PHD

Mobile Service Awareness via Auditory Notifications

Garzonis, Stavros

Award date:
2010

Awarding institution:
University of Bath

MOBILE SERVICE AWARENESS VIA AUDITORY NOTIFICATIONS

Stavros Garzonis

A thesis submitted for the degree of Doctor of Philosophy

University of Bath
Department of Computer Science

June 2010

……………………………………..

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible without the kindness, support, and direct or indirect contributions from a number of people. First and foremost, I am thankful to my supervisor Dr Eamonn O'Neill for igniting and nourishing my inquisitive nature in HCI. He has been my wisest guide in the ways of scientific research, thinking and writing. In this respect, his contribution in the shaping of this thesis is not only unique and fundamental, but also omnipresent throughout its contents.

Furthermore, my HCI family at the University of Bath has contributed to my research with endless conversations and debates. More specifically, I thank Simon Jones for coding the mobile application presented on Chapter 6, and for his overall input and presence throughout the study; Peter Goodison and Leo Lara for their contributions to PHP coding for the online surveys presented on Chapter 5; Chris Bevan and Tim Jay for their invaluable help with factor analysis and with the use of SPSS; and Fabio Nemetz for being my trusted and wise 'pilot' for many of the studies presented in the thesis. Moreover, people like Vassilis Kostakos, Andrew Warr, Jay Kaenampornpan, Tim Coughlan, Rachid Hourizi, Chris Middup and Em Srikulwong have been inspirational and supportive, each in their own special way.

I want to thank all the people who participated in my studies and gave me feedback for improvement. In particular, I would also like to thank Andrew Shakespear, Francis Binns, Daniel Goldstien, Erxiong (Ivan) Xu, Michael Crocker, Iain Kingston for their help in the study presented in Section 4.1.

Finally, there are plenty of people that I am hugely grateful towards for supporting my research in more indirect ways, mainly by supporting me through the challenges of research and life in general. Of course, amongst the most prominent of those are my family, who have always been there in many ways. The loving person most responsible for making sure I did not go mad during the writing up of this thesis (and tolerating me in the process) is Katherine Wood, whom I sincerely thank.

This research was funded by EPSRC and Vodafone Group R&D, with the support of Nokia and the Cityware project.

PUBLICATIONS

Two publications have been generated as a direct result of the research presented in this thesis.

Related to Chapter 4:

Garzonis, S., Bevan, C., O'Neill, E. (2008). Mobile Service Audio Notifications: intuitive semantics and noises. Proceedings of OZCHI08, the CHISIG Annual Conference on Human-Computer Interaction, December 2008, Cairns, Australia. 156-163.

Related to Chapters 5 and 6:

Garzonis, S., Jones, S., Jay, T., O'Neill, E. (2009). Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. In Proceedings of CHI 2009 conference on human factors in computing systems, April 2009, Boston, USA, 1513-1522.

ABSTRACT

Placed within the realms of Human Computer Interaction, this thesis contributes towards the goals of Ubiquitous Computing, where mobile devices can provide anywhere, anytime support to people's everyday activities. With interconnected computing devices distributed in our habitat, services relevant to any situation may be always available to address our needs. However, despite the enhanced capabilities of mobile phones, users had been reluctant to adopt any services other than calls and messaging. This has been changing more recently, especially since the launch of the iPhone, with users getting access to hundreds of services. The original question motivating the research presented in this thesis "How can we improve mobile service usage?" is in the interest of enthusiasts of mobile services as well as slow adopters.

We propose the concept of 'mobile service awareness' and operationalise it through the more focused research question: "How can we design for non-intrusive yet informative auditory mobile service notifications?" We design and conduct a series of surveys, laboratory experiments and longitudinal field studies to address this question. Our results, also informed by literature on context-aware computing, awareness, notification systems and auditory interface design, produce two distinct major contributions. First, we provide a set of conclusions on the relative efficiency of auditory icons and earcons as auditory notifications. Second, we produce a set of design guidelines for the two types of notifications, based on the critical evaluation of the methodologies we develop and adapt from the literature. Although these contributions were made with mobile service notification in mind, they are arguably useful for designers of any auditory interfaces conveying complex concepts (such as mobile services) and are used in attention demanding contexts.

Chapter 1

Introduction

Ubiquitous computing[1] is an emerging and promising technological environment, where multiple fixed and mobile devices or virtually augmented everyday objects are seamlessly interconnected. With the computing power distributed away from the desktop computer and into our natural habitat, information and services are becoming available to everyone, everywhere, anytime. The goal of ubiquitous computing is to provide the technological means to unobtrusively support people of any age or background in their daily activities where and when needed.

Human Computer Interaction (HCI) is one of the disciplines integral to the development of systems to address this goal. HCI is defined as "a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them"[2]. It involves the study of computer systems components (such as input and output devices) as well as the human nature (such as phenomena and theories on cognitive procedures in information processing, sensory and attentional limitations, and communication capabilities). Through their studies, HCI researchers often focus on designing user-centred technologies that do not demand unnecessary attentional and cognitive resources from users during the interaction. Some of the results of HCI research can be theories, descriptive and/or predictive models, and user-centred systems' development. Other contributions steer more towards the development of design or evaluation methodologies. The context of these contributions of HCI can be in the direction of improving the interaction with current technologies, or in exploring new paradigms of interaction with upcoming technologies.

The research presented in this thesis is situated within the realms of HCI, and more specifically contributes towards usable mobile interfaces that can unobtrusively enable everywhere, anytime information access for everyone. Mobile devices are getting more powerful and multi-functional, more distributed and interconnected. These new properties of mobile devices, in conjunction with their ever-shrinking physical size, are creating new interaction challenges.

On the other hand, mobile phones' evolution offers a virtual window onto a plethora of stand-alone applications and networked services. Mobile operators offer a pervasive infrastructure of everywhere/anytime connectivity in all developed countries ("Worldwide mobile telephone subscriptions reached 3.3 billion -- equivalent to half

---

[1] Term originally coined by Mark Weiser – see [Weiser, 1991]

[2] ACM SIGCHI Curricula for Human-Computer Interaction - http://www.sigchi.org/cdg/cdg2.html (last accessed 30/06/2009)

the global population"[1]), while developing countries are also demonstrating staggering uptake rates. For example, reports in popular media in 2004 reported a "wireless revolution that has made Africa the world's fastest-growing mobile phone market"[2]. Often skipping the technological generation of PC or even fixed phone lines, the mobile phone is dominant in daily interactions around the world, connecting and supporting people in their professional and personal lives. In other words, "the mobile phone can be seen as the first successful ubiquitous computer" [Brown & Randell, 2004].

Despite this continuous global growth of the mobile industry, users had been reluctant to adopt services other than voice calls and text messaging. According to a survey in 2005, 77% of UK mobile phone users had never used any of the data services, such as multimedia messaging and ringtone downloads[3]. In a different survey later the same year, 56% of users were using data services at least once a month[4]. More recently, there has been a rapid increase of widgets (task-specific applications) for mobile phones, mainly driven by the introduction of Apple's iPhone:

> "While it has been possible to create and download applications on to handsets for years, the arrival of the iPhone has revolutionised the market […] leaving the network itself as little more than a broadband 'pipe in the air'. […] Since its launch last year, more than 1bn applications, from games and online newspapers to taxi-finding services and music players, have been downloaded from the iTunes store"[5]

The motivation for the work presented in this thesis was originally rooted in the observed users' reluctance to adopt mobile data services. Previous work has identified that the methods users have at their disposal to find and initiate the appropriate services ('mobile service discovery') is one of the major obstacles towards mobile service usage [Garzonis & O'Neill, 2006]. Therefore, the work presented here focuses on delivering 'mobile service awareness' as a more efficient supporting tool for adopting mobile services. More importantly, it is also a tool aiming to support the more enthusiastic users of the 'iPhone era' to identify and manage the relevant services from the plethora on offer.

The major contribution of the thesis comes from our work on auditory mobile notifications, which are utilised as the main mechanism for delivering mobile service awareness. We develop an effective and systematic auditory notification design proc-

---

[1] Global cell phone use at 50 percent - Reuters 29 Nov 2007: http://www.reuters.com/article/technologyNews/idUSL2917209520071129 (last accessed 30/06/2009)

[2] "Phone revolution makes Africa upwardly mobile" - Times Online, 04 Mar 2006: http://www.timesonline.co.uk/tol/news/world/article737130.ece (last accessed 30/06/2009)

[3] "77% of UK mobile phone users shy away from using mobile data…" - M2 Presswire, 20 May 2005: http://goliath.ecnext.com/coms2/gi_0199-4278153/77-of-UK-mobile-phone.html (last accessed 30/06/2009)

[4] "Mobile data usage is on the rise" - InfoWorld, 07 Nov 2005: http://www.infoworld.com/d/networking/mobile-data-usage-rise-041 (last accessed 30/06/2009)

[5] "Vodafone enters mobile phone applications market" - guardian.co.uk, 12 May 2009: http://www.guardian.co.uk/business/2009/may/12/vodafone-applications-store-mobile-phones (last accessed 30/06/2009)

ess, which can have a positive and pragmatic impact for thousands of people who perform numerous interactions with their mobile phones on a daily basis. In this way, our research also makes a substantial contribution towards mobile usability and the (further) understanding and realisation of ubiquitous computing.

The remainder of this chapter is structured as follows. In the next section, we set the scene of our research by describing in more detail the domain of ubiquitous computing. In Section 1.2 we examine the motivating real world problem and describe the transition from the challenge of designing for service discovery to the one of service awareness. Then, in Section 1.3, we scope the issue of service awareness and focus on the aspects this thesis is going to address. This path of investigation leads to a clear and concise research question, which is presented in Section 1.4. The goals and methods applied to answer the research question are presented in Section 1.5. Finally, the contributions arising from our research are presented along with the chapter structure of the thesis in Section 1.6.

1.1 Setting the Scene

A number of research streams and a plethora of terms have emerged to explore and describe ubiquitous computing. The concepts of 'post-PC computing era', 'invisible computing', 'pervasive computing', 'ambient intelligence' and 'context-aware computing' overlap considerably, each one stresses different elements or viewpoints of ubiquitous computing.

'Post-PC computing era' refers to the fact that the computing power of the desktop PC is being gradually diffused to a variety of distributed devices. Therefore, more task-specific, interconnected devices exist to support people's daily activities. Our mobile devices can inform us about our peers' whereabouts, our fridges can order milk online and our cars can guide us to the nearest free parking spot. These examples of existing or upcoming technologies highlight the transition from 'one person – one computer' era to the 'one person – many computers' era (Figure 1-1), with many of these computers being everyday objects with computational capabilities.

'Invisible computing' purports to represent the desirability of ubiquitous computing systems when seamlessly interwoven in the physical environment. If distributed technology intrudes our environment and offensively demands our attention, ubiquitous computing fails to meet its goals. Instead, the use of technology should be effortless and unobtrusive enough to support rather than interrupt our daily activities. In other words, technology needs to hide its complexities and disappear into the background, so that it is not perceived as "technology" but as what is naturally there. For example, switching the lights, editing a document or placing a call are perceived as everyday activities that require no effort or focus on the technological media involved in the process (infrastructures, devices and interfaces). On the other hand, infrastructure limitations, unexpected software behaviour or complex menus and settings take our focus away from our goals and onto the technological medium. For example, attempting to "convince" the text editor to insert a footnote or picture on the correct page takes the focus away from the task of document editing and onto the erratic software.

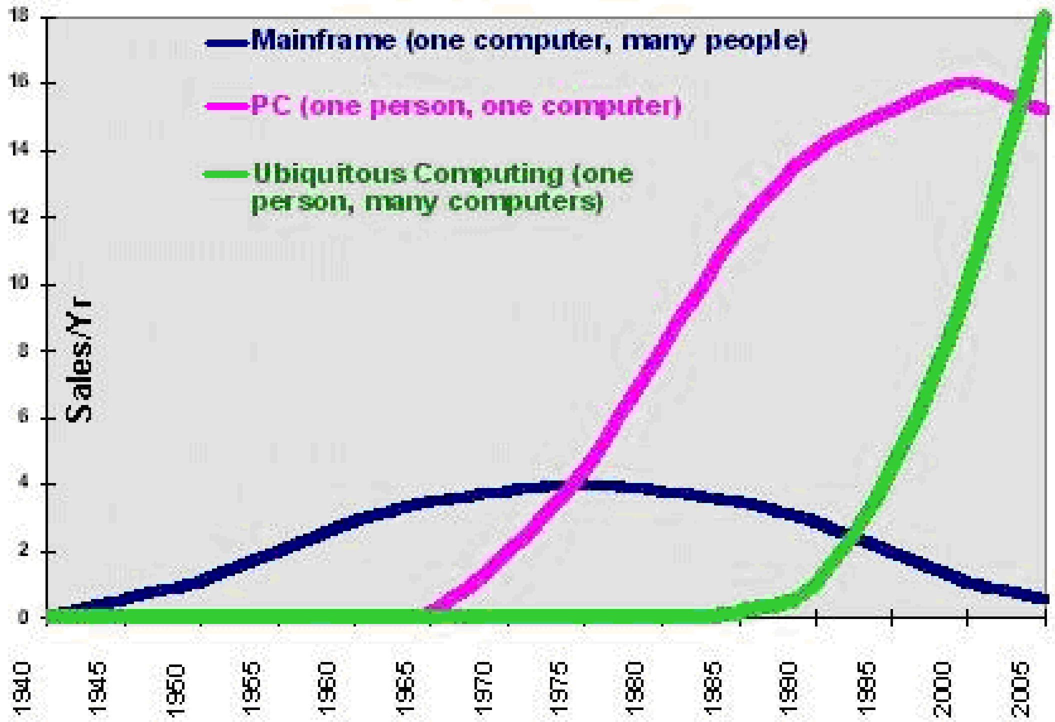


Figure 1-1: Major Trends in Computing[1]

Invisible computing has also been referred to as 'calm technology' or 'the disappearing computer' [Weiser & Brown, 1995].

'Pervasive computing' was introduced as "convenient access, through a new class of appliances, to relevant information with the ability to take action on when or where you need it" [Hansmann, 2003]. It is one of the terms that are most commonly being used interchangeably with ubiquitous computing, although others consider Pervasive as more fixed and Ubiquitous as mobile [Lyytinen & Yoo, 2002].

'Context-aware computing' (also known as 'ambient intelligence') facilitates tailored support to users of mobile and pervasive systems, by sensing and adapting to the dynamic context of use. This can add value to pervasive systems by staying invisible when not needed but offering the relevant services when and where needed most. Context-aware computing is as an integral part of the ubiquitous computing paradigm, and can be described in 3 stages. First, information regarding the user, the task at hand and the environment are gathered and monitored. Environmental properties (e.g. noise levels, lighting conditions) and computational environment (e.g. printers, wireless networks) are usually gathered by distributed (fixed or mobile) sensors, while user information (i.e. user profile) is provided by users explicitly or implicitly (e.g. history of interactions). Second, the gathered information is put together to infer the current situation of the user including (when possible) activity and intention. Third, the relevant devices present this information to users ('passive context awareness') or

[1] Image adapted from Mark Weiser: http://www.ubiq.com/hypertext/weiser/UbiHome.html (last accessed 30/06/2009)

automatically adapt ('active context awareness') in order to accommodate their needs. Examples of active context-aware systems include applications such as mobile phones' profile changing from 'outdoors' to 'silent' during a meeting, and information screens on mobile tourist guides refreshing to describe the new landmark in sight. An example of passive context-aware computing are the triggered mobile alerts, which timely notify us about relevant information (e.g. a reminder to 'buy milk' when this is on our 'to-do' list and we are driving past the local supermarket on the way home).

However, making sense of users' higher levels of context, such as activities and intentions, is particularly difficult. It is not surprising that we cannot develop systems that can infer user intentions, when indeed humans often fall short of correctly interpreting each other's intentions. Although artificial intelligence has made some considerable steps in the last couple of decades, it seems we are far from the vision of infallible context-aware computing. For example, a simple context-aware application such as the mobile phone profile switching takes up to 30 seconds to recognise the new context with an average success rate of 87% [Gellersen, 2002].

When active context awareness fails, systems are perceived to behave unpredictably and may hinder our activities. On the other hand, when passive context awareness fails, we are inconvenienced by unnecessary and distracting interruptions. This latter challenge is the focus of research areas such as 'ambient notification systems' and 'situation awareness'. These areas investigate how interfaces can efficiently make users aware of the current status or changes of computing systems. This efficiency depends on addressing challenges such as estimating user interruptibility against information urgency, and balancing it to the type and method of information presentation. For example, design decisions may involve presenting the information at the centre or the periphery of users' attention, via the appropriate sensory channel (visual, auditory, haptic), and appropriately encoded in a meaningful yet non-intrusive way. Notification or situation awareness systems can be utilised within 'Computer Supported Collaborative Work' (CSCW) applications, 'Information and Communication Technologies' (ICT), critical-safety systems (such as aircraft cockpits and nuclear plant controllers), and mobile applications.

This thesis explores the problems and solutions around mobile service usage, by combining and extending previous knowledge mainly from the domains of context-aware computing, awareness and notification systems.

1.2 Mobile Service Discovery or Mobile Service Awareness?

In previous work it has been argued that mobile service uptake has been slow due to three main factors [Garzonis & O'Neill, 2006]. Firstly, networks are not offering services that address the needs of most mobile users. Secondly, when the context that a service would be useful arises, users are unaware of its existence and/or how to initiate it. Thirdly, the usability problems in the process of initiating and/or using a service discourage (or prevent) users from initiating or completing the intended task on their mobile device.

User requirements and usability factors are common reasons for making users reluctant to engage with any computing device or application. In other words, the less a piece of software is needed and the more difficult it is to use, the less it is going to be used. However, paramount is the importance of the third dimension to the problem of mobile service usage, which arises from the difficulty users may have to discover the desired services. In the desktop environment it is common for users to apply exploratory learning to find out what applications are available and how to initiate them. However, this primarily trial-and-error browsing technique is less appropriate for mobile service discovery, mainly due to the limited resources, such as input and output means available, slow connection speed and browsing costs. Furthermore, these difficulties are further heightened by the context of use, in which users are expected to share their attentional/cognitive resources, eyes, ears and hands between their everyday activities and the mobile device. In such a complex and dynamic context of use, it is difficult to know what computational and/or user resources are going to be available in any given time.

Therefore, alternative interaction paradigms should be applied in order to support mobile service discovery. One such paradigm could emerge from the research domain of 'tangible computing'. Instead of a directly transferring (shrunk) user interfaces from PCs to mobile phones and PDAs, functionality and computational power can be distributed in digitally augmented everyday objects. Mobile services could therefore be "found" outside the mobile device and could be initiated by appropriately manipulating a picture frame or a box of chocolates. A current commercially available technology is based on barcodes and RFID (Radio Frequency Identification) tags. Mobile phones can be used as scanners and pick up information from any everyday object that has been equipped with such a tag. For example, one could buy cinema tickets by touching their phone on the relevant poster, or place a call to the plumber by touching it on their home boiler. Provided that these objects are visibly marked as virtually augmented, they can function both as a sign announcing the presence of services, and as an immediate method to initiate these services.

One of the problems with these physical portals to the virtual world though, is that they cannot always be marked or noticed amongst the plethora of everyday objects. Crowded and visually cluttered environments can make such objects and tagged posters difficult to spot and/or approach, while privately owned and regulated spaces could forbid them. Furthermore, with the rise of context-aware computing, certain services might be available only under certain conditions. For example, a tourist information service might be available only on weekends, at important city landmarks, or for non-locals. However, the physical counterparts of these 'contextually available' services cannot appear and disappear at different times of the day, or depending on who is looking at them. Apart from location, time and user, more complex context conditions can significantly limit the efficiency of virtually augmented objects as mobile service indicators and initiators. For example, services can become available in an *ad hoc* manner depending on the computational surrounding that is shaped at any given moment/place by the availability of mobile and fixed devices. Finally, the extra interaction step with the physical world may pose usability and /or privacy issues (as for example with NFC and 2D barcodes as reported in [O'Neill et al., 2007]).

Therefore, interaction with the mobile interface or digitally augmented objects can only provide partial and limited solutions to the problem of mobile service discovery. 'Mobile service awareness' is therefore proposed here as an alternative perspective on mobile service underuse. Instead of applying ineffective methods to facilitate mobile service discovery, users could become aware of the services that are relevant to a given context and activity, via a context-aware notification system. Upon service suggestion, the appropriate link would appear on the mobile device to facilitate instant initiation. One of the main benefits of such a system is that it could alleviate much of the attentional and cognitive resources an individual would need in order to scan his surroundings to locate and interact with tagged objects and eventually with their mobile device. Another major benefit of mobile service awareness is that it can effectively deal with a plethora of potentially available services, including any contextually available services. Regardless of how or when they become available, users will be informed of their availability if they become relevant to their activities. On the other hand, it has already been discussed how context-aware systems cannot always successfully infer users' intentions or interruptibility. Consequently, the mobile service awareness approach is bound to result in some undesired interruptions, either due to the irrelevance of the suggested services, or the inappropriate timing of the notification.

Although a system combining mobile service discovery and mobile service awareness techniques should provide a well-balanced and more effective method to support mobile service usage, this thesis will focus on addressing the issues surrounding mobile service awareness, especially via the auditory interface. We find this area to be relatively unexplored, and the contributions to be made necessary and exciting for mobile service usage and management.

1.3 Scoping Mobile Service Awareness

The main problem with mobile service awareness approach is that its context-aware system is likely to produce notifications for irrelevant services. One way to address this issue is to equip the context-aware system with a mechanism that can transform unwanted interruptions to discreet and implicit awareness information. The goal therefore is to design and develop a delivery mechanism that can provide service notifications that are informative yet unobtrusive, in accordance to the vision of the 'invisible computing'.

A simplified black box representation of a context-aware system would include an input stream, a 'context engine' and an output stream. The input stream consists of a collection of context data (gathered by sensors or given by users), such as current location, time, activity and user preferences. The context engine will then infer the user context in a more meaningful expression (e.g. 'driving to work'), usually selecting the most appropriate of the options it is programmed to recognise. In our case, the context engine will also infer which services are relevant to the specific context of a user.

The output of the system can consist either of the inferred information (passive context awareness) or of the relevant commands that will instruct the automatic adaptation of the interface or another part of the wider system (active context awareness).

A context-aware notification system can then take the output of the context engine and feed it into a 'notification engine', which responsibility is to select the appropriate notification for the suggested service, in the given context. The procedure of delivering mobile service awareness, from data gathering to service notification is presented in Figure 1-2.

The relationship between context-aware systems and notification systems is further underlined by McCrickard & Chewar's explicit decision not to include situational context as one of the critical parameters in their framework for notification systems. Although they argue that social context is "certainly an important facet in the success of a notification goal", the challenges of anticipating, measuring or manipulating context variables led them to the decision to keep them separate from the characterisation of a notification system [McCrickard & Chewar, 2003]. Our model supports this conceptual separation, but we argue that context descriptions need to be passed on the notification engine as one of the parameters that will improve the chance of presenting the most suitable notifications.

The first level of scoping for the work presented in this thesis excludes the investigation of the context-aware system. The improvement of algorithms that infer user context and intentions is important but separable from notification design. Our interest remains in the types of notifications that can be utilised once the relevant services have been chosen by the context engine.

The second level of scoping is with regard to the variety and diversity of all the notification stimuli that could be possibly produced. In a fully developed mobile service awareness system, the notification engine could decide which of the available interface modalities to utilise: visual, auditory, tactile and combinations thereof (smell and taste have hardly been utilised as modalities in human computer interaction). For example, if we assume three distinct stimuli for each modality (e.g. 3 colours, 3 sounds, 3 vibration patterns), we can create 27 unique (simultaneous) multimodal signals.

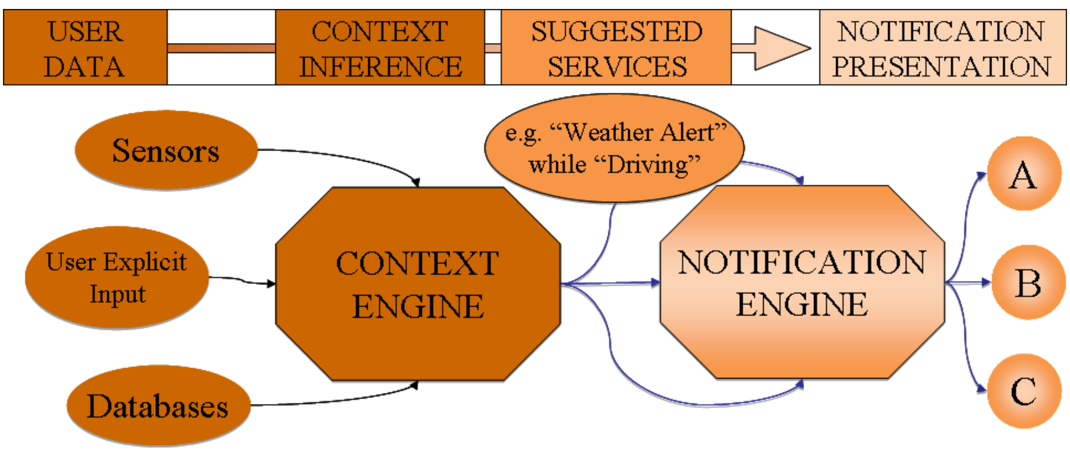


Figure 1-2: Delivering mobile service awareness

Each one of these modalities has its limitations. Visual stimuli from devices on the periphery of our vision can be easy to miss, especially in the mobile context where we heavily rely on our vision for everyday tasks. On the other hand, hearing has the ability to process multiple auditory signals at the same time, given enough diversity amongst them (e.g. speech and non-speech signals) [Fitch & Kramer, 1994], or focus on a single speaker amongst multiple conversations or noise (known as the cocktail party effect) [Cherry, 1953]. However, audio notifications are not always welcome (e.g. in a meeting or cinema) or can be missed (e.g. in a very noisy environment). Contrary to both vision and hearing, tactile input is a modality that is in very limited use in everyday activities and therefore available to convey information to users at any time (provided they are in physical contact with the notification device). Although it is still a research area relatively unexplored, it has been found that we can successfully recognise 9 distinct tactile signals with 71% accuracy (with one actuator in direct contact with the skin) [Brown et al., 2005]. There are also indications that this rate can go up if multiple actuators are distributed on the user's body, and spatial vibration is used as one of the information encoding parameters [Hoggan & Brewster, 2007]. On the other hand, the visual and auditory channels demonstrate much more flexibility in their ability to encode information, from semantically symbolic flash patterns and bursts of sounds, to the semantically rich text and speech stimuli.

Given these advantages and limitations of each modality, notification systems could maximise their efficiency by appropriately combining them depending on context of use. For example, if a modality is unavailable in a certain situation, a multimodal notification has a better chance in succeeding to convey the message through the available modalities (e.g. vibration and flashing light in the absence of sound when the user is at the cinema theatre). Furthermore, designers could distribute the encoding of different attributes of the notification information (e.g. type of service, urgency, identification of sender) on all available modalities, as long as the meaning of the notification does not get misinterpreted in the absence of some of them.

However, mobile services notifications are relatively unexplored and before a multimodal approach is explored, there are many challenges to be addressed within each modality. A better, deeper understanding of single-modality notifications is a necessary prerequisite towards efficient multimodal notifications. In this thesis the auditory channel was selected on two grounds. First, it demonstrates great flexibility in its ability to encode information in varying levels of richness (i.e. how much information it can convey). Since our goal is to provide meaningful yet non-interruptive notifications, this flexibility of the auditory channel can be utilised to manage the trade-off between awareness and intrusion. Second, mobile phones are often hidden away in pockets or bags, or otherwise left out of sight and reach. Therefore, visual stimuli (which also demonstrate great richness flexibility) and tactile stimuli can often go unattended. In these situations, auditory notifications are easier to perceive and, if designed efficiently, inform without requesting user's redirection of attention. Once efficient auditory notifications for mobile services have been established via a systematic design process, further research in multimodal notifications should follow to explore the implications of simultaneous presentations on different sensory channels.

1.4 Research Question

The initial question that motivated this research was:

Q1) "How can we improve mobile service usage?"

After identifying mobile service discovery as one of the factors responsible [Garzonis & O'Neill, 2006], we investigated ways to address it. However, the existing and upcoming methods of interaction we investigated were found inherently limited solutions [O'Neill et al., 2007], especially with the uptake of contextually available services. At the same time, mobile users started being more willing to download and use task-specific applications (widgets), which increased mobile service usage (and it is predicted to keep increasing in the future) . As a result, our interest focused on mobile service awareness, which could address both Q1 and also manage the influx of services being used. The research question was therefore evolved into:

Q2) "How can we increase service awareness of contextually available mobile services?"

However, providing a system that fully addresses this question would be beyond the resource limitations of this thesis. Having sketched a context-aware notification system to address Q2, we focus only on the notification side of the system, leaving the workings and improvements of the context engine outside the scope of the thesis. Given a current or future context-aware system that can adequately decide on which services are relevant in different situations, we address the question:

Q3) "How can we design for non-intrusive yet informative mobile service notifications?"

Finally, scoping the project on the most suitable and promising modality, especially for out of sight and out of reach devices, the research question has taken its final form:

Q4) "How can we design for non-intrusive yet informative auditory mobile service notifications?"

1.5 Goals and Methods

The initial motivation behind this research has been to understand mobile service underuse, and find ways to address it. During our investigations in this area, our understanding has evolved and the direction of research has been adapted accordingly. Furthermore, the mobile industry and market have been growing and changing rapidly, with mobile services benefitting by faster connections (3G) and open developing platforms (e.g. Android[1] by Google). In this section, we briefly summarise and present the evolution of our lower level goals and the research methods applied to achieve them.

