren, Joke Kort, and Marc Hassenzahl. Towards a shared definition of user experience. In *CHI '08 extended abstracts on Human factors in computing systems*, CHI EA '08, pages 2395–2398, New York, NY, USA, 2008. ACM.

Effie Lai-Chong Law, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 719–728, New York, NY, USA, 2009. ACM.

Saija Lemmelä, Akos Vetek, Kaj Mäkelä, and Dari Trendafilov. Designing and evaluating multimodal interaction for mobile contexts. In *Proceedings of the 10th international conference on Multimodal interfaces*, ICMI '08, pages 265–272, New York, NY, USA, 2008. ACM.

Tuck W. Leong, Lalya Gaye, Atau Tanaka, Robyn Taylor, and Peter C. Wright. The user in flux: bringing HCI and digital arts together to interrogate shifting roles in interactive media. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 45–48, New York, NY, USA, 2011. ACM.

D.K. Lewis. *Convention: a philosophical study*. Blackwell, 2002.

M. Lewis and T. Staehler. *Phenomenology: An Introduction*. Continuum, 2011.

Thomas Luckmann. *Phenomenology and Sociology*. Penguin Books, 1978.

Yogesh Malhotra and Dennis F. Galletta. Extending the technology acceptance model to account for social influence: Theoretical bases and empirical validation. *Hawaii International Conference on System Sciences*, 1:6–14, 1999.

Rob McCarney, James Warner, Steve Iliffe, Robbert van Haselen, Mark Griffin, and Peter Fisher. The hawthorne effect: a randomised, controlled trial. *BMC Medical Research Methodology*, 7 (1):30, 2007.

D. McNeill. *Hand and mind: what gestures reveal about thought*. University of Chicago Press, 1996.

S. Milgram. Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67 (4):371–378, October 1963.

Zhenyao Mo and Ulrich Neumann. Lexical gesture interface. In *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems*, pages 7–14, Washington, DC, USA, 2006. IEEE Computer Society.

Calkin S. Montero, Jason Alexander, Mark T. Marshall, and Sriram Subramanian. Would you do that?: understanding social acceptance of gestural interfaces. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 275–278, New York, NY, USA, 2010. ACM.

Roderick Murray-Smith, Andrew Ramsay, Simon Garrod, Melissa Jackson, and Bojan Musizza. Gait alignment in mobile phone conversations. In *Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, MobileHCI '07, pages 214–221, New York, NY, USA, 2007. ACM.

J.G. Neal, C.Y. Thielman, D.J. Funke, and J.S. Byoun. Multi-modal output composition for human-computer dialogues. In *AI Systems in Government Conference, 1989.,Proceedings of the Annual*, pages 250 –257, 1989.

J.L. Nespoulous, P. Perron, A.R. Lecours, and Toronto Semiotic Circle. *The Biological foundations of gestures: motor and semiotic aspects*. Neuropsychology and neurolinguistics. L. Erlbaum Associates, 1986.

Marianna Obrist, Effie Law, Kaisa Väänänen-Vainio-Mattila, Virpi Roto, Arnold Vermeeren, and Kari Kuutti. UX research: what theoretical roots do we build on – if any? In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 165–168, New York, NY, USA, 2011. ACM.

Marianna Obrist, Virpi Roto, and Kaisa Väänänen-Vainio-Mattila. User experience evaluation: do you know which method to use? In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, CHI EA '09, pages 2763–2766, New York, NY, USA, 2009. ACM.

Wanda J. Orlikowski. Learning from notes: organizational issues in groupware implementation. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, CSCW '92, pages 362–369, New York, NY, USA, 1992. ACM.

Sharon Oviatt. *The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications*, chapter Multimodal Interfaces. Human factors and ergonomics. Lawrence Erlbaum Associates, 2003.

Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford. When do we interact multimodally?: cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces*, ICMI '04, pages 129–136, New York, NY, USA, 2004. ACM.

Mark Perry, Steve Beckett, Kenton O'Hara, and Sriram Subramanian. Wavewindow: public, performative gestural interaction. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 109–112, New York, NY, USA, 2010. ACM.

Stuart Reeves, Steve Benford, Claire O'Malley, and Mike Fraser. Designing the spectator experience. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 741–750, New York, NY, USA, 2005. ACM.

Julie Rico and Stephen Brewster. Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, New York, NY, USA, 2010a. ACM.

Julie Rico and Stephen Brewster. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 887–896, New York, NY, USA, 2010b. ACM.

E.M. Rogers. *Diffusion of Innovations, 5th Edition*. Simon & Schuster, 2003.