[1] www.android.com (last accessed 30/06/2009)

As our focus turned towards mobile service awareness and notification systems, our first goal was to develop in-depth understanding of auditory mobile service notifications. Therefore, we conducted an exhaustive ***literature review*** on context-aware and notification systems, as well as particular types of audio notifications, which informed our first ***empirical laboratory experiment*** on assessing auditory mobile notifications.

Our two major goals were to assess the effectiveness of non-speech auditory notifications and develop guidelines for their design. The methods applied to develop design methodologies for two types of sounds included ***analytical brainstorming*** sessions, two ***online surveys*** and two further ***laboratory experiments***. Also, analytical literature review of commercial services and two ***card-sorting studies*** were conducted in order to derive in a service classification. The sounds and their respective design processes were assessed in a ***longitudinal study***, which included two more ***laboratory experiments***, a weeklong ***field study*** and two ***web-based experiments***.

1.6 Thesis Overview and Contributions

The thesis as a whole contributes towards supporting and managing mobile service usage. By introducing the concept of mobile service awareness, it provides the first systematic investigation in auditory mobile service notifications. In particular, we describe the design processes for two types of auditory notifications for mobile services: auditory icons and earcons. By comparing the two sets of sounds we produced via the respective methodologies, we supply effectiveness results of the notifications and a set of design guidelines for them. The value of these results and guidelines extends beyond mobile services and can be applied in notification systems that utilise an auditory interface in similar contexts of use. Next, we give an overview of the thesis, which presents the contributions in more detail as they appear in each chapter.

Chapter 2 introduces the concept of mobile service awareness and it provides an extensive literature review on the related areas, as a response to our redefined research question (Q2). Context-aware computing background is provided to demonstrate how it would feature in the complete solution of mobile service awareness, as depicted in Figure 1-2. We then present the literature of research areas that developed theories and systems to keep users aware of background or extra incoming information, systems' states and events: peripheral awareness, situation awareness and notification systems. Issues and solutions identified in the literature set the requirements and lay the road for mobile service awareness through notifications.

The literature review continues on Chapter 3, with specific focus on auditory notifications. We first support our decision to select the auditory modality (scoping from Q3 to Q4), by presenting properties of the human hearing channel and literature on the context of mobile use. Then we focus on speech and two non-speech types of notifications established in the literature: earcons and auditory icons. We present background work on designing and using each one separately, and many studies that have attempted to compare their effectiveness in a variety of contexts. Furthermore, we extend the literature by providing an insightful representation of the relationships between different sound types that are generally utilised in auditory interface research.

Chapter 4 is split in two parts, both of which make unique contributions to the literature, and serve as a preparation step towards the studies that will address our research question (Q4). First, we describe our first investigation into auditory mobile service notifications, comparing the three types of sound (speech, auditory icons, earcons). Results and discussion on the intuitiveness of the notifications are presented for the first time in the literature. Also, the experience gained in designing and evaluating these notifications, as well as the results we obtained, informed our studies presented in Chapters 5 and 6. In the second part of Chapter 4, we describe the process by which we created a novel, meaningful classification of mobile services, also to be utilised in the studies following. By applying analytical and empirical methods, we designed a hierarchy of existing (and some upcoming) services, so that different notifications can be assigned to each cluster of services.

In Chapter 5, we describe the design processes for auditory icon and earcon mobile service notifications, which make use of the classification derived in Chapter 4. In order to compare the effectiveness and appropriateness of the two types of notifications, we needed to ensure that we have obtained reasonably good representative instances for each sound family. In particular, the auditory icon process is designed by combining and adapting elements of the relevant literature in a novel and more complete way. Our experience in designing both types of sounds provides invaluable input for the final study.

Chapter 6 presents a 4-stage study that was designed to compare the effectiveness of the two sets of mobile notifications we obtained in the previous chapter. The studies of the previous chapters informed the design of the study presented here, which focused on four metrics: intuitiveness, learnability, memorability and user preference. These factors were measured during two laboratory experiments, a weeklong field study, and two web-based experiments (one and four weeks later). This in-depth longitudinal study contributes to the auditory interface and mobile notifications research communities in multiple ways. First, no previous auditory notification comparison studies have gone in such breadth (four metrics) and depth (longitudinally). Furthermore, we are not aware of any studies that have compared auditory icons and earcons that have been obtained through systematic and empirical methods. Our results are therefore more reliable as we ensured that the sound instances were good representatives of the sound types. Therefore, our results on the effectiveness of the sound types are generalisable enough to be translated into conclusions for auditory icon and earcon notifications. Our final and most valuable contribution is a set of design guidelines for auditory notifications, derived by assessing the processes of Chapter 6. These provide a unique and novel guidance for designing efficient mobile service notifications, and significantly extend the literature suggestions on auditory icon and earcon design.

Finally, Chapter 7 provides a summary of the thesis and conclusions on mobile service awareness are drawn. We also discuss future work directions, such as streamlining the design processes and extending them for speech and non-audio notifications.

The outline of the thesis chapters and their relationships are presented in Figure 1-3.

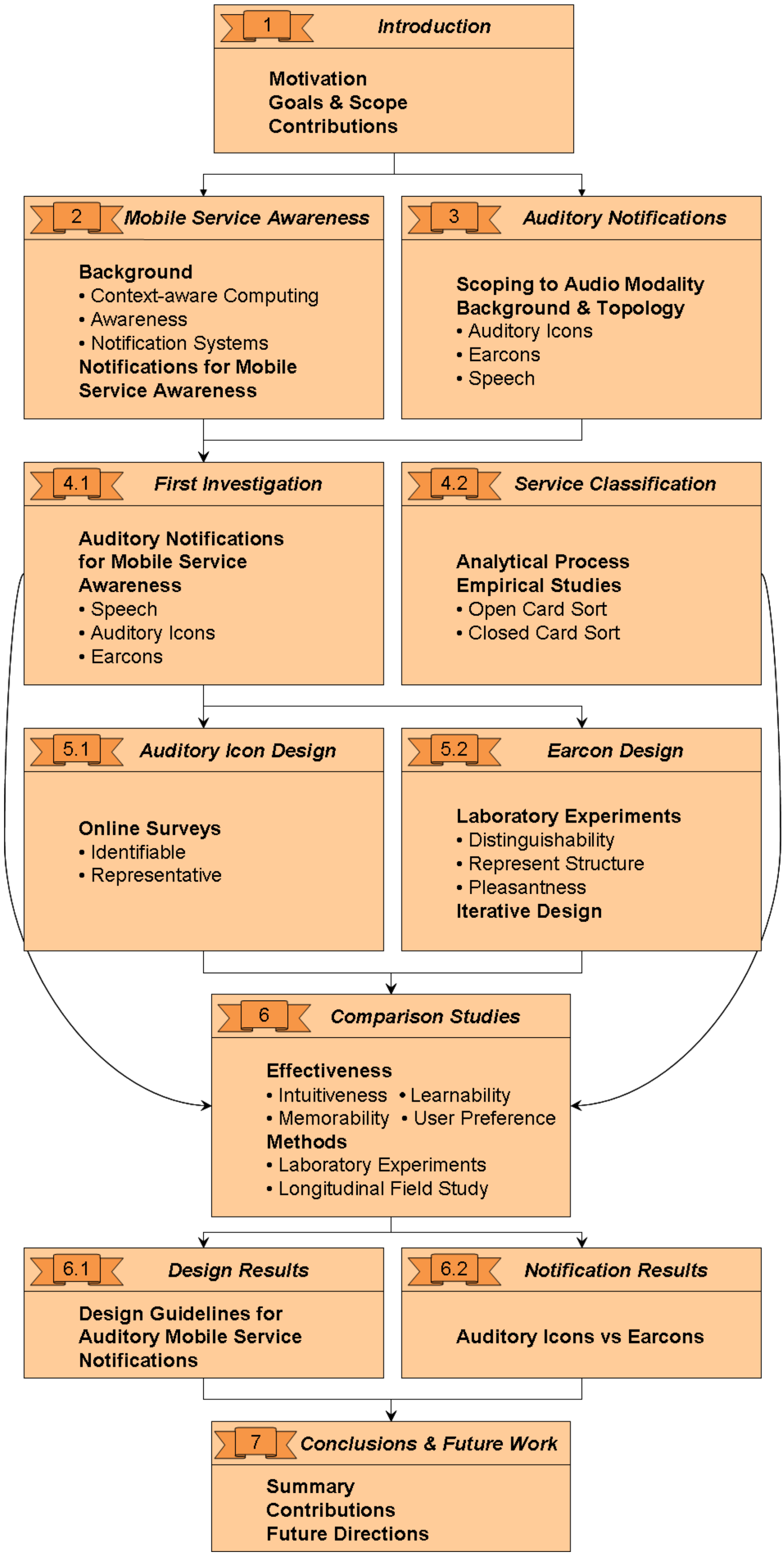


Figure 1-3: Thesis outline

Chapter 2

Mobile Service Awareness

2.1 Mobile Services

The term 'service' can have very wide and diverse meanings. The definition of services we build upon assumes some kind of interaction between a technology user and a provider, such as: "specialized, software-based functionality provided by network servers…"[1]. If we generalise this definition to describe services as a process, but specify it for the mobile context with a user-centred approach, we form the following definition: ***Mobile Service** is the process of utilising a handheld device (with or without other artefacts and/or actors), in order to help a user achieve a certain goal.* An 'actor' will in most cases be an individual, a company or a software component (agent). The handheld device can be a mobile phone, a PDA or any other highly portable device, which are usually operated in a dynamic context of use, where user attention may be split between everyday activities and the device. Of course, other artefacts, such as devices (mobile or fixed) or components (e.g. wireless hub), can be part of a particular mobile service, whether they are owned by the user or other actors. There can be one or more providers that provide the artefacts not belonging to the user. Without the provider(s), the service could not be invoked and this is why there is usually a direct or indirect charge for services.

A simple everyday example of a service is placing a phone call via a mobile phone: an individual (user) wants to communicate with another person (goal) through one's mobile phone (own artefact). The network provider provides all the other elements (e.g. servers, antennas) to facilitate the process of voice transfer. Other examples of mobile services include receiving and sending text messages, weather or sports alerts, downloading or playing games, calendar or to-do reminders, online purchases from the mobile phone, and device personalisation, such as wallpapers and ringtones (for a short history on the use of ringtones, see Appendix I.1). Furthermore, as mobile technology becomes more context-aware, services expand to triggered reminders (e.g. reminders from the to-do list to buy groceries when walking past a grocery shop), friends' current location, live travel updates etc.

PC users can normally find their desired service by applying a variety of techniques, such as browsing, using search engines or following indices of services. However, these explicit user actions are harder to perform over a slow connection and through the limited input/output mechanisms available on mobile devices. Moreover, certain mobile services may be only available in specific contexts ('contextual availability'). Therefore, the process of service discovery on mobile devices has been found to be a major factor contributing to low usage of mobile services [Garzonis & O'Neill, 2006].

---

[1] Glossary of IT & Networking Terms - http://newforestsystems.com/network_acr.aspx (last accessed: 30/06/2009)

Furthermore, the availability of these services depends to some degree on the context of use. Aspects of context that may determine service availability include personal, spatial and temporal dimensions. Different services will be available to different people, in different places and at different times. For example, at locations where there is no network coverage or at times that the network is overloaded, phone calls cannot be placed. The availability of conventional services is to a great extent well known and predictable. However, as services increase in numbers and become more contextually available, it becomes harder to know which services are available in any given moment, location and to any particular user. For example, one might have access to a local voting service upon entering a local forum or the ability to print or exchange documents in an *ad hoc* manner depending on the presence of other devices. In these cases, changes of location, social and technical context affect the availability of a service, and hence our knowledge of this. Furthermore, certain services might be broadly available (anyplace, anytime, any user) but only useful to invoke them at certain contexts. Therefore, some mechanism to provide service awareness is needed for services to be utilised where and when needed most.

2.2 Designing for Mobile Service Awareness

There have been consistent anecdotal data and surveys indicating that users have been reluctant in adopting new services extending beyond calls and messaging. One survey in 2005 portrayed 77% of UK mobile phone users to have never used any mobile data services[1]. Later that year, it was reported that 56% of users (in the Americas, Europe and Asia) connected to the Internet, used email or MMS via their mobile phone at least once per month[2]. However, in 2009, iPhone popularised the use of widgets (task-specific applications and services) and it has been reported that "since its launch last year, more than 1bn applications, from games and online newspapers to taxi-finding services and music players, have been downloaded from the iTunes store"[3]. Despite this increase, mobile data services have been found "vulnerable" in terms of loyal use. For example, one survey reports that 48% of American users "would drop mobile data service completely" if they had to reduce household expenditures[4]. This demonstrates that although outside the scope of this thesis, pricing and marketing are also important factors determining the uptake of these services.

Therefore, a design solution is needed to support mobile service usage for both groups of users: those who are still reluctant to expand on the type of mobile services they use, and those who are already ridding the iPhone application 'wave'. The former

[1] "77% of UK mobile phone users shy away from using mobile data reveals new research" - M2 Presswire, 20 May 2005: http://goliath.ecnext.com/coms2/gi_0199-4278153/77-of-UK-mobile-phone.html (last accessed 30/06/2009)

[2] "Mobile data usage is on the rise" - InfoWorld, 07 Nov 2005: http://www.infoworld.com/d/networking/mobile-data-usage-rise-041 (last accessed 30/06/2009)

[3] "Vodafone enters mobile phone applications market", - guardian.co.uk, 12 May 2009: http://www.guardian.co.uk/business/2009/may/12/vodafone-applications-store-mobile-phones (last accessed: 30/06/2009)

[4] 48% of Americans Would Drop Mobile Data Service Completely - Strategy Analytics, 15 Jun 2009: http://www.strategyanalytics.com/default.aspx?mod=PressReleaseViewer&a0=4751 (last accessed: 30/06/2009)

group of users may be unaware of services' existence and the potential value they have to offer in their everyday activities (there is of course the possibility that these services are not needed or valued – for implications on user needs see Appendix I.2). The latter group of users may be unaware of which of the hundreds of services are most relevant to a given situation/activity, or when and where the most appropriate services are available. Nevertheless, enthusiastic and reluctant users alike often lack the information needed for the smooth initiation of the relevant services. These two issues (being unaware of relevant services' availability and the process of their initiation) can also be described as gaps that exist between our intentions and our ability to act towards the corresponding goals. Three types of such gaps have been identified by [Gershman et al., 1999]: physical discontinuity, lack of awareness, and lack of information. In other words, if a service is needed but not used, it could be due to the fact that we are either at the wrong place, unaware of the service availability or unaware of how to initiate it. It has been argued that these gaps can be bridged if we "enhance situations in the physical world with relevant connections to the virtual world" [Gershman et al., 1999].

There are two obvious ways for mobile users to determine which services are available in any given context: explore and look for cues in the immediate physical environment, or explore the virtual world through their mobile device. However, there are indications that there are inherent problems in providing these connections between the physical and virtual worlds with both in-environment [e.g. O'Neill et al., 2007] and on-device [e.g. Garzonis & O'Neill, 2006] mobile service discovery (for more on mobile service discovery and the related usability issues see Appendix II).

Here, we present the concept of 'Mobile Service Awareness' as an alternative approach to bridge these gaps. We define mobile service awareness any user has at any given moment as "the knowledge of the available relevant mobile services and how to initiate them". Therefore, mobile service usage could improve by providing users with *knowledge* of availability (awareness gap) of *relevant* services (physical discontinuity gap) and the ability for *instant initiation* (information gap). Therefore our initial research question "How can we improve mobile service usage?" is refocused to "How can we increase service awareness of contextually available mobile services?"

We argue that a context-aware notification system (with an architecture such as the one depicted in Figure 1-2) can provide mobile service awareness by operating in two levels. First, the context engine derives which services are relevant to the users' current context, such as location, activity and intentions. Then, the notification engine decides on the appropriate medium and timing to make users aware of the service availability and provide with the corresponding initiation link.

Therefore, there are two distinct challenges for the proposed mobile service awareness system: first to choose the "correct" services and then to provide the "correct" awareness. Keeping systems aware in order to engage the relevant technology is integral to ubiquitous computing and has been the main focus of *context-aware computing*. Keeping users aware of systems' states or events has been investigated by the *aware-*

ness and *notification systems* research domains. In the remainder of this chapter, we present these three research areas and we discuss their relevancy and input towards a system that could provide mobile service awareness.

2.3 Keeping Systems Aware: Context-aware Computing Background

Context-aware computing has the goal of providing tailored support to mobile users by taking into account the dynamic context of use. The term has been interchangeably used with ubiquitous computing, which was introduced as a concept by Mark Weiser in 1991 in his paper on ubiquitous computing [Weiser, 1991]. Since then, a variety of definitions, conceptual approaches and system implementations have been developed. We will first present some of the most notable systems of the literature and then aggregate and compare them along three factors: the elements of context they consider, the level of explicitness in user input, and the level of automation in their reaction in environmental changes.

2.3.1 Context-Aware Systems

The first widely acknowledged prototypes of location-aware computing is the "active badge location system" developed at the Olivetti Research Lab [Want et al., 1992]. A custom-built mobile device was utilised to track office workers' locations within a building and the system would forward incoming calls to the fixed telephone device nearest to the intended recipient. Also prominent in the history of context-aware systems, the "Tabs, Pads and Boards" were developed by the Xerox PARC group in 1993 [Weiser, 1993], which utilised multiple devices of different sizes to capture and present colleagues' locations. Later, "Active Map" also had a similar focus but disseminated the users' location on mobile hosts rather on fixed communal screens [Schilit & Theimer, 1994].

More recently, many location-aware research prototypes and some commercial systems have been developed. Many of them are mobile guides that focus around the tourist experience by giving relevant information according to the current location of the user [e.g. Davies et al., 2001; for a comparison survey see Kray & Baus, 2001]. Other systems focus on indoor tourist activities such as proximity to museum artefacts [e.g. Aoki et al., 2002; Luyten & Coninx, 2004]. On the commercial side, services such as "TomTom Navigator"[1] (or any other driving navigation application), "Find and Seek"[2] by Vodafone (or any other service providing points of interest around the user) are location-aware and are often branded as Location-Based Services (LBS). A more recent example is Google Latitude[3], which presents the location of one's friends (or more precisely the location of their mobile phones) on a map (on PC or mobile).

---

[1] http://www.tomtom.com (last accessed 30/06/2009)

[2] http://www.3g.co.uk/PR/Jan2005/8953.htm (last accessed 30/06/2009)

[3] www.google.com/latitude (last accessed 30/06/2009)

Besides just location-awareness, there are a number of context-aware prototypes that take into account more elements of the users' context (e.g. social or computation environment, time, user preferences). This is expected to increase the value and usability of a system by constructing a better understanding of the users' tasks and objectives [Baldauf et al., 2006]. For example:

- The "GUIDE" system utilises user preferences apart from location to deliver a tailored tourist guide experience [Davies et al., 2001];
- "Squeeze me, hold me, tilt me" makes use of pressure and tilt sensors to imitate naturalistic gesture navigation within a virtual book or sequential lists [Harrison et al., 1998];
- "SmartRestaurant" takes account of customers' current time and location to match their arrival to a restaurant with the preparation of their food order [Lukkari et al., 2004];
- "Ubiquitous Multimedia Information Delivering Service (U-MIDS)" delivers the appropriate multimedia information, such as music, news, and internet radios, based on user, location, schedule and preferences [Hsu et al., 2007];
- "StartleCam" is a wearable system that uses skin conductivity sensors in order to "mimic the wearer's own selective memory response" and record pictures near events that cause involuntary startle responses [Healey and Picard, 1998];
- "SenSay" [Siewiorek, 2003] and "TEA" [Gellersen, 2002] utilise sensors such as accelerometer, light sensor and microphone on the user or the devise respectively, to infer user or device state and adjust the type of the mobile notifications accordingly;
- "Context-aware mobile communication in hospitals" is a handheld system that extends the instant messaging paradigm by adding location, time, recipient's role, and artefacts' location and state to the conditions for the delivery of the message [Munoz et al., 2003];
- "CybreMinder" [Dey & Abowd, 2000] and "EventManager" [McCarthy and Anagnost, 2000] are context-aware tools that allow users to create conditional reminders (and adjust their delivery) combining several contextual elements, such as time, location, users' co-location, or other complex conditions;
- "Everywhere messaging" is demonstrated through a series of research projects, which prioritise messages (e.g. emails or voice mail) and deliver them to the most appropriate medium, location, time and modality (to minimise interruption) by monitoring several elements of context such as calendar entries, to-do lists, sent mail, outgoing calls, usage history, ambient noise and conversational sounds [Schmandt, 1999].

2.3.2 Classifications of Context-aware Systems

Different approaches to context and context-awareness are taken within the plethora of systems. First, they differ not only in which parts of the context they take into account, but also in what actually they consider as context. The most basic form of context is the user location in absolute (e.g. GPS coordinates) or relative terms (e.g. distance from ‘x’). Others have included computational environment, absolute time, history of interactions or other elements of the user identity, such as personal or interaction preferences. Some researchers have argued that the user activities while interacting to the context-aware system (including this interaction), are themselves part of the context. Table 2-1 classifies the context elements and demonstrates the diversity in context focus between 16 established research groups in the area (adapted from [Kaenampornpan, 2009]). It also depicts some of the labelling inconsistencies amongst common context elements.

The second major difference amongst context-aware applications in the literature is that there are different levels of user involvement in the context input. When the context and its changes are detected by sensors that inform the system automatically, ‘implicit input’ is used. On the other extreme, when the context needs to be directly input by the user, ‘explicit input’ paradigm is in place. Most context-aware systems will have a level of implicit input but may also require some explicit input from users. The perceived explicitness in context input can be shaped not just by the amount of user input, but also by how much this interaction is seen as context input. A hybrid approach between implicit and explicit context input is when users input information to the system as part of their interaction, but this information is also captured to inform the system of any context changes. For example, when a user buys a product online, this information can be used to update users’ interests in products and present it accordingly (e.g. “users who have bought this product also have bought this one”).

	Location	Conditions	Infrastructure (Computing Environment)	Information on User	Social	User Activity	Time	Device Characteristics
1	Physical Environment			Cultural Context				
2	Physical Environment		Infrastructure	User Environment				
3	Physical Environment			Human Factor			Time	
4	User Environment	Physical Environment	Infrastructure	User Environment				
5		Physical Environment		Information on User				Device Characteristics
6	Location		Infrastructure		Social	User Activity		Device Characteristics
7	Physical Environment		Information Context					Device Characteristics
8	Location			Identity		User Activity	Time	Identity
9	Location	Environment		User		Subset of Information on User	Time	Device
10	Location			Identity			Time	Identity
11	Location		Quality of Service	Identity			Time	Identity
12	Where		What	Who	What	How	When	What
13	Location		Tools	People			Time	Tools
14	Physical		Infrastructure	Domain/System				System
15	Active/Passive							

Table 2-1: Context classification systems (adapted from [Kaenampornpan, 2009])

Third, there are different approaches in the literature with regard to technology adaptation, intervention or interactivity once the context has changed. The most basic differentiation is ‘active’ vs. ‘passive’ context-awareness [e.g. Chen & Kotz, 2000]. The former instructs an application to adapt to the discovered context as needed by changing its behaviour. One simple example is the automatic change of mobile profiles (e.g. “silent” or “outdoors”) or ringer volume when moving between different contexts (e.g. from attending a meeting to driving) [e.g. Gellersen, 2002; Schmidt et al., 1999; Siewiorek, 2003]. Passive context-awareness, on the other hand, presents context information and changes to users so that they can decide how to make best use of it. For example, navigation or tourist guide systems may provide with a user’s current location and points of interest around her on the map. Another example is “Context-Call” [Schmidt, 2000], which retrieves the current status of other mobile users (which is manually entered) before they decide to call. More advanced systems such as “Connecto” [Barkhuus, 2008] continuously share the location and the profile of pre-selected friends. In these examples, the relevant context is displayed to the users for them to decide the appropriate action. The granularity of this information ranges from sensor raw data (e.g. GPS coordinates, temperature readings) to more meaningful interpretation of such data (e.g. “home” or “very cold”). In some cases, the system can suggest an adaptation (or a number of adaptation options) and allow users to approve or decide when (and/or which) adaptation should take place to support their activity in the new context. This can be described as somewhat hybrid context-awareness between active and passive. For example, the GUIDE system was upgraded from a push to a hybrid interaction paradigm [Davies et al., 2001]. When new (tourist) information was available, a little icon appeared and it was left to the user to choose if or when to retrieve this information and replace the previous text.

Communicating context to users and allowing them to make sense of it is also one of the guidelines suggested for “context sensitive computing” [Brown & Randell, 2004]. The main benefit of passive over active context-awareness is that it avoids misunderstandings between a system’s current state and users’ perception of the system’s state. It also avoids confusion of why system decisions were made, and minimises disruptions caused when the user is already engaged with the system. However, a disadvantage of passive context-aware systems is that they may become too demanding in terms of user attention and frequency of interruptions (depending on how obtrusively they present the context changes).

There are a number of different classifications of context-aware computing that present more specific types than the crude distinction between passive and active types. Table 2-2 is a comprehensive meta-classification of different context-awareness found in the literature (adapted from [Kaenampornpan, 2009]).

In summary, context aware systems create different interaction paradigms by the elements of context they monitor, how context is input in the system (explicit, implicit or hybrid), and the type of action the system takes to support users (active, passive or hybrid). However, orthogonal to the two dimensions of input and output of context is the design decision of how the adaptation or the context information is presented to

Types of Context Aware Computing				
Automatic contextual reconfiguration	Contextual-trigger actions	Proximate selection application	Contextual information and commands	
Contextual resource discovery	Contextual adaptation	Contextual sensing		Contextual augmentation
Automatic execution of a service		Presentation of information and services to a user		Tagging of context to information for later retrieval
Active		Passive		

Table 2-2: Meta-classification of context-aware computing [adapted from Kaenampornpan, 2009]

users. For example, an application adaptation might take place (i.e. active context awareness) with or without a mechanism to inform the user about this change. Similarly, the presentation of context (i.e. passive context awareness) can be done subtly or in a more interruptive way, depending on factors such as urgency and importance. The integration of this 'notification dimension' in context-aware systems creates the idea of context-aware notification systems that can form different interaction paradigms. For example, if we assume three crude levels for each one of the three dimensions of context-aware notifications systems, twenty seven different paradigms can be designed (although it is more useful to see each one of these dimensions as a continuum). Designers are therefore able to create the appropriate level of user involvement (for both input and output), depending on the specific requirements of the system. The three dimensions of context-aware notification systems (and how their different values form distinct interaction paradigms) are graphically summarised in Table 2-3.

Next, we explore the continuum of the notification dimension, and how user interfaces have been utilised to keep users aware of systems' states and events.

	INPUT	OUTPUT	
		NOTIFICATION	FINAL OUTPUT
Interaction Paradigm	Implicit/Explicit/Hybrid	Interruptive/Subtle/Absent	Active/Passive/Hybrid

Table 2-3: Three dimensions of context-aware notification systems (with three crude values for each dimension).

2.4 Keeping Humans Aware: Awareness Background

Context-aware systems can play a central role in providing mobile service awareness by inferring which services are relevant to users' activities. However, the mechanism by which users are made aware of these services is also crucial. These kinds of mechanisms and their implications are explored in 'awareness' literature. Being 'aware' is broadly defined as "having or showing realization, perception, or knowledge" [1] of a situation or fact. However, 'awareness' has been used in a "bewildering" and diverse number of ways in academia [Robertson, 2002]. In fact, Schmidt has gathered an extensive list of studies in Computer Supported Collaborative Work (CSCW) that specify different types of awareness, such as "general", "collaboration", "peripheral", "background", "passive", "reciprocal", "mutual" and "workspace" [Schmidt, 2002]. He points out that this plethora of adjectives arose from the ambiguity of the term 'awareness' and the need to qualify it in a meaningful and useful way.

Domains that awareness appears in the HCI literature are:

- Awareness of tasks secondary to the primary focus of attention, such as email, calendar entries etc. [e.g. van Dantzich, 2002], or ongoing computing processes [e.g. Weiser & Brown, 1995].
- Awareness of peers' presence in the virtual or physical space [e.g. Gaver, 1991].
- Awareness of potential dangerous incidents in vehicle collision avoidance systems [e.g. Graham, 1999; Patterson, 1990], power plant control rooms [e.g. Hickling, 1994] and aircraft cockpits [e.g. Hourizi & Johnson, 2004].
- Awareness of the presence of relevant and/or interesting objects, services or people while on the move [for a review see Filho, 2002].

Schmidt [Schmidt, 2002] classified systems that support awareness within the domain of CSCW in two major categories: those that support the awareness within specific tasks, so that "actors tacitly and seamlessly align and integrate their distributed and yet interdependent activities" [e.g. Gutwin & Greenberg, 2002]; and those that support the social context and the informal interaction of dispersed colleagues, friends, family or lovers, inside or outside the workplace [e.g. Bly et al., 1993; Dourish & Bly, 1992; Gaver, 2002]. These two types of awareness are somewhat equivalent to "micro-level awareness" and "macro-level awareness" respectively, as defined by [Vertegaal et al., 1997]. The first type deals with aspects of a virtual meeting in synchronous interactive systems, while the second type focuses on the context outside the meeting. This second category of awareness is often referred to as "peripheral awareness", and is more relevant to the goal of mobile service awareness.

2.4.1 Peripheral Awareness

Peripheral awareness can be defined as the ability to "maintain and constantly update a sense of our social and physical context" [Pedersen & Sokoler, 1997]. Systems that support it convey information about "the context of activity", while "awareness is peripheral insofar as it concerns activities that are not foreground tasks" [Gaver, 2002]. Others relate it to a "preattentive process" and "subliminal perception and intuitive conduct" [Pedersen & Sokoler, 1997]. However, these references refer more to the internal human cognitive processes, rather the way information is presented to users.

[1] Merriam-Webster Online: http://www.merriam-webster.com/dictionary/aware (last accessed 30/06/2009)

In other words, we consider preattentive processes to relate to peripheral attention rather than peripheral awareness. This distinction is further supported by Treisman's remark on visual processing: "preattentive processing is done quickly, effortlessly and in parallel without any attention being focused on the display" [Treisman, 1985].

The tradeoff between raising awareness and causing unwanted disruptions or distractions is established in the literature [e.g. Gutwin & Greenberg, 1995; Hudson & Smith, 1996b]. Some have chosen to address this tradeoff by abstracting the information that provides peripheral awareness [Gaver, 2002; Hudson & Smith, 1996b; Pedersen & Sokoler, 1997; Weiser & Brown, 1995; Wisneski et al., 1998], while others have opted or recommended richer or more meaningful information representations [e.g. Alexanderson, 2004; Bly et al., 1993; van Dantzich, 2002]. These two approaches have respectively been described as "lightweight" and "heavyweight" methods of obtaining peripheral awareness [De Guzman et al., 2004]. The contradiction of those approaches is better understood when we consider the level and type of attention they require in order to deliver peripheral awareness. Systems that constantly present information to the periphery of the attention of users tend to have more lightweight representations, as very rich information (such as speech) would cause large and constant distraction from the primary task. For example, the "Dangling String" [Weiser & Brown, 1995] is a plastic string that was placed hanging at a workplace hallway and was designed to rotate at a speed relative to the network's traffic. The authors summarise its benefits to common screen displays as such: "While screen displays of traffic are common, their symbols require interpretation and attention, and do not peripheralize well" [Weiser & Brown, 1995]. Auditory examples of lightweight awareness include the continuous "group pulse" sound denoting colleagues presence and activity [Mynatt et al., 1998]; and the ambient "flowing water" sound denoting one's message centre activity [Sawhney & Schmandt, 2000]. On the other hand, systems that occasionally require users' more focused attention tend to present the information in more heavyweight fashion so that they minimise the time spent to understand them. For example, glanceable displays such as "Scope" [van Dantzich, 2002] are designed to inform about secondary tasks by presenting rich information to users. In this case, the disruption is kept to the minimum by occasional glances to the secondary display, which are kept short due to its meaningfulness. An auditory equivalent is VoiceCues, which utilise snippets of someone's voice to denote there is an incoming email from them [Sawhney & Schmandt, 2000].

In short, peripheral awareness refers to the goal of providing contextual information different from the users' primary tasks. The richness or abstraction of this information depends on the frequency of the attention shift and the level of attention required. However, the presentation of information can be more flexible and adjustable both in terms of richness and in terms of periphery of attention. For example, "calm technology" is described to have the ability "to move easily from the periphery of our attention, to the center, and back" [Weiser & Brown, 1995]. Similarly, Schmidt claims that there is no categorical distinction between obtrusive interruptions and peripheral awareness, as people "regulate their monitoring quite delicately so as to adjust the degree of obtrusiveness to the requirements of the situation" [Schmidt, 2002]. Understanding these elements of peripheral awareness is critical for achieving our goal as stated in our research question: "How can we design for non-intrusive yet informative mobile service notifications".

2.4.2 Situation Awareness

Apart from the types of awareness discussed in CSCW research, there is a third type that usually refers to an operator's understanding of the state of a complex and/or safety critical system. It is commonly referred to as "situation awareness" and can simply be defined as "knowing what is going on so you can figure out what to do" [Adam, 1993]. A more elaborated definition often used in the literature is: "situation awareness is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [Endsley, 1995]. According to this definition, there are three steps that are required for a person to achieve situation awareness (or an operator to achieve system awareness). First, they need to perceive the elements of the environment that are indicative of the situation. Second, they need to understand them in relation to their relevant goals and form a holistic mental picture of the current situation, incorporating all the elements and their significance. The third and final step is to be able to predict what is likely to happen next, and therefore be "ahead" of the current situation. We later adapt these steps to address mobile service awareness (Section 2.6)

In contrast to peripheral awareness, achieving situation awareness can be a demanding process and "involves significant perceptual and cognitive resources" [Adams & Pew, 1990]. The demand on the higher level of cognitive resources (e.g. behaviour prediction) may vary according to the domain. For example, it is expected to be particularly high in decision-making processes in dynamic, complex, safety critical environments, such as power plant control rooms and aircraft cockpits. In fact, it has been found that operators' performance is directly related to the quality or completeness of their situation awareness [e.g. Entin et al., 1995; Nullmeyer et al., 2005]. Furthermore, task familiarity is found to affect the level of automaticity of execution, with more novel tasks requiring more effort [Rasmussen, 1986]. This is further supported by the finding that the relationship between working memory and situation awareness diminishes with task practice [Gonzalez & Wimisberg, 2007].

2.4.3 Attention

Regardless of the complexity of a system, perception and comprehension of the relevant environmental elements are related more to users' attention capacity and the nature of the stimuli. The terms 'attention' and 'consciousness' are as much controversial than the term 'awareness', and have been vibrantly debated in the literature. This thesis will follow Johnston and Heinz's definition of attention: "the systematic admission of perceptual data into consciousness" [Johnston & Heinz, 1978]. We also agree that it "implies withdrawal from some things in order to deal effectively with others" [James, 1890], and it can be either willingly or instinctively directed [Sodnik et al., 2008]. Several models ('bottleneck' and 'capacity models') have been proposed to explain the process of selection of the stimuli people analyse semantically [e.g. Deutsch & Deutsch, 1963; Treisman, 1964]. Although their details and differences are outside the scope of this thesis, the most establish theories seem to converge in two points. First, there is some mechanism of attenuating the incoming stimuli based on their physical characteristics and the relevance they have to the given situation. Second, there is some level of pre-conscious cognition (i.e. semantic analysis of unattended stimuli) that can affect conscious processes [Dixon, 1981]. Although the explanation of the results can be debatable, Velmans reported many studies where participants had galvanic skin reactions to the (unattended) meaning or sound of words previously associated with electric shocks, even though they did not recall hearing these words [Velmans, 1999]. Furthermore, Moray demonstrated that even when a strong enough attentional barrier is set by a cognitive demanding task, the subject's own name was the only stimulus to that was attended to [Moray, 1959]. This can also be seen as an example of the cocktail party effect [Cherry, 1953], the ability of listeners to focus on a single speaker amongst multiple conversations and/or noise.

In terms of HCI, users can build awareness of a system and its state by choosing to turn their attention to (or parts of) the interface. Alternatively, they can build peripheral awareness by perceiving the relevant environmental elements from ambient interfaces in the periphery of their attention, and in some cases with some level of pre-conscious cognition. Finally, they may receive explicit interruptive notifications when a system changes states, which will exclusively demand their attention or trigger the cocktail party effect. The speed and level of comprehension of these notifications depend on the inherent meaning of the stimuli in relation to the cognitive capacities of the user (including memory and past experiences).

In the next section we present and discuss literature on notification systems, as they constitute one of the mechanisms that important information (and ultimately awareness) can be conveyed to the users (in a more or less interruptive way). Then we present our suggestion towards mobile service awareness, which combines elements from peripheral awareness, attention and notification systems literature.

2.5 Notification Systems Background

Notification systems can be defined as "interfaces specifically designed to support user access to additional digital information from sources secondary to current activities" [McCrickard & Chewar, 2003]. This information needs to be relevant and important, and to be delivered to users in "an efficient and effective manner without unwanted distraction to ongoing tasks" [McCrickard & Chewar, 2003]. In other words, they need to be able to support attention allocation between a (primary) task or activity and extra incoming information concerning a different (secondary) task or activity. Notification systems can therefore be utilised to support peripheral or situation awareness in a variety of diverse applications, such as email notifications on glanceable displays, social interaction via augmented physical objects or warning signals in operating theatres and aircraft cockpits.

Despite this diversity in domains of application, according to McCrickard & Chewar's framework [McCrickard & Chewar, 2003], all notification systems share three common critical parameters that can describe user notification goals: interruption, reaction and comprehension. These parameters can be manipulated (and empirically measured) to drive design decisions to address the particular requirements of different notification systems. 'Interruption' is caused by an event that requires from users to partially or fully shift their attention from the primary task to the notification. This event can vary in subtleness and persistence in order to match properties of the secondary task, such as urgency. 'Reaction' is the immediate response of users to the interruptive event. This reaction can be guided by manipulating the inherent semantics of the stimuli the notification system presents (e.g. red colour notifications for more urgent need of rapid response). 'Comprehension' refers (beyond the systems' ability to guide to the appropriate immediate reaction) to the ability to help users understand the deeper underlying and perhaps complex reasons the notification is invoked. This supports users to utilise this contextual information to assess their long-term strategies with regard to the system operation. These three critical parameters (interruption, reaction, comprehension) should be appropriately balanced according to the particular goals of each notification system [McCrickard & Chewar, 2003].

2.5.1 Interruptibility

The decisions of when or how much to interrupt people from their daily activities is very challenging to address. The basis of this challenge lays on the fact that even humans are not perfect in inferring interruptibility of their peers. For example, novice designers of notification systems identified appropriate interruption levels with a 17% deviation from expert designers (although expert estimation was not validated or tested in real life scenarios) [Chewar et al., 2004a]. In another study, it was found that people were able to distinguish "highly non-interruptible situations from other situations with accuracy of 76.9%" [Fogarty, 2005]. Similarly, Kern and Schiele showed that this inconsistency seems to extend in interruptibility estimation both of a particular user or her environment, as only about 65% of the cases they tested were estimated correctly [Kern & Schiele, 2006]. Since it seems we are lacking the understanding and the model to infer the degree to which an interruption would be welcome, context-aware systems are unlikely to be able to achieve rates better than humans. In fact, in the same study, a relatively simple system (recording only ambient sound, time, telephone and keyboard/mouse usage) was able to infer interruptibility exactly as well as humans did.

If our best efforts (as humans or through context-aware systems) to infer interruptibility are not always successful, we need to design for damage limitation when estimations are wrong. In fact, there is plenty of evidence that people are very intolerant of errors that lead to unwanted interruptions. For example, annoying or intrusive notifications are rejected and deactivated altogether whether it is about email [Serenko, 2006] or alarms in the operating theatre [Block, 1999]. In a study by Schiaffino, it was found that 93% of participants felt angry when interrupted with assistance irrelevant to the situation [Schiaffino, 2003]. However, 76% of them did not mind to be interrupted if the information presented to them was "important and relevant to them". Hudson et al. conclude that "attitudes toward interruption are marked by a complex tension between wanting to avoid interruption and appreciating its usefulness" [Hudson et al., 2002].

In terms of mobile service awareness, the dimensions of interruption and reaction are the most important. The scenario used in [McCrickard & Chewar's, 2003] for highlighting the usefulness of a carefully balanced interruption/reaction notification system is that of a businessman on the move, who relies on calendar and email alerts. They conclude that for such a system, "redirecting activity to the right place, at the right time is the only important consideration". However, designing a mobile notification system to achieve the ideal level of interruption is particularly challenging. Put simply, "deciding whether to ring or not is a very difficult problem" and "[such attempts] are unlikely to be successful" [Brown & Randell, 2004]. This apparent failure is not only routed on the technological challenges but on the highly conflicting user requirements too. Such a phone should avoid overwhelming users with unnecessary interruptions, while keeping them aware of important calls or services. For this, the right balance between incoming information urgency and (estimated) user interruptibility is needed.

2.5.2 Attentive User Interfaces

"Attentive user interfaces" are notification systems that extend their functionality to sense and infer users' attentive states and priorities to provide peripheral awareness. "They structure their communication such that the limited resource of user attention is allocated optimally across the user's tasks" [Vertegaal, 2003]. They are usually referred to as "attention-centric systems" [McCrickard & Chewar, 2005] when they adapt the information presentation and/or the modality of delivery. Also, when they attempt to infer urgency and relevance of presented information, they have been called "attentional user interfaces" [Vertegaal, 2003]. Some representative examples are:

- "AudioAura" [Mynatt et al., 1998] provides serendipitous information via background auditory cues, depending on the identity and approximate location of users within a building.
- "Nomadic Radio" [Sawhney & Schmandt, 2000] is more context-sensitive as it provides audio notifications depending on inferred importance of the incoming information (such as email, news broadcasts, calendar events), whether the user is engaged in a conversation, and on recent history usage of the device. The system can also manipulate the level of obtrusiveness of the notifications by presenting them in a scalable way, from ambient sound to speech generation.
- "SenSay" [Siewiorek et al., 2003] estimates user interruptibility based on conversation, ambient noise and light, calendar entries and movement. According to the level of interruptibility (and device location in relation to the user), it adjusts the ringer volume, the operation of vibration, or gives call-back suggestions.
- "CaBN" [Wahid 2006] utilises 5 priority levels in its alerts, and provides the appropriate level by comparing the priorities of incoming information (such as emails and calendar events) to the priority of primary user tasks. It chooses amongst (or a combination of) audio and visual modalities.

- "Notification Platform" [Horvitz et al., 2003] is a similar but more sophisticated attention-sensitive system. It calculates the value and urgency of incoming information (from multiple sources) against the probability and the cost of disruption, based on the current location and attention state of the user (utilising a Bayesian attention model). As a result of this analysis it decides the modality (two levels of audio and/or visual) and the device (e.g. PDA, voicemail, PC) to deliver the notification, based on the fidelity, the likelihood the notification will be attended to, and user implicit and/or explicit preferences. The components of the Notification Platform are depicted in Figure 2-1.

Some of these notification systems take into consideration the appropriateness of different modalities or intensity of notification, based on contextual elements. Therefore, they seem to fit the high-level architecture we have proposed for mobile service awareness (Figure 1-2). However, the adaptation of the notifications is done according to the users' availability rather than the nature or the relevancy of the suggested service. The efficiency or annoyance of notifications is not only affected by the users' interruptibility and overall state, but also from the nature of the stimuli presented. If the stimuli are easy to interpret, interruptions could be avoided when unnecessary, but attended to when relevant. One such limited example presents the Nomadic Radio, where a notification for a "short email message sounds like a splash while a two-minute audio news summary is heard as faster flowing water while being downloaded" [Sawhney & Schmandt, 2000]. However, this decision appears to rely on the researchers' personal preference, without any explicit interest in an objective investigation of the meaningfulness, learnability or otherwise appropriateness of the sounds chosen for the particular services.

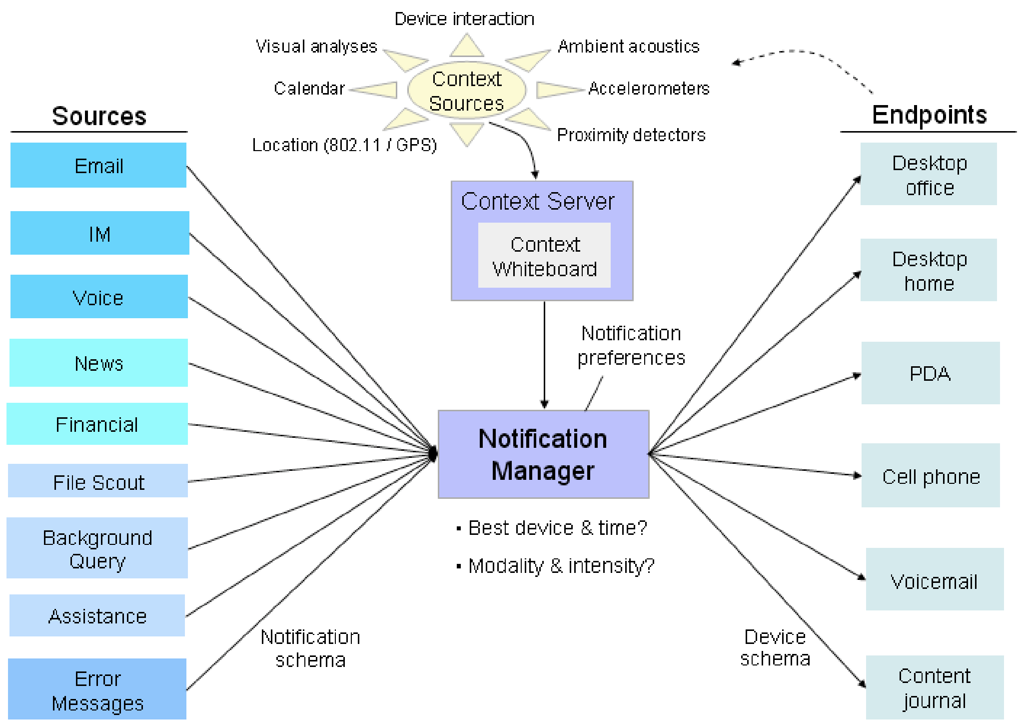


Figure 2-1: Constellation of components of the Notification Platform, depicting the subscription architecture [Horvitz et al. 2003]

2.6 Notifications for Mobile Service Awareness

In this thesis, we argue that mobile service awareness can be achieved by providing users with knowledge of availability of relevant services and the ability for instant initiation. First, a context engine (similar to those in the context-aware systems presented in Section 2.3) would be responsible for selecting the services that are relevant to the users' situation every time. Then, a notification engine would be responsible to select the appropriate information presentation to make users aware of the availability of the service (similar to peripheral awareness systems). The attentive interfaces presented in Section 2.5.2 could further enhance this system by tailoring the notification according to users' interruptibility or location. Finally, the notification could contain the appropriate virtual information that is needed for quick and easy initiation of the service.

The literature of context-aware computing and attentive interfaces has many research prototype context engines to demonstrate, which would be reasonably successful in achieving the estimation of relevancy for mobile services. For example, in the simplified scenario of technology adaptation where the mobile phone is changing profiles according to the environment, the success rate has been 87% [Gellersen 2002]. Once a service has been suggested, a virtual link on the mobile device and the presentation of the relevant information could easily address the instant initiation requirement. However, the appropriate presentation of information is more difficult to address.

Since services may become available at any time or place, there is a requirement for smooth output mechanisms that keep users aware about the relevant services. If notifications are interruptive and cause frustration, it can lead users to attentive exclusion or deactivation of the notification systems [e.g. Block, 1999; Serenko, 2006]. Therefore, it seems that peripheral awareness offers a more desirable paradigm for mobile service awareness. However, mobile device capabilities are much more restricted than the technological means that have been utilised in the peripheral awareness systems discussed. They are mainly designed to deliver interruptive notifications. The absence of peripheral mechanisms constitutes the challenge of notification encoding (enriched or abstracted) even more challenging.

We suggest that this challenge of information presentation is addressed by notifications which are intuitive[1] and/or easy to learn. More specifically, we define intuitiveness of service notifications as the likelihood (above some error threshold) of a naïve user correctly inferring their interactional meaning (without having encountered them previously). This approach combines the best of heavyweight and lightweight peripheral awareness approaches (as described in Section 2.4.1). As notifications will appear occasionally (rather than as constant information in the periphery of the users' attention), the shift of attention away from the primary task will be infrequent and therefore disruptions will be kept to minimum (as in heavyweight approach). In addi-

[1] Merriam-Webster Online definition: "readily learned or understood" http://www.merriam-webster.com/dictionary/intuitively (last accessed 30/06/2009)

tion, this shift would be short-lived and effortless, due to the intuitiveness of the information presented (as in lightweight approach). Furthermore, meaningful notifications could compensate for false positives of the context engine. If users can intuitively recognise which (irrelevant) services are (incorrectly) suggested, then they will not have to shift their focus away from their activities (e.g. driving, conversing or working) to the mobile device. Thus, by reducing the level of interruption, we reduce the level of annoyance. This is consistent with the context-aware guideline, according to which, context sensitive systems should be used defensively, where incorrect behaviour is tolerable and easy to fix [Brown & Randell 2004]. Therefore, our research question evolves from "How can we increase service awareness of contextually available mobile services?" to "How can we design for non-intrusive yet informative mobile service notifications?"

Our starting point in answering this question comes from the situation awareness literature. If we adapt the steps of Endsley's definition of situation awareness [Endsley, 1995] (see Section 2.4.2), users will have to go through the following steps in order to gain awareness of the available mobile services:

1. Perceive the information available to them in the surrounding environment or the mobile interface.
2. Understand this information so that they build a mental picture of which services are available.
3. Relate those services to their current goals.
4. Predict how and when the service availability scene may change.

The notification mechanism we propose would address the first step. The active exploration for the discovery of available mobile services is inherently difficult, mainly due to the interaction limitations (see "issues with on-device mobile service discovery" – Appendix II.1). On the other hand, inferring availability based purely on information from peripheral attention entails the risk of not perceiving the relevant environmental elements, or imposing dominating cues in inappropriate environments (see "issues with in-environment mobile service discovery" – Appendix II.2). However, presentation of information through a notification has better chances of diverting mobile users' attention enough to inform them of the relevant and available services.

The fact that we propose meaningful notifications (or ones that can be easily and quickly learned) is addressing the second step. The more comprehensible the notifications are (or become through usage), the more preattentive or pre-conscious processing may occur (similar to the goals of peripheral awareness systems, and the automaticity occurring in situation awareness systems). The investigation of the meaning and learnability of auditory notifications is the focus of the auditory interface research community, and is discussed in Chapter 3.

Step 3 should be particularly easy for users if the context engine selects the correct service and timing to present the notification (and assuming they are familiar with the functionality of the service). However, even when it fails and suggests irrelevant services, users should be able to understand the contents of the suggested services from the notification stimulus and filter it out with minimal cognitive effort (and hence avert the unnecessary interruption and attention shift).

Finally, step 4 is outside the immediate scope of this thesis. However, it is likely that users will develop an understanding of the typical conditions that cause service availability changes, if they are timely and unobtrusively informed every time such changes do occur.

2.7 Chapter Summary

In this chapter we introduced the concept of mobile service awareness as an alternative perspective to mobile service discovery with the goal of improving mobile service usage. We defined it as "the knowledge (any user has at any given moment) of the available relevant mobile services and how to initiate them". The system we proposed to provide mobile service awareness is a context-aware notification system, which needs to meet two specific requirements.

First, it needs to select the services that are truly relevant in supporting users' everyday activities. Although this is not in the immediate focus of this thesis, we presented relevant background work of context-aware computing. A series of context-aware systems and classifications of context and context-awareness were presented. The understanding gained from this literature review asserts us that the requirement of our context engine can be met with adequate success.

The second requirement is to select the "right" way of presenting the service availability information to users. Drawing from the literature review of awareness and notification systems, we argued that the best presentation of this information is meaningful, or easy to learn, notifications. Similar to the approach of peripheral awareness but adapted for the mobile phone capabilities, easy to comprehend notifications will cause minimum disruption as attentional shifts away from the primary task will be infrequent, short and effortless. This should also mitigate for the occasional errors of the context engine, which would otherwise demand from users to attend to the irrelevant notifications. Therefore, our research focus for the rest of the thesis is around the question: "How can we design for non-intrusive yet informative mobile service notifications?"

Chapter 3

Auditory Notifications

In order to address our new research question ("How can we design for non-intrusive yet informative mobile service notifications?"), we first need to consider the available notification modalities. Notification stimuli could be perceived through any of the five senses but so far only three of them have systematically been utilised as modalities in human computer interaction: sight, hearing and touch. Some could argue that multimodal interaction (utilising two or three modalities) would be the most natural option. It is true that we are used to make sense of the world by constantly using a combination of our senses. For example, seeing and hearing an object falling on the ground is giving us information about its weight and the damage it might have suffered. Only hearing the sound could leave us wondering what caused it, seeing only the fall gives reduced information on its properties, such as weight and material. However, when designers combine stimuli from different modalities in less natural ways, there is a risk that the extra information presented may overwhelm or confuse users rather than add to the semantic richness.

Furthermore, certain modalities might be inappropriate or unavailable in certain contexts. Therefore, a multimodal notification (where modalities are used complementary) will be appropriate or complete in fewer contexts than each individual modality (since all modalities will be required to convey the full meaning of the notification). For example, while in the cinema it is culturally expected that auditory notifications of mobile phone will be switched off. In fact, a cinema company in Dublin went as far as installing a signal blocker to stop customers from receiving any calls or messages for the duration of the movie, which was met with mixed responses from the public[1]. (The somewhat extreme approach of the signal blocker was eventually banned as it was found illegal by Ireland's communications regulator). There is clearly a demand for social conformity with regard to ringtone nuisance in the cinema, but there is also a need for anytime/anywhere communication (e.g. for urgent calls). A more viable and acceptable solution could be based on the appropriate (and possibly automatic) choice of notification modality. A discreet vibration in the cinema can inform the user of that important incoming text message, without disturbing the other patrons.

Each of three available modalities can be utilised in a number of ways to convey information to users, and in particular notifications to mobile users. For example, the visual channel can be utilised to present text, images, or video on the mobile display and/or on displays that might be present in the users' environment. Furthermore, simpler forms of information can be encoded in the flashing patterns or colours of lights (e.g. LEDs). For example, a blue flashing light on a mobile phone is usually associi-

[1] http://news.bbc.co.uk/1/hi/entertainment/film/2991451.stm (last accessed 30/06/2009)

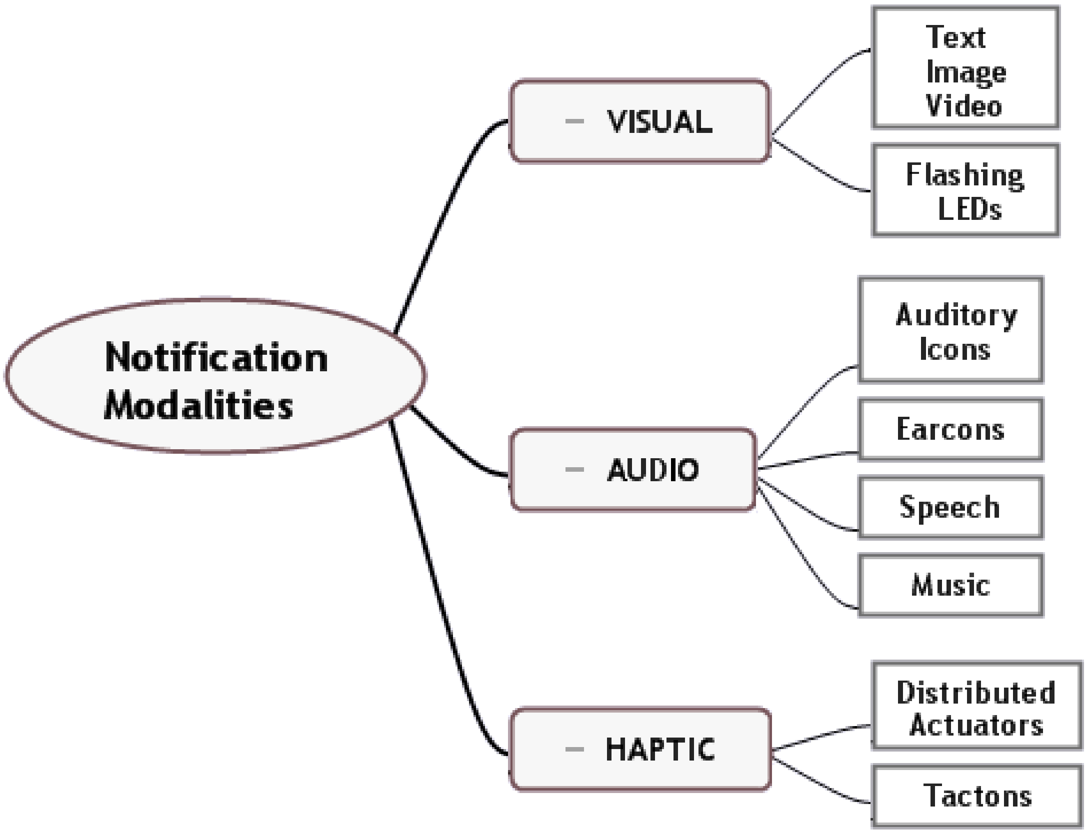


Figure 3-1: The available notification modalities for mobile services and examples of their potential use

ated with Bluetooth functionality, while different flashing patterns of a red coloured light could be notifying about missed calls or low battery warnings. The auditory channel can be used to transmit different types of sounds that also vary in the richness of the information they can convey. For example, speech (synthesised or pre-recorded) is very rich and can be as specific as needed, while sounds like Auditory Icons and Earcons are less informative (these types of sound are discussed in detail in Section 3.2). Finally, the haptic channel can be utilised to encode some information in the vibration patterns (known as "Tactons" [Brewster & Brown, 2004]) and/or frequencies by actuators on the mobile phone or distributed on a user's body through wearable computing. The three modalities and some examples of the way they can be utilised are graphically depicted in Figure 3-1.

3.1 Scoping Notifications to the Audio Modality

Designing multimodal notifications is out of the scope of this thesis. Before one systematically investigates the combinations and compatibility of signals from multiple modalities, more research is needed in the construction of meaningful stimuli for each modality. Besides, this research can contribute in crossmodal interaction, as there is early evidence indicating that knowledge of information coded in auditory and haptic modalities is highly transferable from one to the other [Hoggan & Brewster, 2007]. We have chosen to focus on the auditory interface, which many researchers have pointed out as more appropriate or measured higher user preference in mobile contexts [Brewster, 2002; Fitch & Kramer, 1994; Hudson & Smith, 1996a; Sawhney & Schmandt, 2000; Sodnik et al., 2007]. In particular, we opted to focus on audio mobile service notifications based on four main reasons: users' dependency on vision while on the move; the increased flexibility of auditory interfaces in semantic richness; human ability to habituate or attend to multiple audio streams; position of mobile devices in relation to their owners. We briefly discuss each one of these factors below.