Sami Ronkainen, Jonna Häkkilä, Saana Kaleva, Ashley Colley, and Jukka Linjama. Tap input as an embedded interaction method for mobile devices. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 263–270, New York, NY, USA, 2007. ACM.

Nithya Sambasivan, Leena Ventä, Jani Mäntyjärvi, Minna Isomursu, and Jonna Häkkilä. Designing for social context of mobility: mobile applications for always-on users. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, OZCHI '09, pages 137–144, New York, NY, USA, 2009. ACM.

S.L. Schensul, J.J. Schensul, and M.D. LeCompte. *Essential ethnographic methods: observations, interviews, and questionnaires*. Ethnographer's toolkit. AltaMira Press, 1999.

A. Schütz. *The phenomenology of the social world*. Northwestern University Press, 1967.

Jennifer Sheridan, Nick Bryan-Kinns, Stuart Reeves, Joe Marshall, and Giles Lane. Graffito: crowd-based performative interaction at festivals. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 1129–1134, New York, NY, USA, 2011. ACM.

Jennifer G. Sheridan, Nick Bryan-Kinns, and Alice Bayliss. Encouraging witting participation and performance in digital live art. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1*, BCS-HCI '07, pages 13–23, Swinton, UK, UK, 2007. British Computer Society.

S. Sherwood, S. Reeves, J. Maitland, A. Morrison, and M. Chalmers. Adapting evaluation to study behaviour in context. *International Journal of Mobile Human Computer Interaction*, 1(2): 37–55, 2009.

Gregory Smith. Enacted others: Specifying goffman's phenomenological omissions and sociological accomplishments. *Human Studies*, 28:397–415, 2005.

Adam J. Sporka, Sri H. Kurniawan, Murni Mahmud, and Pavel Slavík. Non-speech input and speech recognition for real-time control of computer games. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, Assets '06, pages 213–220, New York, NY, USA, 2006. ACM.

Adam J. Sporka, Sri H. Kurniawan, Murni Mahmud, and Pavel Slavik. Longitudinal study of continuous non-speech operated mouse pointer. In *CHI '07 extended abstracts on Human factors in computing systems*, CHI EA '07, pages 2669–2674, New York, NY, USA, 2007. ACM.

D.W. Stewart, P.N. Shamdasani, and D.W. Rook. *Focus groups: theory and practice*. Applied social research methods series. SAGE Publications, 2007.

Steven Strachan, Roderick Murray-Smith, and Sile O'Modhrain. Bodyspace: inferring body pose for natural control of a music player. In *CHI '07 extended abstracts on Human factors in computing systems*, CHI EA '07, pages 2001–2006, New York, NY, USA, 2007. ACM.

A.L. Strauss and J.M. Corbin. *Basics of qualitative research: techniques and procedures for developing grounded theory*. Sage Publications, 1998.

Bobby Swar, Jedong Kim, Duk Hee Lee, and Junghoon Moon. Impact of culture on mobile phone service adoption and diffusion: A cross-country analysis. In *Proceedings of the 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology*, ICCIT '09, pages 1442–1444, Washington, DC, USA, 2009. IEEE Computer Society.

K. Väänänen and K. Böhm. *Virtual Reality Systems*, chapter Gesture Driven Interaction as a Human Factor in Virtual Environments – An Approach to Neural Networks. Academic Press Limited, 1993.

Arnold P. O. S. Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila. User experience evaluation methods: current state and development needs. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 521–530, New York, NY, USA, 2010. ACM.

Tanja Walsh, Piia Nurkka, and Rod Walsh. Cultural differences in smartphone user experience evaluation. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, MUM '10, New York, NY, USA, 2010. ACM.

Alan Wexelblat. Research challenges in gesture: Open issues and unsolved problems. In *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, pages 1–11, London, UK, 1998. Springer-Verlag.

J. R. Williamson, A. Crossan, and S. Brewster. Multimodal mobile interactions: Usability studies in real world settings. In *ICMI 2011*, 2011.

John Williamson. *Continuous Uncertain Interaction*. PhD thesis, University of Glasgow, 2006.

John Williamson, Roderick Murray-Smith, and Stephen Hughes. Shoogle: excitatory multimodal interaction on mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 121–124, New York, NY, USA, 2007. ACM.

John Williamson, Simon Robinson, Craig Stewart, Roderick Murray-Smith, Matt Jones, and Stephen Brewster. Social gravity: a virtual elastic tether for casual, privacy-preserving pedestrian rendezvous. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 1485–1494, New York, NY, USA, 2010. ACM.

Julie Rico Williamson. Send me bubbles: multimodal performance and social acceptability. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, CHI EA '11, pages 899–904, New York, NY, USA, 2011. ACM.