First, users heavily depend on their vision while on the move and therefore have less visual attention to spend on the mobile interface. Looking at a small display can seriously increase cognitive load and reduce performance of the primary task, such as navigation as pedestrians [e.g. Brewster, 2002] or drivers [Sodnik et al., 2007]. Besides, in studies comparing visual and auditory displays in attention demanding or cognitive demanding environments, participants performed and preferred better the auditory displays [Fitch & Kramer, 1994; Sodnik et al., 2007; Tzelgov et al., 1987]. Therefore, meaningful audio service notifications are promising in terms of performance and preference, without engaging the users' vision or hands.

Second, auditory interfaces are flexible in the amount of information they can convey over time. Ranging from arbitrary sounds, to loose or strong metaphors and speech, the semantic richness can vary according to the needs of each user, application and/or context of use. Since there is a trade-off between raising awareness and causing disruptions [Gutwin & Greenberg, 1995; Hudson & Smith, 1996b], this flexibility can be utilised and be fine-tuned by abstracting or enriching the auditory stimuli accordingly (see Section 2.4.1 for peripheral awareness literature review). In addition, such adjustments in the richness of the auditory semantics can also address the trade-off between awareness and privacy, as users might want to regulate the amount/type of personal information shared [Hudson & Smith, 1996b]. On the contrary, non-textual visual notifications (e.g. flashing patterns) and tactile notifications are limited in the amount of information that they can convey. These modalities may be fairly successful with small sets of notifications (e.g. users were able to learn to recognise with 90% accuracy three mobile services from three vibration patterns [Brown & Brewster, 2005]), but they are difficult to scale up. On the other hand, a large number of studies have demonstrated the effectiveness of different audio types as notifications, even with much larger sets of notifications (discussed in Section 3.2).

Third, the semantic flexibility of the auditory interface can be complemented and coupled with the auditory pre-attentive abilities that humans seem to have (as demonstrated in a number of studies summarised by [Velmans, 1999]). Many models of limited attention suggest that incoming stimuli are being attenuating based on their physical and semantic properties [e.g. Deutsch & Deutsch, 1963; Treisman, 1964]. In particular with the auditory channel, it has been observed that listeners can focus on a single speaker amongst multiple conversations or noise (known as the cocktail party effect) [Cherry, 1953], or extract multiple parallel audio streams at the same time [Fitch & Kramer, 1994]. This ability to cognitively ignore or habituate uninteresting or irrelevant sounds better than visual stimuli could be explained by our anatomical inability to block out these sounds [Fitch & Kramer, 1994]. Since our goal is to provide mobile notifications that can disappear in the background when irrelevant or not needed, the auditory channel seems most appropriate.

Fourth, mobile devices are often out of sight as they are put in pockets or bags, are left in different rooms, or obstructed by other objects. Chipchase et al. found that mobile essentials (mainly mobile phones, keys and cash) achieve the maximum distance from their owners while within their home, where security is not a problem [Chipchase et al., 2005]. Similar are the results from Patel et al., who conclude that users overestimate the proximity and availability of their mobile phones, when in reality it was found that only 50% of the time the phones were at arm's reach (and turned on) [Patel et al., 2006]. Both these studies report that when outdoors, people usually keep their devices closer to themselves. More specifically, another survey of 419 participants found that the most popular location for mobile phones when outdoors was the front pocket of trousers (mainly for men), followed closely by shoulder bags (mainly women) [Ichikawa et al., 2005]. Together they accounted for 67% of the participants, who would not perceive any visual notifications unless they made the effort to reach for the device (if notified by another modality). Not surprisingly, only 46% of the participants who kept their mobile phone in a bag felt that notifications were successful, while for pocket users this number reached 70%. Although vibration notifications are expected to be easier to feel when the mobile is in the pocket than in the bag, no significant difference was found between vibration and non-vibration users' perception on notification success [Ichikawa et al., 2005].

Therefore, there are two dimensions to the user context, which should inform the notification design and modality. The first is the user's context (e.g. location and activity), which determines their level of interruptibility (and that of others around them). The second is the location of the mobile phone in relation to the user (e.g. in the bag), which determines which modalities can be perceived. When the device is out of sight, touch or audible range, the corresponding modalities will provide ineffective notifications. Table 3-1 demonstrates how these two dimensions of context affect the appropriateness of modality. If we assume four crude categories for user context at any time, they can be described in terms of the four profiles appearing in most modern mobile phones: 'Outdoors' (flashing screen, loud noise and vibration), e.g. 'watching a football game' or 'socialising in a pub or a club'; 'General' (flashing screen, normal volume noise and vibration on), e.g. 'working in the office', 'having a chat over lunch' or 'watching TV at home'; 'Meeting' (flashing screen, low volume noise and vibration on), e.g. 'attending a lecture or a meeting' or 'watching a play or a movie at the cinema'; 'Pager' (flashing screen, no noise and vibration on), in any context where no sound is accepted. Similarly, if we assume four crude positions of the mobile phone in relation to the user, they could be clustered in: 'in a pocket' or 'in a bag' (the two most popular locations found by [Ichikawa et al., 2005] when users are outdoors), or 'within arm's reach' (e.g. on a desk near-by) or 'in another room'. The latter two locations are similar to 'within the same room (within 5-6 meters)' and 'unavailable (beyond 6 meters)', as identified by [Patel et al., 2006].

Phone Location / **User Context**	*In Pocket*			*In Bag*			*Arm's Reach*			*Other Room*		
	V	**A**	**T**	**V**	**A**	**T**	**V**	**A**	**T**	**V**	**A**	**T**
Outdoors	X	✓	✓	X	✓	X	✓	✓	X	X	✓	X
General	X	✓	✓	X	✓	X	✓	✓	✓	X	✓	X
Meeting	X	✓	✓	X	✓	X	✓	✓	✓	X	✓	X
Pager	X	X	✓	X	X	X	✓	X	✓	X	X	X

Table 3-1: Effectiveness of notification modalities in different examples of user context and phone location (V: Visual, A: Auditory, T: Tactile, ✓: Perceivable, X: Unperceivable)

One of the interesting observations from Table 3-1 is that the visual modality is rarely available as a notification. It can be informative and used in any context if the user attends to the device, but can only attract attention in very few contexts. On the contrary, the auditory channel can be utilised as a notification in most contexts outside the Pager profile. Whether the phone is in another room, in the pocket or in a bag, audio notifications are likely to be perceived and understood without the need to take the eyes or the hands away from concurrent activities. In some of these situations the effectiveness of the audio modality will depend on the level of ambient noise, the appropriateness of sound interruptions, the particular stimuli utilised and their amplitude. Similarly, the phone vibration might become audible depending on the ambient noise, the surface the device is resting and its distance from the user. Overall however, the audio stimuli flexibility (in type, meaning and amplitude) makes it more appropriate than any other modality in the most common contexts (in Table 3-1, visual modality is unperceivable in 12 contexts, tactile in 9, and audio in 3).

From these observations and the inherent advantages of sound in stimuli flexibility, pre-attentive properties, and appropriateness in mobile contexts, we conclude that the most efficient mobile notifications are those that utilise the auditory channel. However, there are some limitations and disadvantages when utilising auditory notifications. First, sound might be socially inappropriate in certain contexts and can lead to embarrassment (e.g. while in a lecture, a meeting or at the theatre). Second, repetitive and frequent audio notifications can also cause annoyance to the intended receiver and those around them. Third, depending on the semantic level of audio notifications and the social context they are received in, users' privacy might be compromised. Fourth, in very noisy or auditory demanding environments audio notifications might be not heard or interfere with the current activity. This is tightly interconnected with the temporal property of sound, making it impossible to scan or review once it is presented. It should also be noted that we fully support the concept of user personalisation to accommodate individual differences and preferences regarding interruptibility, presentation modality and stimuli. However, such resource-demanding personalisable multimodal approach is outside the scope of this thesis.

Therefore, we do not claim that audio should be the only channel for mobile service notifications. A careful choice of the modality (or combination of modalities) should be applied depending on the context of the user and the relative position of the mobile device. Research projects that sense these contexts and automatically set the phone profile [e.g. Gellersen, 2002; Schmidt et al., 1999; Siewiorek, 2003] are efforts towards reducing or making interruptions more discreet. However, some argue that the goal does not need to be towards the reduction of interruptions but towards making them more effective [e.g. Hudson et al., 2002]. We are interested to address this challenge by investigating the properties of the notification stimuli. We are interested to investigate how careful design of audio notifications can minimise some of the negative effects associated with interruptions. For example, different types of sounds or different instances of each type may be perceived as more or less annoying, intrusive, compromising privacy or cause varying levels of distraction. Therefore, our research question takes its final form: "How can we design for non-intrusive yet informative auditory mobile service notifications?"

In the next section we present extensive background work on auditory interfaces and in particular how 'auditory icons', 'earcons' and speech have been utilised as auditory notifications. We then aggregate the theoretical and experimental findings of these three sound types and presented their relationships in a novel topology in Section 3.3. Finally, in Section 3.4, we review numerous studies that have measured and compared the effectiveness of different types of audio notifications. The literature review presented here informs directly our studies conducted and presented in the following chapters.

3.2 Sound Types as Auditory Notifications

Many different types of sounds have been used in a variety of commercial and research auditory interfaces. These sounds can range from simple electronic 'beeps' to digital reproduction of complex sounds such as human speech, according to the particular requirements and limitations of the systems they complement. There exist many different and overlapping classifications of them, as they can vary greatly in their attributes and complexity. One way to classify them is according to their relationship to the interface event they signify. This event could for example be the emptying of the recycling bin on a desktop computer [Gaver, 1989], a warning signal for missile launch in an aircraft cockpit [Leung et al., 1997; Patterson & Mayfield, 1990], the presence of users on a network [Cohen, 1994], or the success (or potential success) of hitting a target button on a mobile interface [Brewster, 2002]. The degree to which these sounds are semantically related to the concepts they represent can be expressed as a continuum, ranging from a completely arbitrary mapping (e.g. a 'beep' sound), to a metaphoric mapping (e.g. raising pitch to indicate escalation of an ongoing process), to a highly meaningful mapping (e.g. speech reproduction). A more detailed classification is presented by [Keller & Stevens, 2004], where indirect mappings between signified and signifier can be of ecological or metaphorical nature. The ecological relations are based on their coexistence in a certain environment (e.g. seagull and ship sounds), while the metaphorical relations derive from feature similarities on a physical or a functional level (e.g. helicopter and mosquito sounds). The authors found that direct mappings were the easiest to learn, and indirect mappings were easier to learn than arbitrary mappings. However, there was no significant difference between the learnability of ecological and metaphorical indirect mappings, although the associative strength of the former was stronger [Keller & Stevens, 2004]. It is worth noting that people tend to construct meanings for the sounds that they hear on an auditory interface, even if they were not designed with those relationships in mind [Brewster, 1998]. Sometimes, they even construct stories in order to give meaning to a series of unrelated sounds, such as sounds provided to monitor background activities on a computer network [Cohen, 1994].

The sparse guidelines for auditory notifications seem to lack generalisability as they are limited by the domain they appear in or the specific sound types they utilise. One such attempt was the ISO 9703-2, which was proposed in 1994 but has now been withdrawn: “Anaesthesia and respiratory care alarm signals - Part 2: Auditory alarm signals”. According to Ulfvengren it suggested that auditory alerts should have “maximum transmission of information at the lowest practicable sound pressure level, ease of learning and retention by operators and perceived urgency” [Ulfvengren, 2003]. In different contexts, other dimensions than urgency might be more appropriate (e.g. relevancy).

A methodology for a user-centred design process for auditory warnings has been proposed by Edworthy & Stanton, based on the ISO standard for evaluating public information systems (ISO/DIS 7001:1979) [Edworthy & Stanton, 1995]. This process breaks down suitability of auditory warnings so that stimuli can be modified by collaborative efforts of designers and end-users accordingly. For example, it accounts for appropriateness in terms of “their confusability, their urgency, their relationship to the actual referent and so on” [Edworthy & Stanton, 1995]. Apart from facilitating representative end-user involvement, two of the main foci of this process are to improve the meaning or ‘nature’ of the warnings to the users, and study them as a cohesive set rather than individual stimuli. Figure 3-2 provides a diagrammatic representation of the design procedure, including the process and its output at each stage.

The methodology suggested by Edworthy & Stanton is not restricted to the design of particular types of sound. However, the types of sounds utilised in auditory interfaces vary, and their distinct characteristics may influence the design of notifications. Although the terminology and definitions found in the literature are diverse, two of the most prevailing non-speech sound types are ‘Earcons’ and ‘Auditory Icons’. Although they have been defined and used in slightly different ways, there seems to be a consensus in the literature with regard to the directness of the mappings they use. Earcons tend to fall on the more arbitrary and metaphorical relations, while auditory icons use more direct or ecological relations. In the following subsections, we present a literature review on the properties of earcons, auditory icons and speech notifications.

3.2.1 Earcons

Blattner et al. define earcons as “nonverbal audio messages used in the user-computer interface to provide information to the user about some computer object, operation or interaction”, and also characterise them as “the aural counterparts of icons” [Blattner

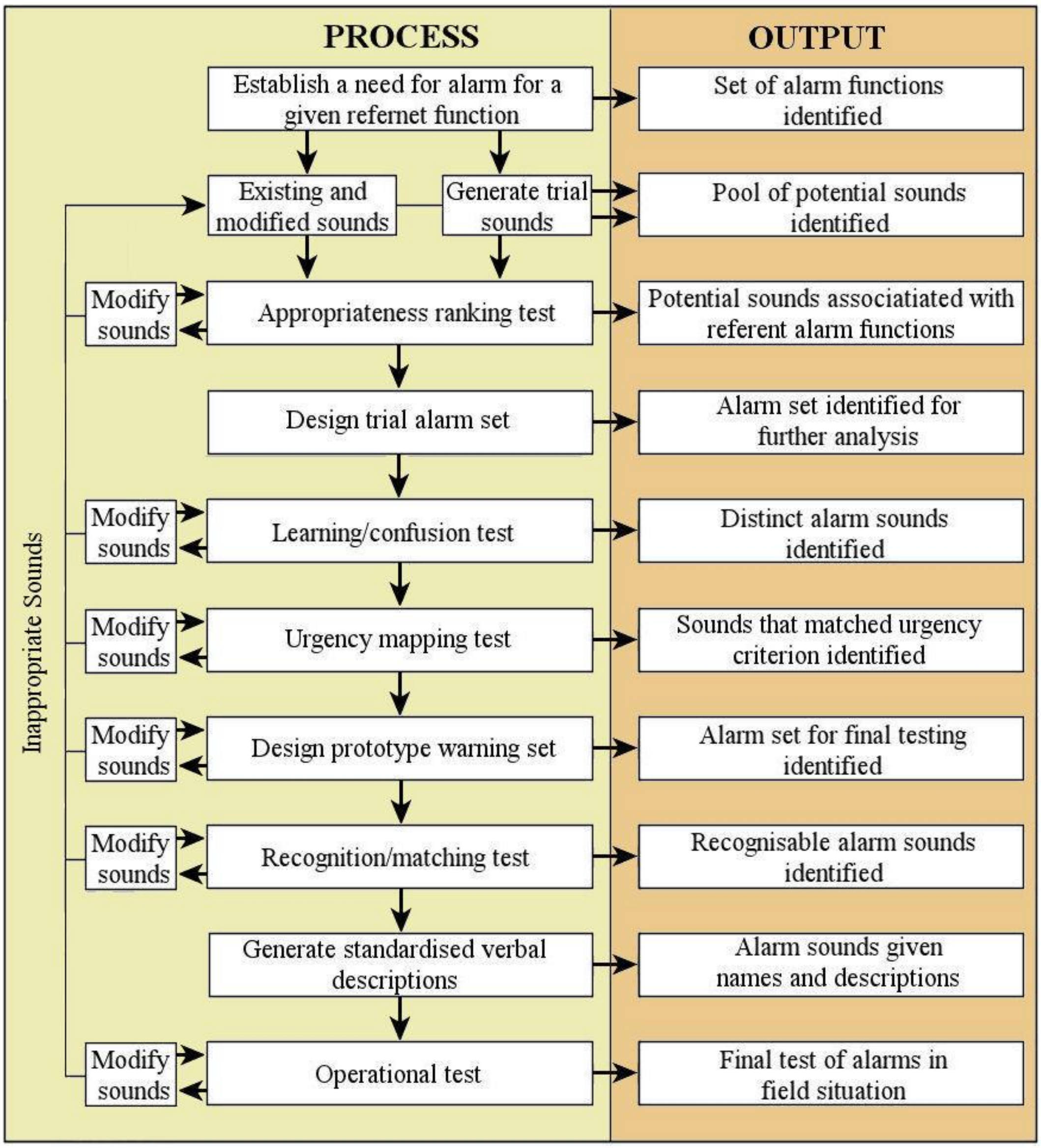


Figure 3-2: Diagrammatic representation of auditory warnings design procedure [Edworthy & Stanton, 1995]

et al., 1989]. Brewster et al. focus on more musical earcons and describe them as "abstract, musical tones that can be used in structured combinations to create auditory messages" and are "composed of short, rhythmic sequences of pitches with variable intensity, timbre and register" [Brewster et al., 1993]. Musical synthetic sounds that are not structured in a particular way have also been utilised in auditory interfaces and referred to as "abstract" [e.g. Leung et al., 1997], "musical" [e.g. Sikora et al., 1995] or "melodic" sounds [e.g. Sanderson et al., 2006], or just "attensons" [e.g. Ulfvengren, 2003b]. Notably, Pirhonen et al. have explicitly expanded the definition of earcons to include all non-speech sounds that do not imitate everyday sounds [Pirhonen et al., 2006].

Flexibility is one of the main advantages of earcons, as they can be tailored to each application or context in which they are used. They can be structured in families, where sounds in each family share at least one property (such as timbre), while sounds within families vary in other properties (such as pitch or rhythm). This way, compounds of individual earcons ('compound earcons') can be used to create a notion of a grammar, and thus deliver more complex messages [Blattner et al., 1989]. For example, similar objects can be represented by one family of earcons, while similar actions that can be performed upon these objects can be represented by a different family [Brewster et al., 1993]. In addition, families of earcons can be created hierarchically ('hierarchical earcons'), where their parameters are changed according to the position within the hierarchy [Blattner et al., 1989]. For example, nodes on the first level of a hierarchy can be differentiated by distinct timbres, which they share with all their children nodes. Also, the deeper one moves within the hierarchy, the lower pitch could be used, while rhythm could be utilised to discriminate amongst sibling nodes (same level and branch).

The major weakness of earcons stems from the absence (or large indirectness) of semantic relationships to their referents. This can be a limitation when designing earcon notifications, as they need to be learned "without benefit of past experience" [Gaver, 1997]. Although there is evidence that creating metaphorical mappings improve their memorability [Brewster, 1998], such metaphorical mappings are constrained by the need to retain their hierarchical or grammatical structure. When no such structure is necessary, there is evidence to suggest that some attributes of sound such as loudness, pitch, tempo and onset can relate to levels of otherwise unrelated concepts such as temperature, pressure, size and rate [Rigas & Alty, 2005; Walker & Kramer, 2005]. However, the rationale and success of these mappings can be too subjective as "data-to-sound mappings that seem intuitive to a sound designer may actually result in less effective performance" [Walker & Kramer, 2005]. Besides, as suggested by the findings of Keller and Stevens, such weaker metaphorical mappings are more difficult to learn [Keller & Stevens, 2004]. Consequently, it is expected that training potential users will be necessary before they are able to successfully use an auditory interface with earcons (earcon learnability and effectiveness are presented in comparative studies in Section 3.2.4).

Designing earcons is not trivial and requires some musical knowledge and experience. This in itself can be seen as a disadvantage when designing an auditory interface as it adds an extra layer of difficulty. Blattner et al. suggest the use of motives for the design of earcons [Blattner et al., 1994]. Motive is defined as "the smallest meaningful unit of musical thought […] - a compact group of notes having a pattern or design sufficiently unique to differentiate it from other groups" [Bernstein, 2005]. Blattner et al. suggest that rhythm and pitch are the most defining parameters for motives, while timbre, register and dynamics should be manipulated to create variances of the same motive. They also comment on the length of an earcon, which should be "sufficient to convey the message effectively, but no longer" [Blattner et al., 1994]. Brewster et al. performed a number of experiments investigating the performance of earcons and derived a set of design guidelines [Brewster et al., 1993]. They generally focus on making sure that there are substantial differences between earcons, as users had difficulty detecting subtle differences. Some of these guidelines are: use musical instruments timbres, especially ones that are easy to tell apart; do not use pitch alone, unless the differences between pitches are very big; make rhythms as different as possible, varying the number of notes in different motives.

The effectiveness of earcons in improving usability of computing systems by conveying extra information has been demonstrated in a number of studies. For example, hierarchical earcons have been found effective for navigating complex menus of mobile phones, as participants completed more tasks successfully with fewer key presses required compared to the silent condition [Brewster, 2002; Leplatre & Brewster, 2000]. Other examples where earcons have been effectively utilised include improving interaction with small soft buttons on mobile interfaces [Brewster, 2002], creating auditory maps [Blattner et al., 1994], and summarising algebra equations for visually impaired users [Stevens et al., 1994].

Furthermore, people seem to be able to learn and recall earcons after adequate, short training sessions. For example, participants were able to identify the correct location of hierarchical earcons in over 80% of the times in a 27-node, 4-level deep file-system hierarchy [Brewster, 1998]. The training procedure used to produce these results was 5 minutes of active learning after a short tutorial explaining the earcon design rationale. Recall rates were equally good even 1 week after the initial training and testing. When compound earcons were used to represent an extended hierarchy of 36 nodes, although earcons were lengthier and more likely to cause increased cognitive effort to learn, recall rates rose at 97%. A smaller number of earcons is even easier to learn. After similar training to the one mentioned above, participants were able to learn 18 earcons with over 90% accuracy in just 2 sessions [Hoggan & Brewster, 2007]. However, in at least 2 separate studies, ability to discriminate amongst earcon rhythms was difficult (and significantly harder than discriminating amongst timbres) [Brewster et al., 1993; Pramana & Leung, 1999], and iterative design was necessary to improve them [Brewster et al., 1993]. Also, others have found that using only pitch of tones or melodies to denote depth in a hierarchy does not significantly improve depth awareness (in comparison to silent conditions), and may also cause confusion to users [Barfield et al., 1991; Sodnik et al., 2008]. Simpler tones have not been as effective as earcons, even when they are combined to construct "melodic alarms" according to literature's standards and guidelines [Block et al., 2000]. For example, 16 melodic alarms were successfully learned by less than 30% of participants, even after two 35-minute long training sessions (on 2 separate days) [Sanderson et al., 2006b]. In the same experiment, even providing mnemonic hints (singing syllables of alarm functions to melody) did not result in any significant difference in learning. In a different experiment with the same sounds and mnemonic tips, participants needed over 43 minutes on average (median 35 min, range 19 to 135 min) to learn the alarm associations at 70% accuracy [Williams & Beatty, 2005].

Some other interesting findings with regard to earcon learnability are the effects of training methods, users' musical ability, and sound quality. For example, explanation of the structure of earcons has a significant positive effect on learnability [Lucas, 1994; Pramana & Leung, 1999], personal tutorial is more effective than textual training, and presentation of the sounds for familiarisation does not add more than active learning (e.g. click and hear) [Brewster, 1998]. On the other hand, there are contradictive data on the effect of participants' musical ability on musical sounds' learnability. For example, participants who knew basic music notations and/or had the ability to play a musical instrument have been found to learn earcons significantly easier (regardless of training) [Pramana & Leung, 1999]. The same was observed for melodic medical equipment alarms (combinations of simple tones) for participants who had at least 1 year of formal musical training [Sanderson et al., 2006b]. In a different study however, the advantage of musically trained participants (ability to read music and play an instrument) was present only for sounds made of simple timbres or sounds without rhythm information, but not for typical earcons [Brewster et al., 1993]. Finally, lowering sound quality (from CD quality to telephone quality) had a negative effect on learnability, but this was counterbalanced by careful redesigning of earcons' rhythms [Brewster, 1998].

3.2.2 Auditory Icons

Auditory icons are sounds designed using the concept of what is described as 'everyday listening'. Everyday listening is "the experience of hearing sounds in terms of their sources", so that "instead of mapping information to sounds, we can map information to events" [Gaver, 1997]. Auditory icons are conceptually more similar to most graphical icons than earcons, as they utilise a metaphor that relates them to their referents. Gaver suggests that "if a good mapping between a source of sound and a source of data can be found, the meaning of an auditory icon should be easily learned and remembered" [Gaver, 1986]. Some of the metaphors used in auditory icons include 'shattering dishes' for dropping an object into the recycle bin [Gaver, 1989], 'door slamming' for remote users logging out of a network [Cohen, 1994], and 'tyre-skidding' for collision warning while operating a driving simulator [Graham, 1999].

The main advantage of auditory icons is that (because of the metaphors they utilise) they are easy to recognise or learn and remember, and therefore training requirements can be kept to minimum. Furthermore, they can easily convey more complex and multi-dimensional information. 'Parameterised auditory icons' are everyday sounds (or synthetic sounds representing them) that can be manipulated along dimensions relevant to the events that produced them, such as weight or material of objects. This way, they can produce families of related sounds. For example, Gaver's SonicFinder application [Gaver, 1989] includes associations of different materials (e.g. wood or metal) for different virtual objects (e.g. files or applications), and distinct sounds of the material manipulations (e.g. knocking or scraping) are used when different actions are performed on the virtual objects (e.g. selecting or dragging). Further mapping can be applied, such as variation of pitch or register to denote variation on size of objects. Parameterised auditory icons are more flexible in auditory interface design, but care must be taken not to become overcomplicated and thus party lose their intuitiveness, or perceived by the listener "to be 'gimmicky' or semi-serious" [Graham, 1999]. The advantage of constructing such sounds is reflected on the level of flexibility they gain in representing a grammar, similar to compound earcons. For example, 4 auditory icons manipulated in parameters such as volume, reverberation and pitch, were able to be recognised for their (iconic) meaning along with 3 more dimensions: size, distance and direction of movement [Stevens at al., 2004]. Although the ease of recognition significantly dropped only after the fourth parameter manipulation was applied, there is evidence that "without a period of substantial training, recall of an icon's identity as well as its three acoustic parameters is difficult and demanding" [Stevens et al., 2004]. It is worth noting that parameterised auditory icons have also been referred as "representational earcons" [Blattner et al., 1989], and were initially introduced by Gaver as "caricatures of naturally occurring sounds" that "don't really sound like the objects they represent, but that capture their essential features" [Gaver, 1986].

On the other hand, the major disadvantage of auditory icons is that virtual events and actions do not always lend themselves to everyday sound metaphors. Therefore, the consistency of the interface will have to be compromised (by using other types of sounds for some events), or far-stretched metaphors will have to be utilised. The subjectivity of the metaphors utilised is in itself a design challenge, as poor metaphors can be misleading and therefore more problematic in learnability and memorability than arbitrary associations. Furthermore, depending on the context of use, it is possible to confuse auditory icons with actual environmental sounds [Cohen, 1994].

Auditory icons can be designed either by recording everyday sounds or constructing synthetic sounds that imitate the everyday sounds. Constructing the sounds is necessary for creating parameterised auditory icons and requires some musical knowledge and experience (as with earcons). Whether recorded or composed, auditory icons' usability is affected by at least 4 factors: identifiability (how easily they are recognised), conceptual mapping (how successful the mapping metaphor is), physical parameters (such as length, frequency range and sound quality) and user preference (e.g. user emotional reaction) [Mynatt, 1994]. Furthermore, and similarly to earcons, as a set they need to be sufficiently distinct from each other. Based on these factors, Mynatt suggests a methodology for creating auditory icons:

1. Choose short sounds which have a wide bandwidth, and where length, intensity, and sound quality are roughly equal.
2. Evaluate the identifiability of the auditory cues using free-form answers.
3. Evaluate the learnability of the auditory cues which are not readily identified.
4. Test possible conceptual mappings for the auditory cues using a repeated measures design where the independent variable is the concept which the cue will represent.
5. Evaluate possible sets of auditory icons for potential problems with masking, discriminability and conflicting mappings.
6. Conduct usability experiments with interfaces using the derived auditory icons.

Therefore, two important steps towards designing effective auditory icons are ease of sound-source identification and intuitive signal-referent mapping (although the link between these two recognition stages becomes automatic through learning [Keller & Stevens, 2004]). Although we are not aware of studies in auditory interface research that test the identification of the auditory icons they utilise, we can build on the results of a number of studies that have systematically and empirically evaluated ease of identification of environmental sounds. They are usually designed to provide sets of sounds that can be utilised in medical, neuropsychological or cognitive sciences. For example, more light could be shed to determine if Alzheimer's or brain-damaged people have the same difficulty of naming sounds as healthy individuals, or explore the generalisability of theoretical frameworks describing semantic long-term memory [Marcell et al., 2000]. In one such study, identification accuracy of short environmental sounds was found to correlate to response times, and self-reported identification difficulty [Ballas, 1994]. Accuracy was also correlated to causal uncertainty, "a function of the logarithm of the number of alternatives that were given as causes for the sound". These results suggest that "causal uncertainty [and hence ease of identification] can be calculated easily from identification responses". These responses were found to be related to "the ease with which a mental picture is formed of the sound, context independence, the familiarity of the sound, the similarity of the sound to a mental stereotype, the ease in using words to describe the sound, and the clarity of the sound" [Ballas, 1994]. In a study investigating "sound events" (i.e. "sequences of closely grouped and temporally related environmental sounds that tell a story or establish a sense of place") identification accuracy (81%) also correlated highly with confidence ratings, response times and sound familiarity. [Marcell et al., 2007]

The process of measuring sound identification is in itself a highly subjective process. It involves choosing the environmental sounds, and judging which of the participants' free-form responses correctly describe the sounds in question. Perhaps, this explains the diversity in absolute accuracy results in the literature. In one study only 15% of 64 short everyday sounds were described at over 80% accuracy by the 83 participants, while the majority of the sounds were given partial or alternative descriptions [Mynatt, 1994]. However, in another study, over 50% of 120 sound events were described at over 90% accuracy by the 25 participants, 33% of the sounds ranged between 50% and 89% accuracy, and only 15% scored accuracy below 50% [Marcell et al., 2007]. Notably, in the latter study alternative, unanticipated descriptions that were "subsequently judged as an acoustically precise alternative interpretation by the unanimous judgments of three independent listeners" were counted as correct responses.

The second step for designing successful auditory icons is to choose appropriate conceptual mappings between the sounds and the interface actions or events they represent. It has been suggested that representative end-users should be involved in interpretation process [e.g. Graham, 1999]. However, this approach is rare in the literature of auditory display and associations are usually based on the researchers' intuition (and therefore are more subjective). To counterbalance this, Pirhonen et al. proposed a designing method for non-speech audio through use scenarios and the involvement of a panel of designers [Pirhonen et al., 2006]. Although their method contributes towards a more analytical approach to sound design and sound ecologies, it is only focusing on producing earcons that are bound to the specific scenarios used. A more detailed design process is an adaptation of the auditory warning design procedure suggested by [Edworthy & Stanton, 1995] (Figure 3-2), with particular focus on "natural warning sounds" (i.e. sounds that occur naturally to a specific environment of operation, and therefore require little or no training by the specific operators) (Figure 3-3) [Ulfvengren, 2003]. The author focuses particularly on end-user participatory conceptual mapping tests, through "soundimagery" and "associability" studies. In soundimagery studies (introduced by [Gardner & Simpson, 1992]), users untrained to the stimuli are asked to match individual sounds to a function and rate their appropriateness. Associability studies test users' ability to learn and retain associations between different sets of sounds and functions. The relative associability success of each set is assumed to derive from their ease to be associated with the functions, thus sounds scoring higher in soundimagery are expected to score high in associability too.

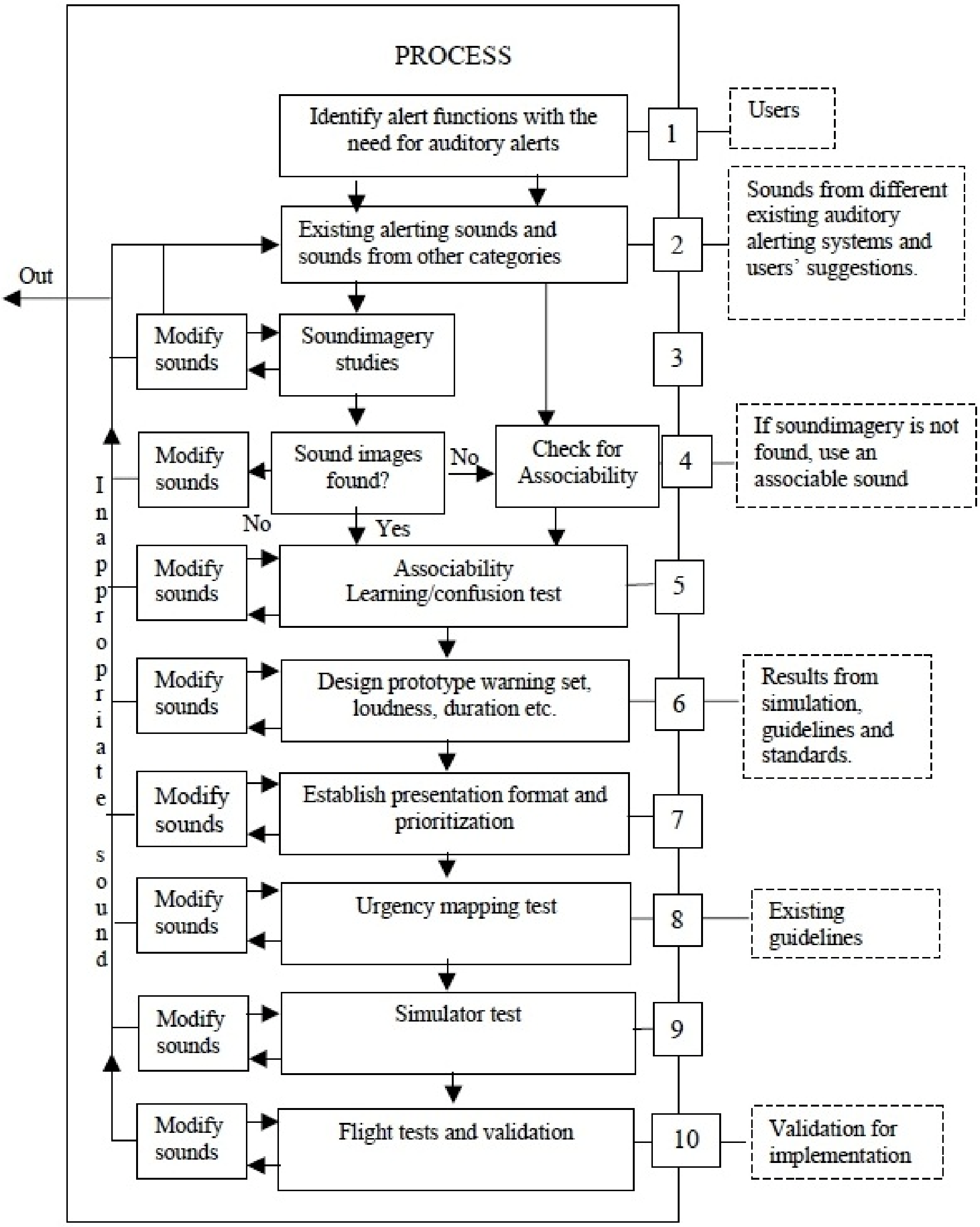


Figure 3-3: A cognitive engineering design process for Natural Warning Sounds and auditory alerting systems [Ulfvengren 2003]

Finally, as with earcons, another factor that affects auditory icon learnability is participants' training. It was found that participants who were informed about the metaphors that auditory icons utilised performed significantly better than those who did not [Lucas, 1994]. Furthermore, giving descriptions or depictions of the auditory stimuli to participants is less effective than allowing them to create their own [Edworthy & Hards, 1999]. Also, it was observed that there is a positive correlation between the range of descriptors used and the learning difficulty [Edworthy & Hards, 1999].

3.2.3 Speech

Speech (in a language known to the user) removes any difficulty in finding direct or metaphorical associations between the sound (signifier) and the system action (signified). Although the mappings between words and meaning are mainly arbitrary in most languages[1], they have been learnt and practised to the extent that their interpretation is usually automatic. This could particularly benefit infrequent, high-priority speech warnings given by computers, but may become annoying for messages that are less important and heard more frequently. Nevertheless, speech notifications can be successfully deployed to give high-level information to several people at the same time - as for example when used in the tannoy systems of airports and train stations. Furthermore, it has been argued that "the human auditory apparatus is focused primarily on distinguishing speech from all other sounds" [Nass & Gong, 2000]. This means that speech auditory warnings have an increased probability to be picked up in a rich and complex acoustic environment.

The main disadvantage of speech notifications is that they are generally not as brief and concise as their non-speech audio counterparts. As icons can present more information (through the use of metaphors) in less space than text, non-speech sounds can take less time to convey more information. Secondly, speech (or text) is not universally understood as it refers only to populations familiar with the particular language, while non-speech audio metaphors can pervade across many cultures. Thirdly, speech notifications can cause interference in speech based communication environments [Edworthy & Stanton, 1995; Patterson, 1982], where non-speech stimuli can be easier to segregate and be attended to. Specifically, Moore suggests that "There is good evidence that speech is a special kind of auditory stimulus and that speech stimuli are perceived and processed in a different way from non-speech stimuli; there is a special "speech mode" of perception" [Moore, 2003, p330]. Fourthly, privacy concerns might make speech the least preferred option for public notifications. One such example can be found in [Baber & Noyes, 1993], where the perceived privacy concerns surrounding a proposed speech-enabled ATM led to it being rejected by its potential users. However, privacy infringement by speech notifications relies on the nature of the information communicated by the notifications. For example, it has been found that speech notifications are preferred for less private information, unlike personal information such as medication reminders [McGee-Lennon et al., 2007].

[1] Sparse and unconfirmed information indicate that the sounds of Sanskrit words represent aspects of the quality of the object they refer to.

Designing speech notifications can be a challenging process as there are many variables that influence auditory interfaces, sometimes in ways that are not yet fully understood (for example, see open questions posed in [Nass & Gong, 2000]). There is evidence that attributes such as gender, emotion tone, personality traits derived by speech, and familiarity of voice affect user preference and performance [Bhatia & McCrickard, 2006; Nass & Gong, 2000]. For example, users found speech notifications of their own voice most interruptive and quickest to react to, while performance was lower for other familiar voices, and even lower for unfamiliar voices [Bhatia & McCrickard, 2006]. In another study, reaction times were slower when synonyms were used instead of the verbatim spoken warnings [Ulfevgren, 2003b]. However, these kinds of design decisions do not affect the comprehension of the notifications, as long as they are audible and in a language (and accent) that people understand. Since the focus of this thesis is to provide mobile service awareness through the comprehension of notifications, details on speech variations will not be discussed further.

3.3 The Meaning of Sounds

In order to better understand the relationships amongst the sound types discussed above, we have created a 2-dimensional topology based on their inherent meaning of sounds and the meaning culturally attributed to them. If we follow the terminology suggested by Gaver [Gaver, 1997], the directness of signal-referent associations is expressed along a spectrum from “symbolic” (completely arbitrary), to “metaphoric” (somewhat meaningful), to “iconic” (self-evident). Along this spectrum, speech would sit towards the symbolic side as there is no readily available evidence or established knowledge of how or why words mean what they do. Earcons would span from symbolic to metaphoric and auditory icons from metaphorical to iconic, depending on the strength of the metaphor used between the sound and its signifier. These relationships in the corresponding sound types are roughly depicted in Figure 3-4.

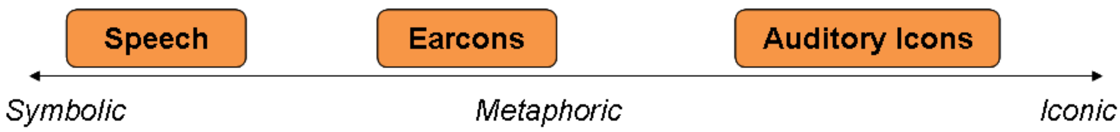


Figure 3-4: Spectrum of nature of signal-referent association and respective sound types

However, orthogonal to this dimension is the meaning that sounds have acquired by usage and learning within a specific culture. Gaver states that symbolic mappings rely on “social convention for their meaning” [Gaver, 1986], but wider cultural elements (such as the social conventions) actually affect the meaning of all sounds along the symbolic-to-iconic spectrum. For example, the English language will be very meaningful to a native English speaker, somewhat meaningful to non-native English speaker (learning the language) and almost meaningless to anyone who have never studied English. On the other extreme, even iconic sounds that use strong metaphors to relate to their function or interface event will be recognised only if people have developed an understanding of these sounds though repeated exposure in their everyday lives. It is therefore suggested that this dimension is unique for each individual, but also dynamic and flexible over time. Ulfvengren describes these sounds as “soundimages", and defines them as sounds that have "a particular meaning to someone without prior training" [Ulfvengren, 2003]. She also points out that soundimages are not restricted to any particular type of sound, and they are dependent on subcultures (e.g. professional pilots) and people's operating contexts (e.g. cockpits). So, if we see the relationships between the different sound types through the lenses of an individual's understanding at a particular snapshot in space and time, it is more likely that it will roughly look like the graph in Figure 3-5. Speech will still be symbolic but meaningful, auditory icons will be generally meaningful because they are constructed from everyday sounds, parameterised auditory icons will be expected to be meaningful only along certain of their dimensions, and earcons will be generally symbolic and meaningless.

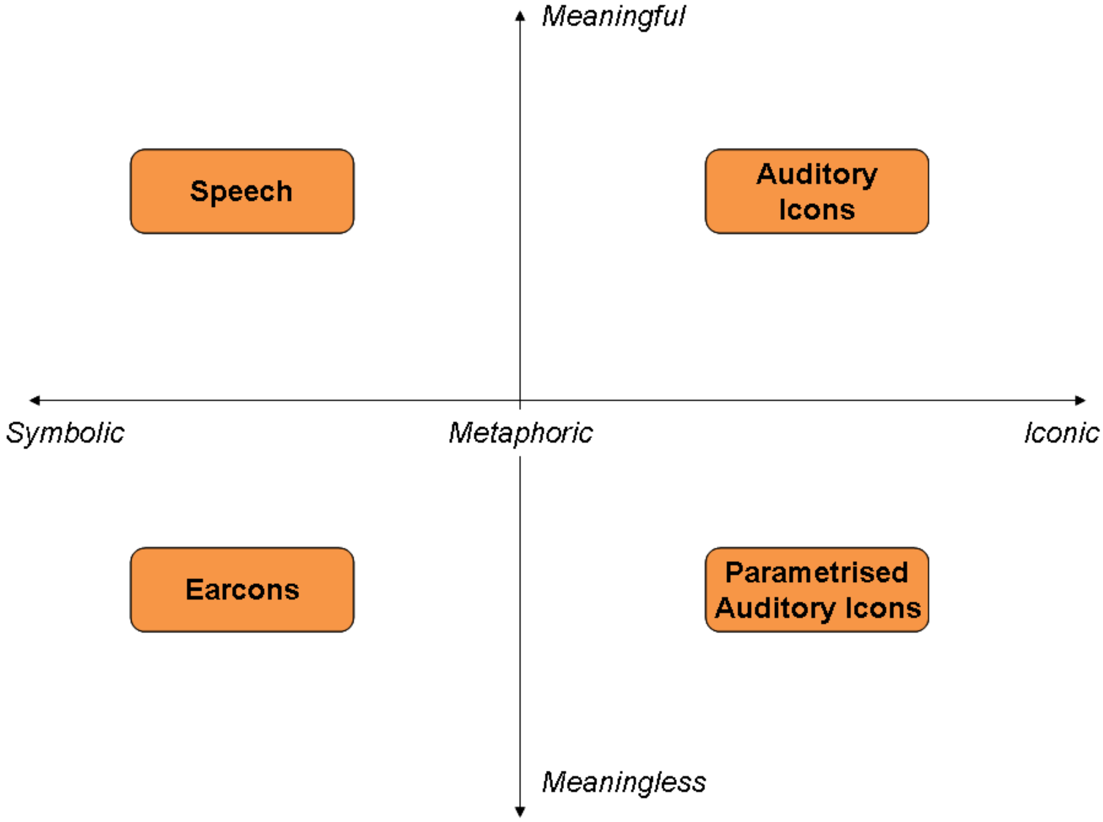


Figure 3-5: Approximate relationship of sounds along their inherent and developed meaning

Of course, this snapshot will be different for particular instances of sounds of each sound type, depending on the amount of exposure and therefore familiarity a person had to these particular sounds. Therefore, the boundaries between earcons and auditory icons are actually much more fluid than presented in the literature and this thesis so far. For example, sounds like the standard Nokia incoming SMS notification were initially assigned in an arbitrary manner but through extended and widespread usage became widely recognisable and established. This earcon should therefore appear higher than most of earcons on the Y-axis of Figure 3-5. As another example, the sound of a siren can be considered as a fairly symbolic auditory icon (and therefore it should appear further left on the X-axis than most auditory icons), which only inherent mapping is between the loudness and pitch of the sound to the urgency of the message conveyed. However, this association is established mainly through cultural usage, from an early age and as implicitly as language is taught to toddlers. Similarly, the Nokia ringtone can be considered as an auditory icon in some cultures, as the association can be learned implicitly. This phenomenon is accurately described by Kramer for continuous sounds whose meaning eventually becomes intuitive: "If the training has brought the user to a place where little or no cognitive effort is required for the link between the data and the sonic representation, so that the representation leads to a "second nature" categorization of the underlying events, we can say that the analogic display is now being used symbolically" [Kramer, 1994a].

Furthermore, a metaphoric auditory icon can be very similar to a metaphoric earcon, depending on the strength of the metaphor that can be used. For example, rhythmic thumps (auditory icon) or rhythmic beeps (earcon) can be used to monitor a patient's heart rate in the operating theatre. The metaphor used is the same but the realism of the sound (similarity to a real heart pumping) differentiates them. One could actually argue that this is an example of an iconic auditory icon and therefore earcons can stretch much further to the right side on the X-axis. On the other hand, auditory icons with more complex metaphors (e.g. shattering dishes for deleting objects) are less iconic, yet not easily represented through an earcon metaphor. Finally, speech can appear more flexible in both dimensions as a foreign language will be less meaningful to an individual and thus appear lower on Y-axis, and words such as "thump" or "boom" are more iconic and should appear further right on the X-axis. These examples, the relationships and similarities of the types of sounds discussed here are presented in Figure 3-6.

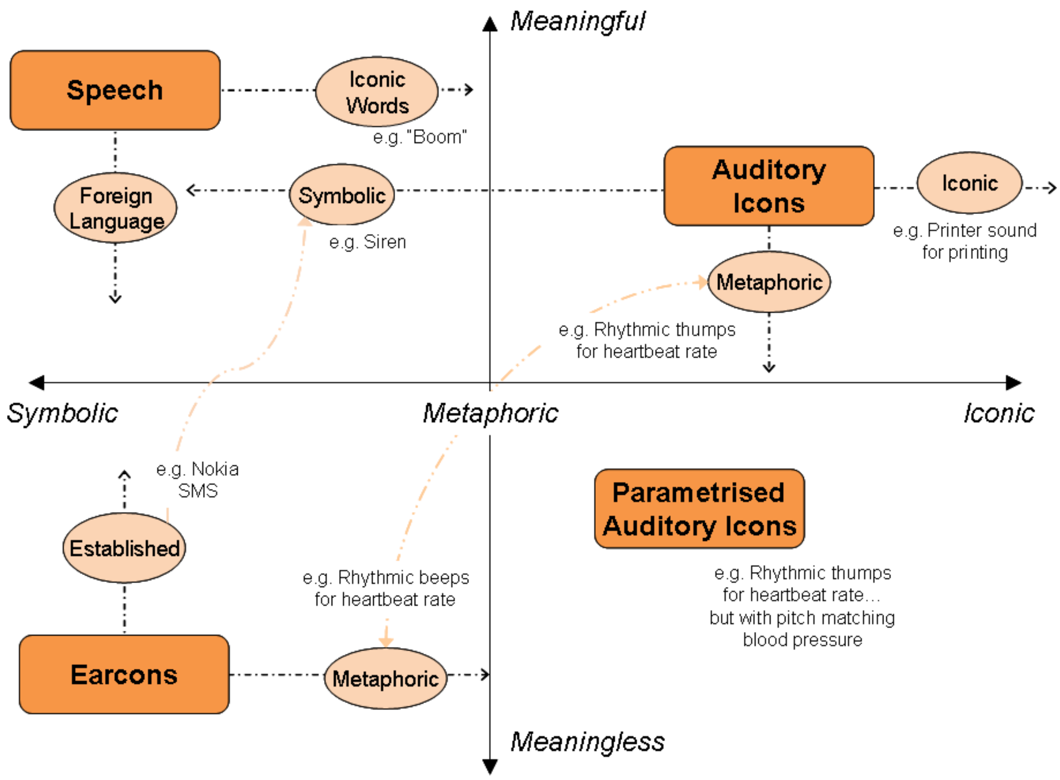


Figure 3-6: Fluidity of sounds' relationships and examples

This fluidity between types of sounds and their meaning, and the impact it has on learnability, is also one of the conclusions of [Edworthy & Hards, 1999]: "…the relationship between a sound and its imagery, whether verbal or otherwise, needs to be considered on a sound-by-sound basis rather than on a type-by-type basis. Thus the ideal method of learning any particular sound is not immediately clarified by knowing which class of sound it might be."

3.4 Effectiveness of Auditory Notifications

The effectiveness of audio notifications or warnings can be measured along several dimensions. It has been argued throughout this thesis that the meaning of the notification is important for the support of mobile service awareness. Since the purpose of a mobile service notification is to unobtrusively inform the user of the availability of services, its meaning needs to be as intuitive (i.e. readily understood) as possible. However, non-speech sounds are less meaningful than speech and the mappings to the events they represent may need to be learned. Therefore, intuitiveness and ease of learning are the first two dimensions of the effectiveness of non-speech audio notifications that we need to focus on. Thirdly, we are interested in measuring users' ability to retain these associations once learned. Memorability is going to be a key aspect of mobile service notifications, especially for services that are not presented frequently to users. Finally, pleasantness and overall user preference over the audio stimuli presented to them is of paramount importance for the success of the notifications. It has often been observed that when notifications are perceived as annoying and generally disruptive, users tend to deactivate them completely [Block et al., 1999; Serenko, 2006].

There are plenty of studies in the auditory interface literature that have partially compared these factors of effectiveness for various sound types as warnings. Some studies are situated in cognitively demanding environments such as power plants [Hickling, 1994], aircraft cockpits [Leung et al., 1997; Patterson & Mayfield, 1990] or motor vehicles [Graham, 1999; Ulfvengren, 2003b], where the accurate and timely interpretation of sound warnings is potentially life-critical. In these contexts, researchers have focused on metrics such as learnability, memorability, reaction times, user preference and perceived urgency. In the context of mobile service notifications however, reaction times are not so relevant, although they could indicate levels of cognitive effort in recognition. The dimension of urgency is essential for warnings prioritisation, but not essential for mobile service notification, and is therefore outside the scope of this thesis. Findings on the metrics of reaction times, cognitive effort, learnability, memorability and user preference are discussed in the following subsections.

3.4.1 Response Times and Cognitive Effort

In studies that speech notifications were utilised, they were found to produce quicker response times than any other sound type [Lucas, 1994], equally quick responses to auditory icons [McKeown, 2005; Ulfvengren, 2003b], or slower responses than auditory icons but quicker than any other type [Graham, 1999]. A possible explanation for these mixed results is that although speech is the most semantically rich medium (and therefore short spoken messages should require less time and cognitive effort to interpret), it is also less concise. This is further supported by the experimental design decision of when to start counting response times in the above studies. In [Lucas, 1994], response times were recorded after the presentation of each cue, thus eliminating the disadvantage of the possibly lengthy speech messages. In [McKeown, 2005] and [Graham, 1999] studies the timer started with the beginning of the presentation of the cues. Furthermore, the [Graham,1999] study compared only 4 sounds with 2 auditory icons (horn and skidding tyres in a driving simulation scenario), which also produced the most false positive reactions. On the other hand, auditory icons seem to produce consistently quicker responses than earcons, musical sounds or simple tones [Bonebright & Nees, 2007; Bussemakers & De Haan, 2000; Graham, 1999; McKeown, 2005; Stevens et al., 2006; Ulfvengren, 2003b]. In one study however, equal response times were produced for auditory icons and earcons [Lucas, 1994]. This exception could be explained by the fact that participants were tested only once for each sound (i.e. each sound was only heard twice including the training phase, and no feedback was given in the testing phase), causing a floor effect for both sound types. Furthermore, in studies where abstract auditory icons (everyday sounds without the application of metaphors to associate them to functions) were used, they were found to produce significantly slower responses than the meaningful auditory icons, but no quicker responses than the simple tones [McKeown, 2005; Ulfvengren, 2003b]. It is therefore likely that the notifications with best response times are the ones which are both most intuitive and shortest in presentation time.

In all of those studies, reaction times were measured as an indication to the cognitive effort required when responding to audio notifications. A more direct measurement of the cognitive effort required to respond to speech, earcon and pager beep notifications was attempted by [McGee-Lennon et al., 2007]. A digit span memorisation task showed no significant difference in the cognitive effort required to react to notifications of the different sound types. However, the option to click on a 'hint' button and read the notification interpretation before responding to it, made the length and effort required to retain the digits uneven between sound types (as the hint function was consistently used only for pager beeps). Also, this design decision makes it impossible to draw any conclusions on notification learnability for the 3 sound types.

3.4.2 Learnability

Learnability is usually measured by the number of trials needed to learn the associations between a set of sounds and the functions or events they represent. In these trials, sounds are displayed in a random fashion (either within categories or all together) and participants are asked to choose the appropriate visual cue from all the functions or labels displayed on the screen (in some exceptions, functions need to be memorised too). Feedback may be given to participants with each response (correct answer given/verified or just correct/incorrect feedback provided) and/or accumulative score at the end of each trial round. In some studies, the overall errors across all trials are also analysed, while in others no feedback is given and only one response per sound is required. The testing trials usually follow an initial training phase, where participants are informed about the rational of the associations. They are usually given the opportunity to hear all sounds and see their visual counterparts and sometimes are given time to explore the associations by click and listen. As described in sections 4.2.1 and 4.2.2, there is evidence that different training methods affect learning and retaining of auditory notifications, for both earcons and auditory icons. Furthermore, the number of associations needed to be learned, training completion criteria and properties of the particular instances of the sound types tested are aspects that can affect the direction and the significance of learnability results in the auditory interface literature.

Despite this diversity of experimental designs and audio types, learnability of auditory notifications seems to be consistent and follows the same trend with reaction times. Speech notifications (and ‘spearcons’, which are speech notifications speeded up to the degree that they are not readily recognised as speech) have been found to be significantly more learnable than auditory icons and earcons [Dingler et al., 2008]. In this particular study however, the experimental design allowed an unlimited (and unrecorded) number of repetition for each sound before responding. Also, some of the auditory icons resembled more abstract auditory icons (where the associations between everyday sounds and their referents are very weak or inexistent). Speech notifications were also easier learned in [Jones & Furner, 1989] (as reported by [Graham, 1999; Lucas, 1994]), [Lucas, 1994] and [Ulfvengren, 2003b], where the number of associations varied from 6 to 20. In some other studies, speech achieved similar learnability to auditory icons [Leung et al., 1997; McKeown, 2005]. The [Leung et al., 1997] study had 8 auditory icons describing events in a military cockpit, which were selected as the most associable for these events by professional aircrew members. This end-user involvement in the design of auditory notifications is rare, and was not applied in the aforementioned studies where speech outperformed auditory icons. The [McKeown, 2005] study did not involve end-users in the design of auditory icons either, but followed a more comprehensive training procedure, with demonstration of the associations followed by 2 training blocks performed with immediate feedback on selection. These observations indicate that auditory icons can perform as well as speech in terms of learnability, if they are associable enough or tested after a couple of training runs.

Furthermore, there seems to be a learnability trend within auditory icons, depending on the directness of the signal-referent relationship. Abstract auditory icons, have been found significantly harder to learn in comparison to auditory icons [Keller & Stevens, 2004; McKeown, 2005; Ulfvengren, 2003b]. Also, nomic auditory icons (where the sound represents the event that caused it) have been found to be more learnable in comparison to indirect auditory icons (where there is a metaphorical or ecological relationship between the sound and the referent) [Keller & Stevens, 2004]. Similarly, auditory icons almost always outperform earcons, musical sounds or simple tones in learnability [Dingler et al., 2008; Edworthy & Hards, 1999; Jones & Furner, 1989 (as reported by [Graham, 1999; Lucas, 1994]), Leung et al., 1997; McKeown, 2005; Stevens et al., 2006; Ulfvengren, 2003b], or self-reported learnability [Bonebright & Nees, 2007]. One exception where auditory icons’ were not significantly better than earcons was in the [Lucas, 1994] study, in which only one response per sound was required and thus not much learning took place. There is also evidence that the trend in which more meaningful sounds demonstrate better learnability continues down the semantic spectrum, with earcons outperforming rhythmic sounds with simple timbres, which in turn outperform simple tones [Brewster et al., 1993].

3.4.3 Memorability

Some of the above studies repeated the tests a number of days later, excluding the training phase, and hence measuring memorability of the associations for the different sound types. Overall, there seems to be a consensus amongst these studies in that all notifications achieved higher or similar accuracy scores and retained the learnability trend amongst them after 2 days [Leung et al., 1997], after 1 week [Leung et al., 1997; Lucas, 1994; McKeown, 2005], and after 2 weeks [Edworthy & Hards, 1999]. Specifically in [Lucas, 1994], the trends in reaction time and accuracy were retained 1 week later, but earcons improved slightly more than auditory icons in reaction times, and auditory icons improved slightly more than earcons in accuracy. In another study, the learnability trend was also retained between two experiments carried out 1 week apart, with the exception of abstract auditory icons, which were slightly forgotten and demonstrated similar performance with abstract tones [Ulfvengren, 2003b]. However, no immediate comparisons can be made as the performance metrics were different between the two experiments (the first measures number of learning trials while the second measures overall errors). In the study by [Brewster et al., 1993], the learnability trend between earcons, rhythmic sounds with simple timbres and abstract tones was completely flattened after 15 to 20 minutes from training (after the distractor task of learning a different set of sounds). This uniformity in performance was the result of an insignificant drop in the performance of earcons.

However, none of the studies above have investigated direct comparisons or reported significant differences between memorability of different sound types such as speech, auditory icons and earcons. Perhaps no such differences were observed because a ceiling effect occurred by a combination of factors, such as the number of associations needed to be learned and the number of days they needed to be retained. It is expected that given sufficient time (or high enough number of associations) the meaning of the notifications should fade in time. An interesting question then would be to investigate which sound type has steeper forgetting curve and which one is retained for longer.

3.4.4 User Preference

Apart from cognitive effort, learnability and memorability, user preference is another important factor that affects auditory notifications' success. Surprisingly, less than one third of the comparison studies mentioned above have focused on user preferences. The ones that have, plus a couple of studies that measured only preference, have focused on metrics such as appropriateness and helpfulness of the notifications, or annoyance and pleasantness of the sounds. Overall, there seems to be more diversity in preference results than in the other factors. Some studies even fail to draw definite conclusions. For example, no significant difference in pleasantness was found between speech, auditory icons, abstract auditory icons and abstract sounds in [McKeown, 2005]. However, the speech was marginally more preferred with no great fluctuation between its instances, while instances of the other sound types were less liked when they were found to be more urgent. Similarly, no significant differences in pleasantness or annoyance were found when comparing speech, earcon and tones notifications in the home environment [McGee-Lennon et al., 2007]. However, speech and earcons were liked equally and marginally more than tones, while speech scored significantly better in helpfulness. Subjective data gathered in this study suggest that the semantic richness of speech notifications in a social context can be perceived as an advantage (since everyone understands it) or a disadvantage (when used for notifications of more private nature).

Of the studies that found consistent differences, earcons or musical sounds seem to be preferred to auditory icons. For example, although auditory icons were found to produce more reliable mappings to communication functions than musical sounds and standard communication tones, auditory icons were rated consistently lower in appropriateness and pleasantness than musical sounds and communications tones [Sikora et al., 1995]. In fact, 11 communication tones, 7 musical sounds and none of the auditory icons were selected as the most preferred sounds for the communication functions. However, there were more musical sounds presented to the participants, and the musical sounds and communication tones were created by a musician while the auditory icons were collected by the researchers, and later reported as of lower quality [Roberts & Sikora, 1997]. These two limitations were remedied in a follow-up study, which compared speech, auditory icon and earcon notifications in a scenario-based simulated communication application [Roberts & Sikora, 1997]. The results showed again significant highest scores in pleasantness and appropriateness for speech notifications, followed by musical sounds and finally auditory icons. It should be noted that both of these studies were targeted at businessmen and tested in a (simulated) business environment. A similar trend in preference (speech to earcons to auditory icons) was found in another study, although auditory icons were easier to learn than earcons [Jones & Furner, 1989] (as reported by [Graham, 1999; Lucas, 1994]).

This trend, however, was not observed in a study comparing 4 notifications (1 speech, 2 auditory icons and 1 simple tone) [Graham, 1999]. It was found that one of the auditory icons (horn) was rated by far as the most appropriate notification for imminent collision. The other auditory icon (tyre-skid) was found less appropriate and was liked less for its poorer quality (it was extracted from a computer game, while the horn was a sound recorded by the authors). Speech was criticised for low urgency and slow speed, and the simple tone was criticised on all counts.

It seems therefore that preference trends between sound types are difficult to establish, and they do not always follow the performance trends observed in learnability or other factors. This diversity and inconclusiveness in preference results can be attributed to a number of reasons: diverse individual differences in perceived pleasantness of sounds in general; contexts of notifications' use and user personas targeted (e.g. businessmen in an office environment or drivers in a collision avoidance system); frequency of notifications (there are unconfirmed reports that natural sounds such as speech and auditory icons become more annoying when repeated frequently); lack of (and inherent difficulty for) standardisation of the processes by which each family of sounds is generated (and hence results being significantly influenced by researchers' subjective evaluations). Furthermore, even when laboratory studies attempt to recreate the natural environment notifications occur, participants are asked to decide within minutes of exposure to sounds to estimate what their preference would be in long term, everyday usage. Perhaps this is another reason why results are inconclusive, in which case it would be interesting to see if there is a clear trend in preference when people are exposed to the sounds longitudinally.

3.4.5 Effectiveness Discussion and Summary

The comparative results in response time, learnability and retention (through trial rounds to criterion or overall accuracy errors), and preference (through appropriateness, helpfulness and pleasantness) are summarised in Table 3-2. Note that in order to maintain the same direction of presentation in all metrics, the response time column has been inverted to reflect response speed. The types of sounds have been grouped and coded as follows:

S: Speech

S_p: Spearcons (speech speeded up)

A: Auditory Icons

A_n: Auditory Icons – nomic (similar to A but sound represents the event that caused it)

A_i: Auditory Icons – indirect (similar to A but signal-referent relationship is indirect either through a weaker metaphor or of ecological relationship)

A_a: Auditory Icons – abstract (everyday sounds with arbitrary signal-referent relationship)

E: Earcons

E_m: Earcons – metaphorical (similar to E but with higher metaphorical signal-referent relationship)

M: Musical Sounds (similar to E but without particular structure as a set)

R: Rhythmic Sounds (similar to M but with simple timbres)

T: Tones (simple beeps or bursts of sounds)

and mathematical comparison operators have been used to demonstrate the relationships between them:

$>$ Significantly better than

$\geq$ Non-significantly better than (or not reported as significant)

$\leq$ Non-significantly worse than (or not reported as significant)

$\approx$ approximately equal performance

Reference	Response Speed	Learnability	Retention	Preference
Bonebright & Nees, 2007	A>E	A>E (self-reported)	--	--
Brewster et al., 1993	--	E>R>T	All Same (≤ E)	--
Bussemakers & De Haan, 2000	A>E	--	--	--
Dingler et al., 2008	--	$S \approx S_p > A > E$	--	--
Edworthy & Hards, 1999	--	$A > E_m > T$	All Improved	--
Graham, 1999	A>S,T	--	--	A>S,T
Jones & Furner, 1989	--	S>A>E	--	S>E>A
Keller & Stevens, 2004	$A_n \approx Ai > Aa$	$A_n > A_i > A_a$	--	--
Leung et al., 1997	--	S ≥ A > M	All Improved	--
Lucas, 1994	S>A ≈ E	S>A ≥ E	All Improved (≥ A)	--
McGee-Lennon et al., 2007	S ≥ E ≈ T (Cognitive Effort)	--	--	S ≈ E ≥ T
McKeown, 2005	$S \approx A > A_a \approx T$	$S \approx A > A_a > T$	A_a & T Improved	$S \geq A, A_a, T$
Sikora et al., 1995	--	--	--	M > T ≥ A
Roberts & Sikora, 1997	--	--	--	S>M>A
Stevens et al., 2006	A>M	A>M	--	--
Ulfvengren, 2003b	$S \approx A > A_a, T$	$S > A > A_a > T$	$A_a \leq T$	--

Table 3-2: Overview of literature on comparative effectiveness of sound types as notifications and alerts.

One ought to be careful drawing generic conclusions from the overview of the studies presented in Table 3-2. There are a number of uncontrolled factors that can influence findings on cognitive effort, learnability, memorability and preference of the auditory notifications. To name a few: type and amount of training, success criteria, quality and other properties of the stimuli presented, labelling and depicting of functions or events represented by sound, signal-referent relationships, number of associations presented, size and typicality of the samples of participants. Although efforts to follow design guidelines of the respective audio types are often reported, it is practically impossible to evaluate the thoroughness of these efforts. Besides, the guidelines themselves are of somewhat abstract nature, since audio signal design involves and requires a level of artistic creativity. Few researchers have attempted to minimise the subjectivity of the stimuli design by assigning the process to professional musicians. However, even fewer have involved samples of (representative) end-users in the design process to empirically validate the sound sets before they compare them.

In spite of the diversity of all those factors, there seems to be a consistent trend in the cognitive effort and ease of learning amongst the studies, with sounds with more direct signal-referent relationships scoring higher. In other words, speech usually performs better than auditory icons, which perform better than earcons, which in turn perform better than simple tones. Also, familiarity of sounds (events causing the everyday sounds utilised in auditory icons, or musical instrument timbres utilised in earcons) increases learnability. On the other hand, there is too little or contradictory data to draw any conclusions on memorability or user preference. Participants usually performed better in remembering the associations after 3, 7 or even 14 days. This suggests that the difficulty of retaining the sounds reached a ceiling effect in performance and hence the respective forgetting curves could not be studied. Moreover, user preference has been widely varying, as it seems to depend more on the user sample and the context of use of the notifications. It should be noted that users have been without exception asked to state their preferences on different types of sounds in short laboratory sessions, away from the real context of use of the notifications.

The inconclusiveness or uncertainty of these findings could also stem from the fact that sound types have divergent advantages and disadvantages. Therefore, some researchers have suggested or prototyped hybrid auditory interfaces, utilising a combination of two or more sound types. For example, it is suggested that auditory icons are used more for system functions that naturally lend themselves to event-produced sounds through metaphors, while earcons are used for the more abstract system functions or when information structure is important [Brewster, 2002b; Gaver, 1997]. A handful of studies have applied some form of hybrid auditory interface with encouraging results and positive user response. [Albers, 1994] suggests a simple 5-step methodology for developing hybrid auditory interfaces that consists of:

1. Choosing the basic auditory interface techniques to combine.

2. Identifying the benefits and limitations of each chosen technique.

3. Combining the benefits to alleviate the limitations of each technique.

4. Realising where merging some aspects of each technique is problematic.

5. Developing strategies to relieve the incompatible combinations.

He then used this methodology to develop an interface for a satellite ground control system that combines auditory icons and sonification (using data to directly manipulate properties of sound, such as frequency). A similar interface was developed for operating theatre notifications consisting of parameterised auditory icons, where relationships between the sounds and their properties (such as pitch) and the parameters they represented ranged from nomic to metaphoric to arbitrary [Fitch & Kramer, 1994]. It was found that participants performed better with the auditory interface than with the alternative visual interface, but only after substantial training (about 1 hour). Also, attempts to combine presentation of earcons and auditory icons together, either in sequence or in parallel, were harder to learn than auditory icons alone [Dingler et al., 2008]. In another example, Nomadic Radio combines speech and auditory icon notifications to create a scalable manner of information presentation (email, calendar events etc.), depending on the interruptibility of the users [Sawhney & Schmandt, 2000]. Similarly, speech and earcons were combined to convey information in a multimedia stock control information system, depending on the amount of information to be conveyed each time [Rigas & Memery, 2002].

The need for further systematic research in hybrid auditory interfaces has been identified by many [e.g. Brewster, 2002b] but is not the immediate focus of this thesis. We believe there still remain many and fundamental questions in relation to the understanding and the effectiveness of different sound types as notifications. In particular, how effectively can the associations be learned implicitly through everyday usage? How is user preference affected over time and how is the forgetting curve shaped for different notification types? Most importantly, how do we involve users in the designing process of the notifications to ensure appropriateness and pleasantness of the stimuli? By exploring and addressing these challenges, we also set a stronger basis for successful design of hybrid auditory interfaces.

3.5 Chapter Summary

In this chapter we arrived to the final form of our research question: "How can we design for non-intrusive yet informative auditory mobile service notifications?" Scoping from multimodal notifications, we decided to focus on the auditory channel for four reasons that were supported with literature and analytical evidence: users' dependency on vision while on the move; flexibility in the semantic richness of auditory interfaces; human ability to habituate or attend to multiple audio streams; position of mobile devices in relation to their owners. These factors make the auditory channel the most appropriate and promising for mobile service notifications.

In the rest of the chapter, we presented an extensive literature review on the three types of sounds most commonly used in auditory notifications: earcons, auditory icons and speech. Furthermore, these sound types were analysed on the two orthogonal dimensions of sound-referent relationship, and culturally attributed meaning of the sounds. By doing so, we were able to graphically demonstrate the relationships of the sounds on a 2-dimensional topology. One of the advantages of this view and representation is that the nature of notifications can be judged regardless of the class they belong to. Also, it demonstrated the fuzziness of the definitions of the sound types.

Finally, studies comparing the effectiveness (response times and cognitive effort, learnability, memorability, user preference) of different notifications types were presented and compared. Although there are many uncontrolled factors that can influence findings (such as type and amount of training, quality and other properties of the stimuli presented), certain trends could be extracted from the literature. There seems to be a consistent trend in the cognitive effort and ease of learning, with sounds with more direct signal-referent relationships scoring higher (speech performing best, followed by auditory icons, followed by earcons, and then simple tones). Also, familiarity of sounds found to increase learnability. On the other hand, there were too little or contradictory data found to draw any conclusions on memorability or user preference.

This lengthy investigation gave us a deep understanding of the area, which informed the studies we conducted and present in the following chapters.

Chapter 4

Auditory Notifications for Mobile Service Awareness

The effectiveness of auditory notifications can be measured along several dimensions, depending on the goals of the notification systems and the contexts of their application. Many of the studies in the relevant literature (see Chapter 3) aim to improve warnings and alarms in cognitive demanding contexts, such as cockpits and operation theatres. Pilots, anaesthetists, satellite ground-control operators, power plant operators etc. are required to work under unusual cognitive effort demands, coordinating several operating and communicating processes. In these contexts, learnability of notifications and immediate, appropriate reactions are potentially life-critical. Therefore, relevant stimuli design guidelines focus on metrics such as response time, ability to learn and retain, perceived urgency, and user preference. However, there has been no systematic investigation of the effectiveness of auditory notifications for mobile services, or any relevant design guidelines.

In this chapter we present a first such investigation, and in particular address two challenges specific to this domain. The first challenge arises from the fact that mobile use context is particularly dynamic as it can wildly vary on aspects such as cognitive and attention demands, and social context. It is therefore required that notifications about mobile services blend naturally in the environment and attract attention only when truly relevant (see Section 2.6). We suggest that one way of achieving this is through meaningful sounds that can be habituated when irrelevant or, similar to the cocktail party effect, pop up to awareness when the information they convey is (preattentively judged as) important or relevant. This design challenge of "smooth notification" has also been identified in other auditory rich environments (e.g. power plants), where the aim should be "to enable sound events to fade into background of awareness if not needed, still be perceptive enough to inform a user about a change" [Alexanderson, 2004]. Furthermore, unlike mobile phone users, skilled operators are expected to undergo substantial training in order to learn and memorise warnings. Similar to walk-up-and-use interfaces (e.g. cash-points), it is expected that mobile phones should be adequately self-explanatory to be readily and easily operated by the general public. On the other hand, if notifications are irritating or unpleasant, they will not serve their purpose of habituating in the background when irrelevant, and (as in many other application domains) they could be deactivated altogether. Interestingly, a 2006 survey portrayed British people being very self-conscious about their choice of ringtones and how it is perceived by others. According to Phonecontent.com[1], 80% of British peo-

[1] "Britons Too Worried About Choosing Ringtones" - phonecontent.com, 02 May 2006: http://www.phonecontent.com/bm/news/1468.shtml (last accessed: 30/06/2009)

ple were "worried what others would think of their choice of ringtones", while 90% of them felt they "have been ridiculed by family and friends for selecting weird ringtones". Therefore, a careful design process is needed to produce intuitive and pleasant mobile service auditory notifications.

A second challenge particularly associated to the domain of mobile service auditory notifications is the classification of services. With recent technological advancements (e.g. more powerful smartphones, expansion of the mobile internet bandwidth), mobile services have multiplied, and they are expected to grow even in larger numbers in the years to come. Even if a small portion of them have or develop notification functionality, it will be virtually impossible to encode the information they wish to convey in meaningful visual, auditory or tactile notifications (excluding speech and text). Although the number of notifications received could be determined by voluntary user subscription and appropriate filtering depending on context, their variety and diversity could impose an impossible challenge in learning and retaining their uniquely coded notifications. Thus, there is a need to organise mobile services in a manageable number of meaningful clusters, which can be represented by the same notification that would convey their common nature.

The remainder of this chapter is divided in two parts. First, we present our initial experimental investigation on the intuitiveness and preference of different sound types as mobile service notifications. The experience gained and the lessons learned are discussed here, and inform the design of our more extensive, longitudinal studies, which are presented in Chapters 5 and 6. The second part of this chapter addresses the challenge specifically relevant to mobile service notification design: the need for a comprehensive service classification. We have developed a rationale that can encompass all existing and upcoming services, and clusters them in a meaningful and manageable way. The development of the classification and the experimental investigation are intertwined. The first version of the classification (Section 4.2.1) was used in the experiment (Section 4.1), and findings from the experiment informed the empirical improvement of the classification (Section 4.2.2), which is used in the studies presented in Chapters 5 and 6.

4.1 A First Investigation in Auditory Notifications for Mobile Service Awareness: Intuitiveness & Preference

In this section we present a first approach to an experimental evaluation on the effectiveness of auditory notifications for mobile service awareness. In particular, we investigate intuitiveness and user preference of a set of speech, auditory icon and earcon sounds as notifications that denote availability for a set of mobile services. The lessons learnt and the insights gained from this first evaluation have informed the service classification presented in the next section and the design of the longitudinal evaluation studies presented in the chapters following.

Intuition is defined as "the power or faculty of attaining to direct knowledge or cognition without evident rational thought and inference" or "immediate apprehension or cognition"[1]. For example, a sound is more intuitive the less cognitive effort and time it requires from a naïve user to interpret its meaning, and the ultimately intuitive sounds would be the ones whose immediate and correct interpretation contributes to the survival of our species (e.g. the roar of a wild animal or a loud bang). These sounds can cause a primal reaction (e.g. flight or fight, instinctive protection of the head) that overrides the cerebral cortex. However, most intuitive sounds are the ones we have learned their meaning through repetitive implicit learning throughout our lives. The sound of a siren, a phone ringing, or a car revving are examples of sounds that we can immediately interpret because we have been exposed to them from our childhood (cultural restrictions apply). These are the sounds that appear higher up on the Y-axis of Figure 3-6.

Intuitiveness in HCI can be quantified and measured as the likelihood of a naïve user correctly inferring the interactional meaning of an identifying feature of an interface without having encountered that identifying feature previously. When a notification sound is being used to represent something different from its intrinsic meaning (even if that meaning is empty), there is a second level of interpretation required. The listener needs not only to understand the meaning of the sound but also the concept it represents in that particular context. Therefore, the most intuitive notifications are the ones where the meaning of the sound and its association to the relevant function is understood with least effort and time. This is exactly what auditory icons are trying to exploit. By assigning a meaningful sound (such as an everyday sound that we already know) to a function that is related to the sound's meaning, the association becomes more intuitive too.

As mobile services rise in numbers, an obvious challenge is to provide notifications whose association to the services are intuitive. Therefore, one should expect that auditory icons should be more intuitive service notifications than earcons. However, one would have to utilise adequately good metaphors connecting signal and referent of auditory icons in an intuitive manner. On the contrary, if these metaphors are misleading, auditory icons could end up being counter-intuitive, and thus require more effort to understand and learn. If users can automatically and consistently identify these associations, it means that their underlying implicit cognitive links are present.

There is strong evidence in the literature comparing learnability of different sound types that auditory icons are easier to learn, possibly because of their meaningful associations to interface objects or functions (see Chapter 3). However, none of these studies have separately measured intuitiveness from learnability. Since it is common to provide some sort of training before any testing takes place, intuitiveness (immediate cognition of the notifications) has not been tested. Training usually consists of a combination of three steps: presentation of the stimuli and their associations, explanations of the signal-referent associations, and active exploration through clicking on

[1] Merriam-Webster Online: http://www.merriam-webster.com/dictionary/intuition (last accessed 30/06/2009)

labels/icons to hear the corresponding sound. All of these processes involve learning and therefore the follow-up initial testing trial measures learnability.

We therefore conducted a controlled experiment to measure recognition rates of auditory icons, earcons and speech notifications for mobile services. This involved no training or familiarisation of the particular stimuli and associations tested. Although the definition of the sound types and the relevant literature suggest that auditory icons will be more intuitive than earcons, no other experiment has verified this long standing assumption. Furthermore, if this direction in recognition accuracy is verified, it will serve as a confirmation that the specific auditory icons we designed utilised appropriate metaphors for the mobile services involved. Also, it could give us the opportunity to observe if there are any tendencies to assign particular sounds to specific services. It has been reported that users apply semantics to associations even if they were not intended by the designers [Brewster, 1998; Cohen, 1994]. This experiment is therefore beginning to address the 'research gap' (also identified by [Simpson, 2007]) between studies that measure effectiveness of sound types and studies that focus on identifying meaningful sound-signal relationships. An additional advantage of not training our participants to the experimental stimuli is that we can gather subjective data in relation to user preference, which are not influenced by their learning performance. Although there are examples in the literature that associability and learnability did not interact with user preference, we are interested to see how naïve listeners would like the different sound types.

4.1.1 Method

Experimental Design

The experiment had a within participants design. The independent variable ('sound type') was the type of audio notification used: *earcons, auditory icons* or *speech*. The dependent variables were *accuracy* (i.e. the number of errors when attempting to choose the correct service from the nine available) and response time. It was predicted that speech notifications would lead to most accurate responses, followed by auditory icons and then earcons (H1). It was also predicted that the same trend would stand for response speed (H2).

Participants

Following pilot trials with 2 participants, the experimental participants were 29 university students: 13 female, 16 male, with an age range of 18-21 (*M=19.7, SD 1.37).* No reward was given for their participation.

Instruments and Measures

The experimental setup was an IBM desktop personal computer (PC) connected to a 42" plasma monitor with touch screen functionality. User responses to our auditory notifications were collected using the touch screen.

Visual Stimuli

The services were first conceptually categorised to mitigate the potential problem of learnability of huge number of notifications, as clustered services can share the same notifications. The process and the rationale of the categorisation of 52 individual services is presented and discussed in next part of this Chapter (Section 4.2.1). The resulting clustering provided three main categories, with three main subcategories within each category. However, this structure was not presented to the users, who were just given the nine subcategories of services. Two of the subcategories were initially named 'streaming media' and 'mobile community' but it was decided that participants would be most likely unfamiliar with these terms. For clarity, we replaced them with the most popular instantiations of them: 'Television' and 'Instant Messaging' respectively. Furthermore, the choice to concentrate on nine services was not an arbitrary one. There is evidence that people can quickly learn and retain four to six different warning sounds, while learning slows down considerably when exposed to 10 warning sounds [Patterson & Mayfield, 1990]. Elsewhere [e.g. Sorkin, 1987] it is suggested that a maximum of four to six sounds is optimal, due to learning time and the requirements of memory in terms of long-term retention. We decided on a slightly higher number than the optimum suggested in the literature to avoid ceiling effects, and thus allow any potential differences between the sounds to show. The nine options presented to the users appear in Table 4-1.

Messaging	Information Retrieval	Downloads
Television	Calls	Instant Messaging
Calendar	To Do	Synchronise Devices

Table 4-1: Services presented on the touch screen during the experimental phase

Auditory Stimuli

Our experimental stimuli were a collection of 27 sounds, 9 per sound type (earcon, auditory icon, speech). To compare the intuitiveness of the three types of sounds, we had to ensure that each sound was related to its corresponding service as directly as possible. Ideas for how best to represent each service were produced during a brainstorming session amongst the experimenters and are described below. The inherent difference between the sound types, and the necessity to find the most appropriate metaphors to relate to the services meant that the duration of the notifications could not be controlled. Therefore, the length of our sounds varied from 0.5 sec to 4.21 sec (M=2.27, SD 0.99). Not surprisingly, the length variance for auditory icons was the highest (SD=1.36), as the appropriate sounds had to vary considerably in order to achieve the realism needed. The sound files were encoded in .wav files at 1411kbps, apart from the auditory icons, which were collected from royalty free online libraries or created by the experimenters, and varied from 88 to 705kbps (M=332.89, SD=281.36). The volume of the sound was not normalised, with auditory icons varying the greatest and speech sounds being on average approximately 8 to 10db louder than the other types. Detailed stimuli characteristics can be found in Appendix III.1, while the sound files can be found on the thesis' accompanying CD (folders "Chapter4.Earcons", "Chapter4.Auditory_Icons" and "Chapter4.Speech").

Earcons

One of the strongest advantages of earcons is that they can be composed to form families of sounds, with each family representing different actions on the same object or similar actions on different objects [Gaver, 1989]. In this case, each one of the three main categories of services was conceptually distinct to each other, and each of them contained sub-categories of services that were semantically related. Therefore, we created three families of earcons that were different enough from one another to represent the three different categories. This distinction was created by the use of three highly distinctive instrument timbres, one for each category (oboe, piano and bells), as suggested by established earcon guidelines [Brewster et al., 1993]. In addition, we designed earcons to be as distinct as possible within each category by varying their pitch and rhythm, again as suggested by the guidelines. A music student was responsible for producing the sounds.

Finding an intuitive way to relate earcons to the concepts of the mobile services was challenging. However, in an attempt to maximise intuitiveness, we strove to make our earcons as metaphorical as possible. In some cases, the metaphor was with respect to the syllables of the words, as suggested by melodic alarms guidelines [Block et al., 2000]. For example, the ‘To-Do’ earcon was represented by the sound of two notes dropping in pitch from one note to the next, while ‘Calendar’ had three sounds. In other cases a conceptually higher metaphor was possible. For example, the ‘synchronizing devices’ earcons was a collection of notes followed by a short rest, and then the same collection of notes repeated. Finally, stronger – albeit culturally specific – metaphors were used: the ‘Television’ earcon, for example, mimicked the intro sound of the popular British television show ‘Eastenders’.

Auditory Icons

Finding iconic mappings between auditory icons and services in some cases was straightforward. For example, an obvious choice for an incoming call is a classic telephone ring, and ‘Television’ was represented by the sound of a TV switching on to white noise. However, in some cases it was not always obvious how to represent a service with an everyday sound. In these cases we tried to apply a more metaphorical mapping. For example, the sound of a consistent dripping sound was used for the ‘Downloads’ service. This was meant to represent the constant stream of data being received while downloading a file.

Speech

The speech notifications were prepared using a pre-recorded and non-synthetic female voice describing each of the services. For example, the ‘Calls’ service was notified as “You have an incoming call” and ‘Downloads’ was notified as “You have a new download”.

Procedure

Each participant was informed about the nature of the experiment and was given a brief definition of the three different sound types. A training session preceded the experimental phase, and participants were invited to complete a short questionnaire immediately subsequent to the experiment. The whole procedure was conducted with at least two experimenters present. Details of each phase of the procedure are described in the following sections, while relevant material (consent form, instructions and questionnaire) can be found in Appendix III.

Training

In most research with audio notifications, participants are informed of the associations between sounds and referents during a training session. However, since we wanted to measure intuitiveness of the notifications, we deemed it necessary to exclude this kind of training. So, although participants were familiarised with the three different types of sounds, they were not exposed to the experimental stimuli during training.

During the training phase, we used animal sounds as they are easily understood and distinct from the sounds used during the experimental phase. Participants were trained in two steps. First, they were given time to become familiar with the three different types of sound and the visual interface. The types of sounds were explained to the participants (auditory icons were referred to as 'natural sounds' throughout the experiment), and their respective principles applied on the animal sounds. The interface was a large screen with 9 buttons (arranged in a 3x2 grid), each one initiating the sound associated with the particular animal (Table 4-2). Participants were asked to interact with the application until they felt comfortable with the interface.

Donkey	Mouse	Sheep
Dolphin	Ape	Pigeon
Elephant	Mule	Snail

Table 4-2. Training interface

In the second step of the training, an example evaluation trial was conducted so that participants became familiar with the evaluation procedure. We presented a series of 9 different animal labels on the screen (Table 4-2) and played the three auditory representations for two of them (total 3x2=6 sounds) in randomised order. For each sound, the user was asked to select the animal that they thought the sound represented. Each sound was heard up to three times (or until the participant made a choice), and the next sound was played after a 2 second silence interval.

Experimental Phase

At the beginning of the experimental phase, each participant was presented with the 9 services (Table 4-1) and was given a short definition for each one of them. During this time, the participant was given time to learn the fixed position of each service on the screen, and was informed that each service had three sound representations, and that the same sounds would be heard more than once during the course of the evaluation.

The experimental phase comprised of three blocks of 27 trials (total: 81 sounds), in which participants were presented with all three types of sound for each of the 9 services. In order to address possible learning effects, the order of presentation of the notifications was pseudo-randomised, ensuring that the same type of sound was not played in two consecutive trials.

Once the test phase was complete, participants were asked to complete a short questionnaire, capturing their preferences and subjective comparisons of the three sound types (Appendix III.4). Each participant took approximately 25 minutes to complete the experiment.

4.1.2 Results

The dependent variables were the number of errors (selection of an incorrect service) and response times of participants across trials (recorded by custom-made logging software) from sound onset.

Quantitative data

The results presented in this section were generated from two measures: *accuracy* and *response time (RT)*. Mean and percentage scores for all measures were calculated and are presented in Table 4-3. Accuracy mean scores for the three sound types are also presented in Figure 4-1 for each service separately. Accuracy scores were not normally distributed but were normalised (prior to any statistical analysis) using a logarithmic transformation.

Comparisons of the accuracy scores and response times across the three *sound type* conditions were performed using *two-tailed* paired-samples *t*-tests. However, due to the accuracy rates for the speech condition (almost 100%), comparisons including the *speech* condition were not considered for further analysis. A statistically significant

Sound Type	Earcons	Auditory Icons	Speech
Accuracy (out of 27)	4.86 (18%)	10.10 (37.4%)	25.90 (95.9%)
Accuracy (SD)	2.43	3.32	2.99
Response time (Sec)	5.11	4.82	3.26
Response time (SD)	2.55	3.08	1.40

Table 4-3. Mean accuracy scores and response times

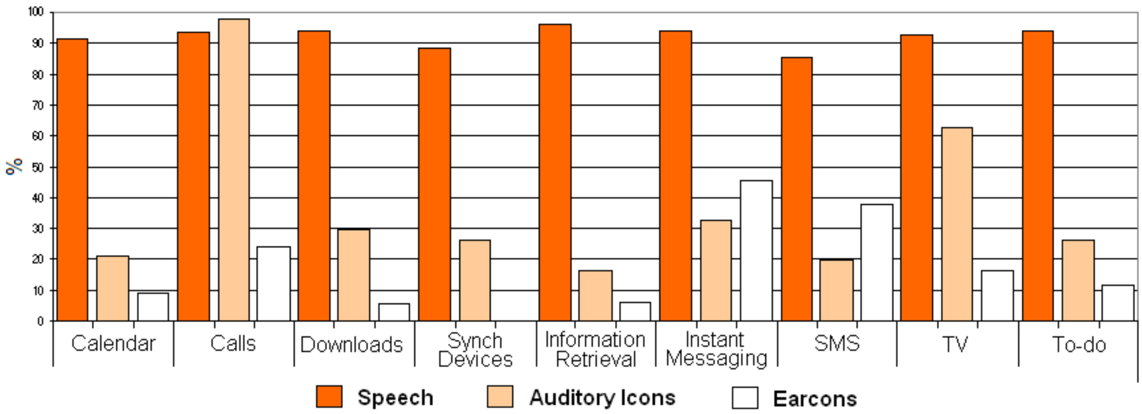


Figure 4-1: Accuracy mean scores for the 3 sound types for each service separately

increase in accuracy scores was observed for *auditory icons* (*M*=0.98, SD=.23) over *earcons* [*M*=0.63, SD=.23, *t*(28)=-7.704, *p*=<0.01]. However, earcons for SMS and instant messaging were more accurate than the corresponding auditory icons. Comparative accuracy rates across the conditions are presented in Figure 4-2.

Response times across the conditions are presented in Figure 4-3. No significant difference was found for the response times between the *earcon* and *auditory icon* conditions [M = 297.74, SD = 902.20, *t*(28)=1.777, *p*=0.086 (n.s)], indicating that the time taken to make an identification was the same regardless of whether the sound was an earcon or an auditory icon. Responses to speech stimuli were significantly faster than either earcons [M = 1852.70, SD = 972.65, t(28) = 10.26, *p*=<0.01] or auditory icons [*M* = 1554.96, SD = 657.32, *t*(28) = 12.74, *p*=<0.01]. Although participants made decisions that were faster and more accurate for every service with speech notifications, the exception to the rule was with the 'Call' service, where the auditory icon scored higher.

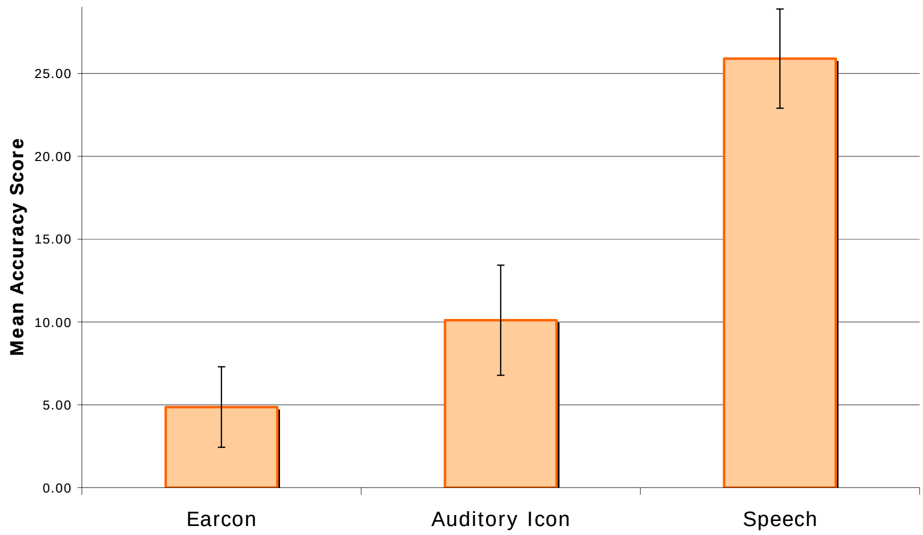


Figure 4-2: Mean accuracy rates across all sound type conditions

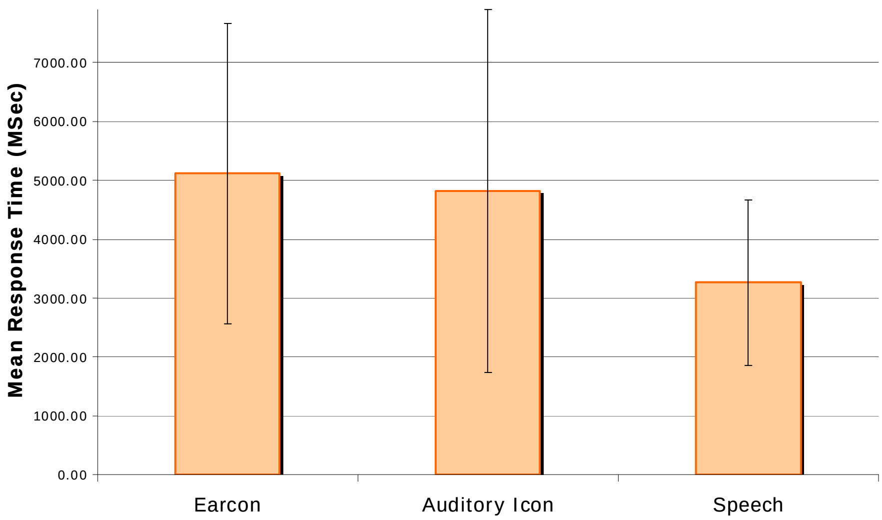


Figure 4-3: Mean response times across all sound type conditions

Qualitative data

Overall, participants' subjective results are in agreement with their performance. In the post-test questionnaire we asked participants to evaluate sound comprehension, their own personal liking for the sounds, and how much they would like to have each notification type on their phones. Also, they were asked to estimate how long it would take them to learn and memorise them if they were using them on their mobile device. Their responses were collected on 5-point Likert scales, which were then converted to a score out of 100. Results for each sound type are reported separately, and are summarised in Table 4-4 (standard deviations provided on the 5-point scale).

Speech

Speech notifications were (unsurprisingly) the easiest to understand. On average participants scored speech at 91% in terms of comprehension and 78% in terms of liking. However, there were mixed views on whether they would like to have them on their own mobile phones. Just over 1/3 of people would want to have them, while just over 1/3 were indifferent to this prospect. Overall, the "would like to have" scale scored 54% on average. Over 40% of the participants reported finding speech notifications boring or annoying or expressed concerns with regard to privacy. Finally, almost 90% of them believed it would take them less than a week to learn and remember the speech associations.

Auditory Icons

Overall, participants rated auditory icon notifications lower than speech but higher than earcons. On ease of comprehension and liking, auditory icons scored exactly 50%, with equal number of participants liking or disliking them. As with speech, there were mixed views on whether they would like to have them on their phones, with an average score of 50% too. On learning and remembering, 83% believed that they would need no more than two weeks.

Earcons

Earcon notifications achieved the lowest scores with comprehension at 31% on average and liking at only 38%. The "would like to have" scale again showed small variation from the other types and earcons scored at 43%, just below auditory icons. However, earcons produced more extremely negative responses comparing to the other types. The perceived ease of learning also scored comparatively poorly (66% on average) with 10% of the participants estimating that they would never learn or retain them.

Scores: % (SD)	Speech	Auditory Icons	Earcons
Comprehension	91 (0.67)	50 (0.85)	31 (1.02)
Preference	78 (0.88)	50 (0.76)	38 (0.99)
Would like to have	54 (0.98)	50 (0.93)	43 (1.03)
Perceived ease to learn	97 (0.44)	78 (0.86)	66 (1.25)

Table 4-4. Mean scores and SD of subjective assessments of sound types

At the end of the questionnaire there were several comparative questions with regard to self-assessment of recognition rate, preference in sound type, and potential usage. The results of these questions were generally in agreement with the actual performance and the ratings for each individual sound type. However, more participants chose earcons over auditory icons as a sound type they would like to use in the future (15 against 11). This contradicts their absolute subjective ratings for each sound type separately, where auditory icons scored higher.

4.1.3 Discussion

Our first hypothesis was supported as auditory icons performed significantly better than earcons in terms of recognition accuracy. Examining individual notifications' results, this advantage was probably due to auditory icons' usually richer semantics. This outcome is not surprising, as it was suggested by the nature of the sound types. Everyday sounds are familiar to people, and they have already been associated with the specific context in which they normally arise. If auditory notifications are to take advantage of this pre-existing relationship, they need to represent a service through a metaphor that relates it to that concept. Users may then recognise these associations with minimum or no training at all, as they did in this experiment.

Furthermore, our results are in agreement with the learnability trend that appears in the literature. Although significance is not reached in every study, all of them have found that more direct signal-referent links are easier to learn (see Section 3.4). Although this is not a surprising finding, we are not aware of other studies that have investigated the effectiveness of notifications without any prior training. We believe that everyday and widespread technology such as mobile devices should be designed to be as intuitive as possible with minimum or no user training requirements.

On the other hand, our second hypothesis that response speed would follow the same trend with accuracy was not accepted. Speech notifications were significantly quicker to respond to comparing to either of non-speech notifications, but no significant difference was found between earcons and auditory icons. These results are generally in agreement with the learnability literature, where speech almost always produces quickest responses than most non-speech sounds, although not always quicker than auditory icons. Similarly, auditory icons sometimes produce significantly quicker responses than musical sounds or tones, although this seems to rely on the strength of the metaphors utilised by auditory icons. Perhaps our results indicate that our auditory icons were not as good as they could have been.

A further interesting result with regard to semantics was that 61% of all earcon notifications were assigned to only three services: messaging, instant messaging and calls. Since the earcons used were designed to be distinguishable, we believe the reason these three services were associated with more than one earcon each may lay elsewhere. Traditionally, all notifications on mobile phones sound closer to earcons (rather than speech or auditory icons) and mostly notify users of incoming messages or calls on mobile phones (i.e. ringtones), or instant messaging applications (such as MSN) on computers. Therefore, it is most likely that participants, based on their experience, related earcons to messaging services and calls. So, it was probably this pre-existing association between earcon-like sounds and the three most popular services that produced the biased responses. This preconception may have interfered with the participants' willingness and ability to create new associations between earcons and other services.

Further evidence supporting this suggestion is the fact that messaging and instant messaging were the only two services where earcons outperformed auditory icons in terms of accuracy (see Figure 4-1). Although these earcons were not identical to the widespread notifications on current popular mobile phones, one could argue that their resemblance made these earcons more effective. Interestingly, such common everyday notifications (e.g. the classic Nokia sound for incoming sms) blur the boundaries between auditory icons and earcons (as discussed in Section 3.3). As the association is established through repetitive widespread usage, they are appropriated and cross over to the auditory icons arena. The same could be argued for the traditional telephone ring itself. When it was first introduced it was nothing more than a noise or arbitrary sound. Through the years it became synonymous with the telephone sound. Therefore, it is not surprising that we found certain auditory icons to be very accurate (e.g. old style phone ringing for 'Calls', and TV turning on with white noise for 'Television'). This further suggests that there are everyday sounds (whether natural or synthetically produced), which if utilised as auditory notifications, can create strong intuitive links to mobile services.

Moreover, the subjective data collected through the post-test questionnaire follow the same trend, with speech being the easiest to comprehend and most liked, followed by auditory icons and then earcons. However, these gaps in preference get smaller when users are asked to envision which notification type they would like to have on their device. It is interesting to note for example that the superiority of speech notifications in accuracy, response speed, and liking, was not reflected in users' willingness to adopt them. This indicates that there are possibly other factors that affect potential adoption of speech notifications. For example, the particular voice and chosen words of the notifications need to be considered and carefully manipulated to design acceptable speech notifications. However, it is unsafe to draw any conclusions on user preference results from a single lab study. The contradictory results we recorded (comparative preference did not match absolute preference of each type) suggest that it might be difficult for lab studies to enable people to envision usage of auditory notifications in the rich and complex audioscape of everyday life.

Limitations

There are several limitations with our experiment, some common to this type of evaluations and some specific to our particular evaluation. First, as with any study in auditory interface, there is certain subjectivity induced by the particular stimuli examined. Even when following literature's guidelines, earcon design involves some level of artistic input and personal interpretation that are difficult to measure, control or standardise. The same applies for the associations chosen between auditory icons and services, which in fact were proven successful (or not) to varying degrees. Also, speech notifications have many parameters, such as gender and emotional expression of the voice, or the wording, which were not explored and may affect their effectiveness (especially the user preference and acceptance). It is possible that results of comparative experiments are affected by all of the aforementioned (and possibly more) uncontrolled parameters of the particular stimuli sets under investigation. One way that this limitation could be partly addressed is through participatory design of the notification stimuli. If empirical evidence is collected about the effectiveness of each stimulus set, then iterative design could be used to improve it. However, this approach is rare in the literature as it is costly in resources (time, money, participants).

A particular limitation in our experiment is that it did not present the hierarchy of the service classification to the participants. This made it impossible to analyse earcon efficiency in conveying information about the service classification structure. This structure, and the equivalent earcon structure, is usually presented to participants during the training session of similar experiments. However, since we were aiming to compare intuitiveness of notifications, we decided to exclude any training on the experimental stimuli. Had we explained the structure to the participants, it would have started the learning process and give earcons an advantage. On the other hand, it is improbable that any participant was able to 'see' this structure either in the services or the earcons. Admittedly, presenting the service structure would have been preferable, as it would give more chances for the timbre-category association to arise. Although our earcons were designed to be hierarchical, it would be perhaps fairer to say that they were only tested in their capacity of being distinguishable and metaphorical, similar to what some experiments in the literature refer to as 'musical sounds'.

Our implementation also posed two further limitations. First, the auditory stimuli were poorly controlled. The length of the different notification types varied out of necessity, as each type requires different length to maximise its efficiency (convey adequate information in limited time). Although this was not taken into account when recording response times (counter started with the beginning of each sound), one could argue that notification conciseness is one factor of its efficiency. However, there was no reason (other than carelessness) that the stimuli were not normalised in volume and quality (kbps). These differences occurred by the way sounds were created or gathered, and it is worth noting that they were not noticeable to the experimenters or reported by any of the participants. Second, our custom-made software presenting and recording the experimental data was sometimes lagging, thus possibly affecting the accuracy of response times. However, this was a random effect, not limited to a single type of sound.

Finally, it was observed that participants were sometimes focusing disproportionally more to the speed of their responses than their accuracy. This could be the result of either feeling pressurised by the presence of the experimenters, or feeling impatient and bored by the length of the procedure. Although the whole of the experiment (including answering the questionnaire) did not take more than approximately 30 minutes, perhaps the 81 trials felt too many for some participants. There is some anecdotal evidence that some participants could not "see the point" of what they were doing, as they received no feedback on their performance and there was no indication bar for their progress. Perhaps fewer trials, overall feedback or a clear indication of their stage in the process would be appropriate. Nevertheless, the possible hastiness some participants may have felt does not invalidate our findings. On the contrary, less conscious thinking could have forced some participants to go for the most intuitive answer.

Despite its limitations, this experiment has achieved two things. First, it has demonstrated how the signal-referent directness can be responsible for intuitiveness (as literature has shown for learnability), and therefore auditory icons can produce more intuitive mobile service notifications. Second, it has given us experience in designing and executing the first experimental comparison of auditory notifications for mobile services. The lessons learned both in stimuli and experimental design will inform more substantial and extended studies on metrics, such as longitudinal learnability and preference, and memorability of empirically evaluated auditory mobile service notifications. These studies are presented in the following chapters of the thesis. Finally, this study has stressed the need and initiated the process of structuring mobile services in meaningful hierarchies and clusters, in order to reduce the problem of requiring learning of impractical numbers of notifications. In the next section, we present the rationale of the hierarchy that instructed the list of services used in this experiment, and the process of its improvement for the follow-up studies.

4.2 Mobile Service Classification

Currently, the categorisations of mobile services found on the web sites of mobile network providers are different, depending on the marketing or other objectives of each carrier. They exist not only to inform users and potential customers of their services, but also to promote and encourage their usage. This has led to categorisations that are not strictly hierarchical (the same services can often be found in multiple subcategories). For example, the ‘Cinema’ service can be seen as an instance of ‘Retrieval Information’ services as it provides information about movies, or as an instance of ‘Downloads’ services as it also provides the possibility of downloading related content (such as ringtones related to a specific movie). The choice of categories can be radically diverse depending on the clustering concepts. For example, a ‘Downloads’ category can include games and ringtones downloads, but the same services could be found under the categories ‘Entertainment’ and ‘Device Personalisation’ respectively.

Suggesting a comprehensive universal categorisation of mobile services is not a trivial task. First, one needs to create a categorisation of services depending on the purpose they will serve. For example, when services are clustered around a common concept (e.g. ‘Downloads’), it is not necessary that this concept (or its label) is the most efficient for notification purposes. The cluster ‘Downloads’ can include a widget to monitor calorie intake, the ringtone based on the theme of the movie, or a car racing game. These three different downloads will be relevant in completely dissimilar contexts (e.g. at a restaurant when ordering, at the cinema when finished watching a movie, and at the airport when the flight is delayed), and therefore should be represented by distinct notifications. Since the existing categorisations do not fit the purpose of notifications, we need to create a new way of classifying services based on how similar (or identical) notifications of similar but distinct services need to be.

A second challenge to mobile service classification is related to the decision of which and how many services to include. According to our definition (see Section 2.1), mobile services can include network connectivity and other devices, or they can be standalone applications on a handheld device. Although most services are currently designed to be initiated by users, we decided to follow a highly inclusive approach. We envision that any of the existing services could potentially be suggested to users, provided they are relevant and desired. In fact, we need to devise a classification that will be able to accommodate not only all the existing services, but also those who are to be conceived and developed.

4.2.1 Analytical Approach

We collected 50 service descriptions and information from the websites of three major UK network operators. From this information, we created service classifications for each operator using a mind-mapping tool, giving services of the same description identical labels. We considered performing cluster analysis by evaluating the distances between all pairs of services in all three classifications, but this idea was discarded for three reasons. First, three classifications were too few to be used as input data for cluster analysis. Second, the classifications provided were not strict hierarchies as services often appear in more than one category. Third, the classifications seem conceptually very dissimilar. For example, some categories were based on the delivery method ('alerts') and some on the content ('entertainment'). Therefore, we felt it was necessary to construct a new classification to accommodate all 50 services. Furthermore, we added two more services that were not commercially available at the time: 'Location Messaging' and 'Friend's Location'. This was done to ensure that our classification rationale would not be restricted by existing services, and it would demonstrate the ability to incorporate new ones. The list of services along with descriptions appears in Appendix IV.1.

The usual classification attributes, such as delivery method or content, do not adequately characterise all mobile services (e.g. 'Calls' have no content and 'Sports Information' can be delivered in many different ways). However, one attribute that did fit our purpose is the 'source' or 'origin' of the incoming services. When a phone rings for example, the first question in mind most likely will be "who is it?" For incoming calls, different ringtones can be assigned to individuals or groups (e.g. 'family'). Of course, applying the same rationale to all services (such as calls, sports information or calendar reminders), we need different audio notifications to denote their origin. A similar approach was followed when encoding multimodal notifications for three mobile messaging services (voicemail, text and email) [Hoggan & Brewster, 2007]. In that study, three types of senders were identified depending on the associated content of the message, and were named: 'Work', 'Personal', and 'Junk'. However, this classification is limited as it can only be directly applied to person-to-person services.

Looking more closely at the services currently provided by network operators, we were able to distinguish three different distinct sources of origin: an impersonal third party (such as a company or a server), another person (known or unknown to the user) or the user herself (e.g. when setting diary reminders). This provided us with a hierarchy that conceptually separated all 52 services into three major categories: 'World-to-User', 'User-to-User', and 'User-to-Self'. Subcategories and examples of services for each category are described below, and the overall structure appears in Figure 4-4.

World-to-User (W2U) services deliver information or other type of content, such as news headlines, traffic information and songs. All the services that are and could be designed to support everyday activities would normally fall in this category. W2U can be further broken down into 'Information Retrieval' services, 'Downloads' and 'Streaming Media', depending whether the content is being browsed, downloaded or streamed live. 'Information Retrieval' subcategory can be further broken down to three categories, according to type of content: 'Fun' (e.g. cinema, sports, music services), 'Work' (e.g. news headlines, stock market, traffic services), and 'Directory Enquiries' (e.g. services for finding people, businesses, or places). The types of services under 'Downloads' were mainly around 'Entertainment' (e.g. music tracks, games) and 'Personalisation' (e.g. ringtones, wallpapers). The available services under 'Steaming Media' were only TV (providing movies, series, documentaries etc.) and Radio (providing music, live news, talk shows etc.).

User-to-User (U2U) services are predominantly communication services amongst individuals, which allow them to connect and share content. U2U can be further broken down to 'Calls' (voice, video, conference calls etc.), 'Messaging' (sms, mms, email etc.) and 'Mobile Community' services that extended current mobile users' connectivity. Two services that were not commercially available at the time were added in this branch: 'Location Messaging' under 'Messaging' category and 'Friends' Location' under 'Mobile Community' sub-category. The first has been developed in several research prototypes [e.g. Munoz et al., 2003], and extends the messaging functionality to include location as one of the parameters of message delivery. 'Friend's Location' was also developed in prototypes (e.g. The Hummingbird [Holmquist et al., 1998]), and allowed users to see their peers' location on a map in real time. This idea has now been commercialised and mainly popularised via Google Latitude[1].

The third and final top-level category, User-to-Self (U2S), includes services created by individuals and targeted to themselves. These are different types of reminders and we have broken them down in three subcategories: 'Calendar' (e.g. meetings, lectures, birthdays etc.), 'To-Do' (items on the to-do list of users, with flexible and dynamic sub-clustering), and 'Synchronise Devices' (e.g. mobile with laptop or online account). These reminders currently have limited connectivity and are usually stand-alone applications on the mobile device. However, reminders that are triggered depending on the users' context are common in the literature of context-aware computing (see Section 2.3.1).

[1] http://www.google.com/latitude/intro.html (last accessed 30/06/2009)

The categorisation discussed here was used in the experiment described in the first part of this chapter (Section 4.1). The 9 services situated at the second level of the hierarchy were selected for the experiment, with the exception of 'Streaming Media' and 'Mobile Community', which were regarded as too abstract. Instead, they were replaced with representative instances of services from within these categories that the participants were expected to be more familiar with ('Television' and 'Instant Messaging' respectively). Although the classification was not presented to the participants in that experiment, its narrow width makes it a good candidate for future experiments, as there is evidence that narrow hierarchies yield better navigation performance on mobile devices [Geven et al., 2006].

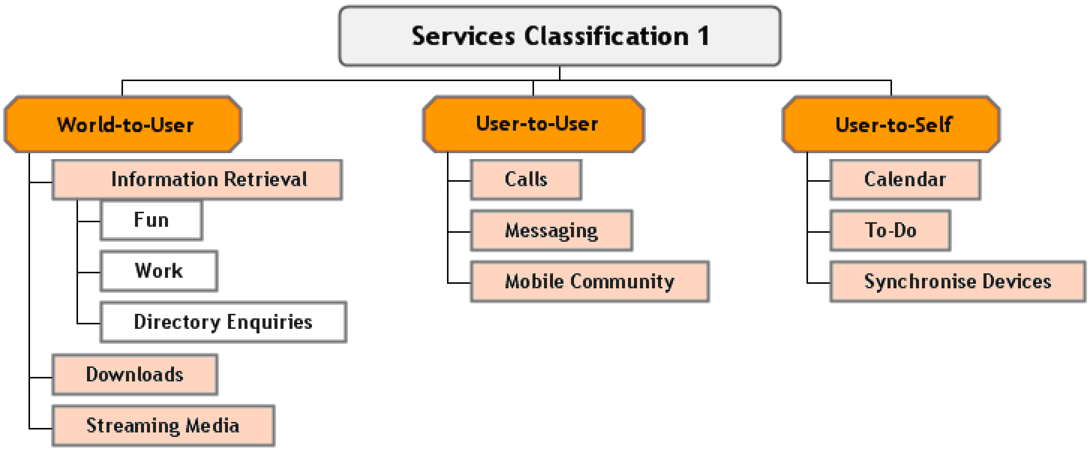


Figure 4-4: Initial attempt to mobile service classification

4.2.2 Empirical Validation and Improvement

In order to empirically validate and improve the service classification for future experiments, we conducted two card-sorting studies. The first one was 'open', as participants were asked to invent and name the categories. These categories and their hierarchies were then reviewed, aggregated and appropriately adjusted before the second card-sorting study. This was a 'closed' card sort as participants were given the categories and asked to place the services within them. This process is similar to sound categorisation studies [e.g. Marcell, 2000], where two different sets of participants perform unconstrained and constrained classifications respectively (but usually no category structure is applied). The first set creates and labels categories by listening to a set of sounds, and the second set of participants assigns the sounds to these categories that have been aggregated and appropriately amended by the researchers.

Open Card Sort

In order to see if the classification we devised in Section 4.2.1 is consistent with users' mental model, we took an empirical approach to see how they would categorise the existing services. Therefore, no categories were presented to the users and an open card-sorting method was used.

Participants

Thirty-seven computer science undergraduate students (5 female and 32 male) in the age range of 18 to 24 (M=19.9, SD 1.87) took part in the study. All but one had owned a mobile phone for a range of 2 to 8 years (M=5.49, SD 1.8), and on average they used it at least once a day for calls and text messages. Their usage of other services and applications on their phone was not as frequent: calendar reminders were used at least once per week, photographs and to-do lists at least once per month, and information retrieval services and downloads even less frequently. A pilot study was run with 12 students of similar demographics and mobile phone usage habits. The study was carried out during a scheduled lab session but no marks or other incentive was given.

Procedure

All 52 services were presented on 52 cards, with the name of the service and a short description (see Appendix IV.1). The order of the cards was randomised before handed out to the participants, who were formed in groups of 3 or 4 people. There were two reasons for grouping participants. First, creating categories for 52 services could prove to be quite daunting for one person to handle, and therefore easy to lose interest and motivation. Second, we wanted to encourage discussion between the participants so that they can collectively devise classification rationales. However, sessions were run in parallel and discussions were not supervised or recorded.

The study was piloted with three groups and the instructions were slightly amended as a result. We instructed participants to group the services together according to similarity of their respective notifications, and to name and describe the categories on the blank cards they were provided with. They were encouraged to create as many categories as they liked in hierarchies of whatever depth and breadth, but there was a requirement for strict hierarchies (i.e. services could belong only to one branch each). Clarification questions were also encouraged throughout the sessions. Once each group had created their classification, its participants filled in an individual questionnaire each, collecting demographic information and their views on the procedure. The relevant material (e.g. consent form, instructions and questionnaires) can be found in Appendix IV.

Results and Discussion

The hierarchies produced by the groups were captured and input in a cluster analysis software program[1], which calculated the relative strengths of relationships between all services. The result of cluster analysis is based on how frequently each pair of cards appears in the same group across all input hierarchies. The output is given in a form of a dendrogram, which is a tree-like graph that depicts these relationships (Figure 4-5). In the far left of the dendrogram is the root, which through a series of bifurcations ends up to the leaves (all 52 services) to the right. The earliest a bifurcation appears (i.e. further to the left), the weakest the relationship between the services that belong

---

[1] Syntagm SynCaps - http://www.syntagm.co.uk/design/cardsortdl.shtml (last accessed 30/06/2009)

in the two corresponding branches. This means that there are stronger relationships within each of the two clusters that resulted from the first bifurcation. Bifurcations that appear last (furthest to the right) signify the strongest relationship between the two leaves (i.e. services).

It is important to stress that a dendrogram does not return a specific number of clusters. It can give any number from 1 to 52 according to where a vertical line is drawn. If we draw the line just after the second bifurcation, we can see that the three resulting clusters of services fit with the rationale of 'World-to-User' (W2U), 'User-to-User' (U2U), and 'User-to-Self' (U2S) with only three exceptions.

First, the 'Postcards' service (ability to upload pictures and receive printouts by post) was grouped very closely with the 'Photography' service (ability to upload pictures for sharing purposes). Although participants identified the common underlying concept of dealing with mobile photographs, arguably they do not belong in the same category of our hierarchy, since one is U2U and the other U2S. Second, 'Online Betting' and 'Adult Content' were group very tightly together, most probably due to their age-restricted accessibility. Similarly, 'Online Account' stood alone as it was consistently found distinctively dissimilar from all other services. However, neither of these two clusters fit the rationale of our hierarchy (Figure 4-4), as they do not share the same origin or similar contents.

Overall, the results of the open card sort are in fair, but not absolute, agreement with our analytical categorisation. Participants' grouping did not directly correspond to the W2U-U2U-U2S rationale but the high level clusters they produced held almost the same services with those in our categories. The data analysis and the differences between the two classifications suggest the following improvements to our categorisation (the resulting classification is depicted on Figure 4-6):

1. We renamed the label 'Information Retrieval' to 'Information Services', and reorganised its subcategories to 'News Information', 'Sports Information' and 'Context-aware Information'. This was done to accommodate the plethora of information services (especially sports information) and organise them more closely to users' mental models.

2. We collapsed the three sub-categories of U2U in only two ('Calls' and 'Messaging') as the two 'Mobile Community' services were coupled very tightly with either the 'Messaging' services ('Location Messaging' service) or the 'Context-aware Information' services ('Friends' Location' service).

3. We collapsed the 'Calendar' and 'To-Do' services of the U2S branch in 'Self Reminders' to reflect the very tight coupling between the two services.

4. We renamed 'Synchronise Devices' of the U2S branch to 'Backup Reminders', which now includes 'Synchronise Devices' and 'Online Account' to accommodate the misunderstood 'Online Account' service.

5. We included 'Other Services' as a fourth branch in our hierarchy that can hold all the services that fit poorly to any of the other clusters. Services such as 'Post-cards', 'Photographs' and 'Online Betting' that seem to be poorly understood and dissimilar to most services, can be included in this category.

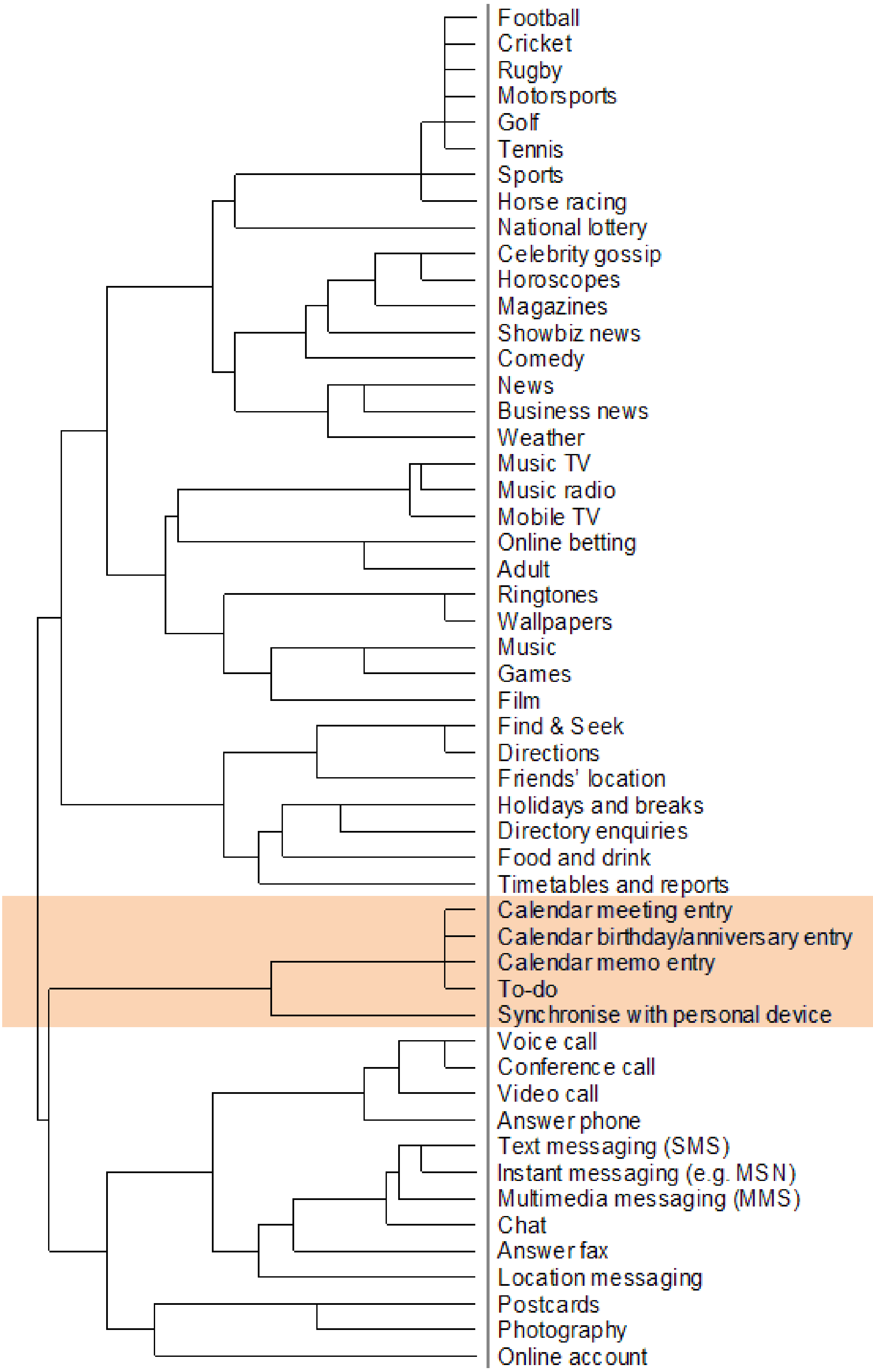


Figure 4-5: Dendrogram of open card sort

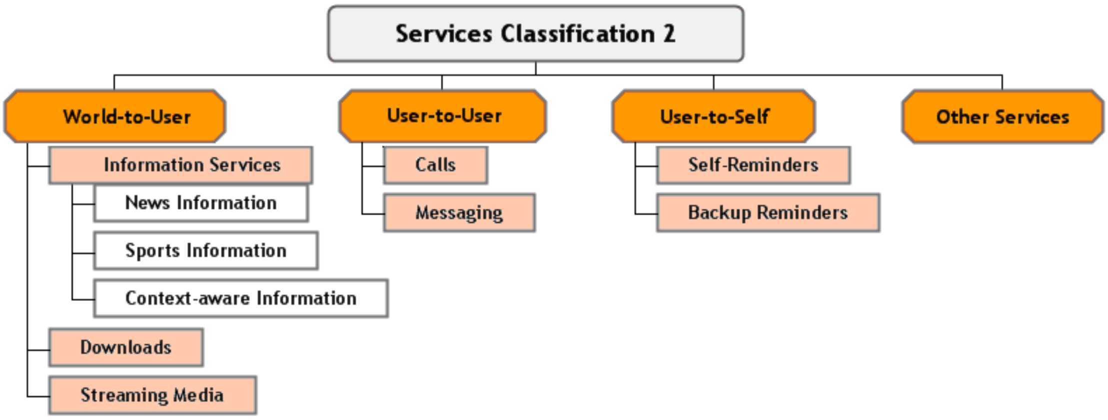


Figure 4-6: The new classification based on the analysis of the open card sort

Closed Card Sort

Our initial classification (Figure 4-4) was amended based on the analysis of the results of the open card sort. The resulting classification (Figure 4-6) still represents the origin-of-service rationale (World-to-User, User-to-User, User-to-Self) but its sub-categories have been re-organised to better reflect users' mental models. However, the new classification needed to be empirically tested and validated, to see if users indeed agreed as to which services belong to each of the sub-categories. To this end, we conducted a closed card-sorting study, which presented the users with the hierarchy and asked them to assign the services to the sub-category that was most appropriate to them.

Participants

Thirteen individuals (5 female and 8 male) in the age range of 23 to 39 (M=28.2, SD 4.38) took part in the study. Most of them were postgraduates or members of staff of the Computer Science Department. Their mobile phone usage was similar to the participants of the open card sort study: 8 years of use, daily use for calls and text messaging, considerably less for calendar reminders, to-do lists and taking photos, and almost never for information services and downloads. A pilot study was run with one individual with similar demographics. No incentive was given for participation.

Procedure

The same 52 cards with the services' names and descriptions were given to the participants (order of the cards was randomised). This time participants completed the study in consecutive sessions with one participant at the time. No groups of participants were formed, as the task would be easy and quick enough for a single participant to complete.

The study was piloted with one participant and one category label was changed as a result. As before, we instructed participants to group the services together according to similarity of their respective notifications. In addition, they were asked to tick on each card one of the boxes signifying how well the service fit in the chosen category (perfect, good, fair). The 10 category cards (along with names and descriptions) were firmly fixed on the working area in a manner that the hierarchy was represented (see

Figure 4-7). Participants were given an explanation of the categories by the experimenter. Clarification questions were encouraged before the task started but not throughout the session, during which participants were encouraged to think aloud. Once they had sorted all the services, a short interview followed in order to capture their thoughts, especially about services they were unsure during sorting. At the end, they were asked to fill in a questionnaire collecting demographic information and their views on the procedure. Material relevant to closed card-sorting (consent form, instructions and questionnaire) can be found in Appendices IV2, IV.5 and IV.6.

(a) (b)

Figure 4-7: Closed card-sorting at the beginning (a) and the end (b) of the classification task

Results and Discussion

The same cluster analysis with the open card sort was performed, with the addition of fit strength as one of the analysis parameters. This is indicated in the dendrogram by the "+" symbols appearing next to the name of each leaf (with more "+" indicating better fit of each leaf in its cluster). The dendrogram in Figure 4-8 depicts the 10 given categories with the services participants assigned to them, along with the relative strength relationships between them. The categorisation was validated with participants reaching 83% agreement on average in assigning the services to the intended categories. As predicted, but with not great strength of fit, services 'Postcards' and 'Photography' were assigned to 'Other Services', and 'Synchronise Devices' and 'Online Account' to 'Backup Reminders'. However, 'Online Betting' was assigned (with a relative poor fit) to 'Sports Information' rather than the expected 'Other Services'. Other services that demonstrated relative poor fit were 'Answer Phone' in 'Calls' and 'Weather' in 'News Information'. From the think aloud and debriefing data collected, the contesting categories for these services were 'Messaging' and 'Here & Now Information' respectively.

'Here & Now Information' demonstrated a consistently poorer fit than other categories, and participants found it difficult to grasp and relate to their own mobile phone usage. In fact, the category was renamed from 'Context-aware Information' after the pilot study, as its name was found to cause confusion (as we anticipated interpretation to vary wildly, depending on participants' familiarity with the term context-aware).

This difficulty was not surprising as 'Here & Now Information' included services currently accessed through active browsing, and were described to provide information "related to you, here and now" (e.g. 'Timetables and reports', 'Directions' and 'Friends' Location'). Results of the usage questionnaire revealed that our participants rarely (if ever) used such services. Finally, from the subjective data collected it was decided to re-label 'Streaming Media' to 'Entertainment Live', as our initial label was found to cause confusion. This change shift the focus from the technical term denoting method of delivery (i.e. streaming) to the more content-oriented, everyday wording of the new label.

As a result of the relabeling of the two categories, the five sub-categories of 'World-to-User' were organised in two intermediate categories: 'Information Services' and 'Entertainment Services'. The 10 clusters with in their final hierarchy and labels are presented in Table 4-5. This classification was carried forward and used in the notification studies presented in Chapters 5 and 6.

A	**Services from the world**	**Information Services**	1	News Information
			2	Sports Information
			3	Here & Now Information
		Entertainment Services	4	Entertainment Downloads
			5	Entertainment Live
B	**Services from other users**		6	Incoming Calls
			7	Incoming Messages
C	**Services from 'myself'**		8	Self Reminders
			9	Backup Reminders
D	**Other services**		10	Other Services

Table 4-5: Final classification of mobile services

Discussion and Limitations of Card-Sorting

We have applied analytical and empirical methods to ensure that our hierarchy of mobile services follows a memorable rationale and fits users' mental model. However, natural language categories (such as the service categories discussed here) come with some level of uncertainty with regard to category membership, in contrast to well-defined categories that have crisp boundaries [e.g. Wittgenstein, 1953]. It would therefore be unreasonable to expect that a single classification would match every user's mental model. This fuzziness between the category boundaries was underlined by the large discrepancies in strength of fitness of services in their respective clusters. For example, for some people 'Sports Information' services are more of 'Entertainment Services' rather than 'Information Services'.

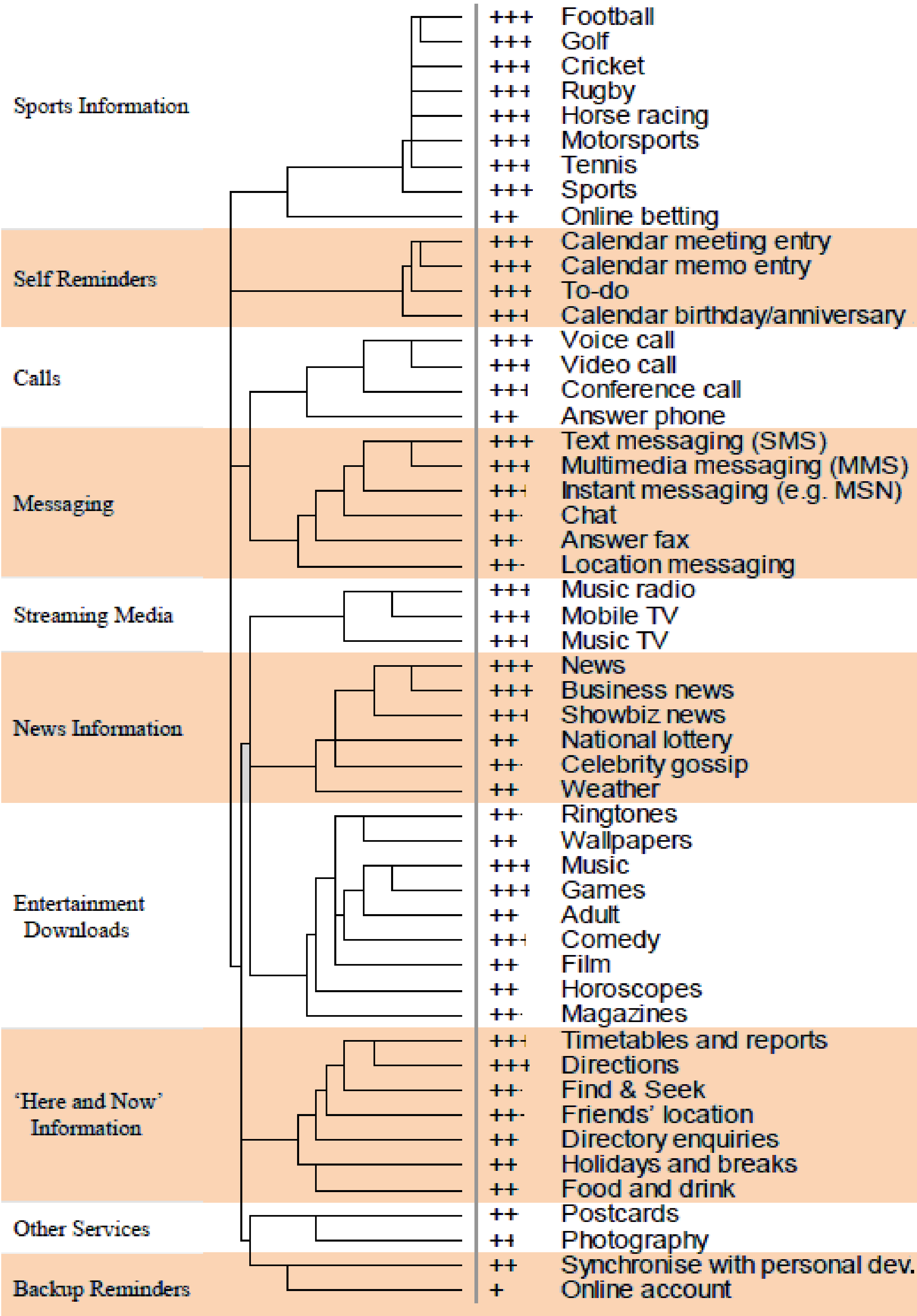


Figure 4-8: Dendrogram of closed card sort

One limitation inherent to card-sorting studies is the potential bias that can be introduced by items' labelling and descriptions. The particular wording that is chosen to represent items (and categories in the case of closed card-sorting) may imply or affect the strength of similarities between items, or between items and categories. For example, if 'Sports Information' was rephrased to 'Fun for Sports', it is likely that it would be associated more with 'Entertainment Services' and less with 'Information Services'. In the studies reported here, we did not explicitly try to avoid or create such a bias. However, we argue that rewording should be used to create or strengthen bias when necessary. It is more important that users understand and remember which services belong to each cluster, rather than worry about some classification inaccuracy on a linguistic level. Therefore, rewording of services with low fit scores should be considered to strengthen the linguistic link with their categories and reduce any uncertainties in the users' mental models of clustering (as long as the new label still accurately reflects the content of the service).

Furthermore, there is strong evidence suggesting that natural language classifications are also influenced by the context [Labov, 1973]. This means that users could classify a service in different categories depending on their current activity and situation. For example, downloading a music track in certain contexts can be part of somebody's work and therefore not classified as 'Entertainment Download'. In addition, categories that involve human activities are often derived as we make plans to achieve certain goals [Barsalou, 1991]. Examples of goal-driven categories are "foods to eat on a diet", "clothing to wear while house painting", or "activities to do on vacation in Japan with one's grandmother" [Barsalou, 1991]. Furthermore, many of these categories are not established in people's memories but are created or amended in an ad hoc manner to achieve a novel goal [Barsalou, 1991].

It is therefore believed that any proposed classification of services should only be suggested as a 'default' setting, and that users should be able to personalise it to their exact preferences. For example, they could classify 'Football' and 'Horoscope' services between information and entertainment categories in a way that it reflects their personality, habits or usual goals. Similarly, they can place under 'Other Services' any services they want to be notified about with a rather generic notification. Our classification (like any other natural language classification) is unable to tackle with the more dynamic nature of ad-hoc and goal-driven categories. It is necessary that relationships between categories and services be established in memory so that users are able to identify, learn and remember the relationships between services and notifications. This remains as an open question for researchers in notification systems and is set outside the scope of this thesis.

4.3 Chapter Summary

In this chapter we have presented the first systematic approach towards the development of a design process for effective auditory mobile service notifications. In particular, we have addressed two challenges unique to this domain: the need to design intuitive and pleasant notifications, and the need to classify services in a meaningful and manageable manner.

For the first challenge, we designed and performed the first experimental investigation comparing speech, auditory icon and earcon sounds as mobile service notifications. In this experiment we tested the intuitiveness of the different audio notifications by asking users to guess the audio-service links, without any prior training to the experimental stimuli. This was done as an initial response to the 'research gap' (also identified by [Simpson, 2007]) between sound-signal intuitiveness studies and notification type effectiveness studies. Our first experimental hypothesis was supported, with auditory icons performing significantly better than earcons in recognition accuracy. Speech not surprisingly outperformed both non-speech sound types. However, our second hypothesis was not supported, as the response speed between the non-speech notifications was not significantly different. User preference and perceived ease of learning of the notifications followed the intuitiveness trend, with speech preferred over auditory icons, which in turn were preferred over earcons. However, answers on choice of sound type for actual potential use by users were contradictory, and therefore preference data were overall inconclusive.

We argue that the success of auditory icons over earcons heavily depends on the success of the metaphors used. Regardless of sound type, associations were more successfully guessed when the mapping between service and notification was more iconic (e.g. auditory icon for 'Television' service and earcon for 'Messaging' service). This suggests that pre-existing semantics should be utilised for intuitive mobile service notifications, just as literature suggests for ease of learning.

There were a number of limitations in relation to the experimental process and the implementation. There was no particular effort to compare sounds that were the most efficient or representative of each type, and the hierarchical earcons were effectively tested only in their capacity as musical sounds, since no information on earcons or services structure was provided. Furthermore, response time records were compromised due to occasional software lags, but not limited to a single type of sound.

Despite these limitations, this experiment achieved two goals. First, it demonstrated the importance of signal-referent directness in intuitiveness. Second, it gave us experience in designing experiments and experimental sounds that can contribute towards the development of design guidelines for auditory mobile service notifications. These lessons guide the more substantial and extended studies presented in the following chapters of the thesis. Finally, it stressed the need of meaningful service classification, which can reduce the numbers of notifications to be learned.

This classification was initially designed through an analytical, systematic approach. Drawing from the relevant literature and the current classifications of network operators, we produced a classification that reflected the three dimensional origin-of-service rationale: 'World-to-User', 'User-to-User' and 'User-to-Self' services. The sub-categorisation of this classification was redesigned according to the empirical results gathered from an open card-sorting study. The resulting categorisation was validated and further improved by a closed card-sorting study. The final classification is used in the studies presented in Chapters 5 and 6.

Chapter 5

Designing Auditory Icons and Earcons for Mobile Service Awareness

There are many challenges that need to be addressed in our attempt to develop a design process for auditory mobile service notifications. In the previous chapter we discussed the challenges that the domain of mobile services imposes on this process. In this chapter we focus on the design challenges arising from the nature of the auditory stimuli. In particular, the procedure of evaluating factors of effectiveness between types of sounds can be highly subjective, as it depends on the particular sets of stimuli tested and their representativeness for each type.

Here, and for the rest of the thesis, we investigate the design process for effective non-speech notifications only. The elimination of speech notifications was decided for three reasons. First, speech notifications present no challenges in terms of intuitiveness, learnability and memorability. In our previous experiment (Section 4.1) and relevant literature (see Table 3-2), speech notifications have performed significantly better in terms of intuitiveness and learnability, as the links between them and their referents are more direct and immediate than with the other sound types. Second, results on user preference and appropriateness for auditory icons and earcons are more mixed, if not inconclusive. It seems that individual preferences, nature of referents and context of use significantly affect users' preferred choice between them. Furthermore, results have been more mixed when the notifications are on mobile devices and used in social contexts [McGee-Lennon et al., 2007]. Third, due to time constraints we decided to carry out more in-depth comparison between non-speech notifications, rather than shallower comparison between speech and non-speech notifications.

Since the literature has design guidelines to suggest for both auditory icons [Mynatt, 1994] and earcons [Brewster et al., 1993], one might assume that by following them we will have answered the subjectivity involved with non-speech auditory mobile service notifications design. However, this assumption would be unfounded for three different reasons. First, although the effectiveness of the guidelines for designing earcons has been established through iterative design and experiments [e.g. Brewster et al., 1993], the guidelines for auditory icons have not been experimentally tested. Second, the effectiveness of the two sound types has not been tested in the domain of mobile service notifications. Therefore, our design process of the stimuli has to take into account the particular challenges of this domain, such as need for minimal training and dynamic context of use. Third, no studies have compared sets of auditory icons and earcons that have both been designed following the respective guidelines and/or experimentally validated.

Therefore, we need to adjust the suggested design processes to match the particular challenges of the domain, and test the (relative) effectiveness of the sounds produced through these processes. This test (presented in Chapter 6) will give results not only on the comparative effectiveness of the sound types, but could also evaluate the effectiveness of the processes described in this chapter. Next, we present the auditory icon mobile service notification design process, followed by the respective earcon design process.

5.1 Designing Auditory Icons

In order for auditory icons to be intuitive mobile service notifications, two steps are needed. First, we need to ensure that most users hear the same thing when listening to the sound. In other words, the events or actors causing each sound should be easily identifiable for all participants. For example, the sound of frying food can be confused with raindrops hitting the ground [Marcell et al., 2000]. The measure of distinctiveness of such sounds is known as 'causal uncertainty' and depends at least on three things: the quality of the recording, the uniqueness of their properties (e.g. frequency, onset, and temporal patterns) and the variety of events causing the same/similar sound. In fact, causal uncertainty has been systematically exploited by 'Foley artists' in film production, who produce sounds in a recording studio (e.g. 'crackling plastic wrapping material') to then include in the soundtrack as sound-alike everyday sounds (e.g. 'wood fire crackling').

The second step towards intuitive auditory icons is related to the choice of successful, direct associations between the sounds and the services. In other words, even if most users hear the same thing (e.g. frying food and not rain), do they associate it with the same service? Directness of the signal-referent associations has been found to create more intuitive (see experiment in Section 4.1) and more learnable (see literature review in Section 3.4.2) auditory icons.

Identifiability and conceptual mappings are two steps that are included in the design methodology suggested by [Mynatt, 1994]. However, we have been unable to find any studies with auditory icons that have performed both steps before they evaluate or compare their effectiveness as notifications. Most studies utilise mappings that are solely based on researchers' subjective choice of metaphor. In one exception where end-users were involved in this stage, it was poorly reported and conducted in a non-systematic manner ("…the [sound] regarded by the majority of [participants] as being the most suitable for each event was chosen") [Leung et al., 1997].

Next, we present the procedure we followed to ensure auditory icon identifiability and service representativeness. End-users participated in the design process via two online surveys that addressed each one of these requirements. The classification of mobile services produced in Section 4.2 (see Table 4-5) was used here to provide the referents of the sounds.

5.1.1 Identification of Auditory Icons

Background

Although sound identification has been suggested as a necessary step in auditory icon design methodologies [Mynatt, 1994] (see Section 3.2.2), it has not been taken up by other similar design procedures (e.g. "natural warnings sounds" design procedure suggested by [Ulfengren, 2003]). In fact, we were unable to find any studies performing this step before they evaluate or compare effectiveness of auditory icons. However, there is some research in everyday sound identification and classification (e.g. [Ballas, 1994; Marcell et al., 2000; Marcell et al., 2007; Vanderveer, 1979]) that informs other research areas (e.g. cognitive science, and diagnosis of medical conditions). In this literature participants are usually asked to identify sounds by describing each stimulus in a short free-text response. Depending on the nature of the research, secondary rating data are gathered about the sounds, such as confidence of response, perceived urgency and pleasantness. Researchers then analyse the responses and decide if they were accurate or not. In case of classification studies, they group together categories that present different wording but essentially describe the same thing, and then may ask a new set of participants to assign the sounds to these categories.

Sound identification can be a highly subjective procedure as it depends on researchers' interpretation of the responses and their accuracy under the light of the particular research. For example, "breaking glass" and "breaking a bottle" could be equally accurate responses if the material manipulation is important, but not if the particular object manipulation is more important. An attempt to standardise this procedure is presented in [Marcell et al., 2000], where a set of guidelines for scoring name responses is given. Amongst other rules, they suggest that responses should be regarded as correct if they use the correct grammatical root (but different ending), if they are obvious misspellings or synonyms. However, subjectivity is introduced in the guidelines as they regard as correct responses that accurately capture the "conceptual nature of the sound", or provide "an unanticipated description that was subsequently judged as an acoustically precise alternative interpretation". Although subjectivity is somewhat reduced by using independent judges, we find that the above guideline is inappropriate for measuring auditory icon notifications' intuitiveness.

A more formal approach for calculating causal uncertainty is followed by [Ballas, 1993], who applied a mathematical formula previously used in picture identification. This formula takes again into account the number of times independent judges assign sounds to categories. This quantification of causal uncertainty gives the opportunity to relate it to other measures, such as identification accuracy, response times [Ballas, 1993], and confidence of responses [Ballas & Howard, 1987, as reported by Ballas, 1993].

Sounds Design and Acquisition

Before we test the identifiability of auditory icons, we needed a pool of candidate sounds. Using the 10 clusters of services (Table 4-5), 4 HCI researchers took part in a brainstorming session where everyday sound-producing events were assigned to the

10 clusters (note that for reasons of simplicity and immediacy, the clusters of services are usually referred to as 'services'). The researchers were presented with the classification of services, and descriptions and examples of the services were given. Then, each researcher individually produced as many ideas of sound-producing events as they could for each service. All ideas were aggregated, followed by discussion on their similarities and appropriateness to represent each of the services. This resulted in 3 to 7 candidate ideas for each one of the 10 services.

Multiple sound instances for all events were collected from royalty free websites or recorded by the author (132 in total). They were all sampled at 44.1KHz with 192Kbits bit rate, and their volume was normalised at -89db. After carefully listening and comparing all of them, the hardest to identify instances were discarded. In cases where identifiability between instances was not obviously different, two independent naïve listeners were asked to help with the identification process. This resulted in a total of 47 sounds (Table 5-1), which were put through an online survey to measure and empirically validate their identifiability.

Survey 1: "Source that Sound!"

The method of online survey was chosen for data collection as it demonstrates a number of advantages. First, it is very efficient in reaching large numbers of people, with virtually no extra costs and in small amount of time. Second, it can easily reach a worldwide, cross-cultural audience, which should be the target of usability studies of widespread devices targeted for the general public. This should assist us to find out if any cultural elements interact with sound identifiability. Third, we took advantage of the flexibility that PHP offers in designing dynamic HTML pages, and presented the stimuli with the appropriate pseudorandom order to avoid learning and fatigue effects.

Convenience sampling was used and respondents were mainly recruited through email, utilising mailing lists of university departments, research communities (with special interests in HCI and/or auditory interfaces), friends and colleagues. Also, the university virtual notice board and a social networking tool (Facebook[1]) were utilised.

Design and Implementation

The survey was built in HTML and PHP. On the first page, some basic information about the task to follow were provided, including estimated time of completion (approximately 30 minutes) and respondents ability to quit at any point. Also, some demographic information was gathered: gender, age, nationality, country of residence and number of years residing there. If respondents were residing in a country other than the UK, they were asked to indicate if they had ever lived there before and for how long. The nationality and residence information was gathered in order to examine if there is any correlation between identifiability and familiarity to the UK soundscape and cultural environment. Also, on the first page of the survey respondents were asked to confirm that they had no hearing difficulties and that they were able to

[1] www.facebook.com

play and hear the auditory instructions presented to them. All fields were required in order to proceed to the survey and they were checked for empty or invalid values (see Appendix V.1 for screenshots of Survey 1).

All 47 sounds were presented only once for each respondent (one per page) in a pseudorandom order, taking special care so that sounds representing the same services or sounds sounding similar were not presented back to back. On the top of each page the same simple instructions were repeated, and respondents were expected to play (and repeat if necessary) the sound on each page using the embedded Flash player. They were asked to identify each sound by responding to the same three questions:

1. "Write *one* short but detailed description of the *most probable* event/situation that you think produced this sound. (E.g. a heavy wooden door slamming in the wind)".

2. "How many *distinct* events/situations could you describe that would produce *exactly the same* sound? (E.g. if a sound sounds like "a wooden door slamming" and like "someone punching a table" and like "a gunshot", then you should enter '3').

3. "Identifying this sound was Very Easy/Easy/Neutral/Hard/Very Hard".

For question 1, a 1-line text box was provided for a free text response. It has been found that descriptions of abstract perceptual properties of the sounds (rather than the event that caused them) are given when the sound source is not recognised [Vanderveer, 1979]. Therefore, we were very careful with the phrasing of the question, so that respondents had a clear understanding of what they were asked to do. The question was revised a few times through iterative design and testing with pilot users.

For question 2, a drop-down menu with options from '1' to '10+' was provided. This question was included because it has been found that the number of alternative responses given in everyday sound identification tasks is significantly correlated to identification accuracy [Ballas, 1994]. We wanted to have a second measure for causal uncertainty to use if the free text response was not filled in. In addition, the effectiveness of the two measures could be compared from the results.

For question 3, a radio button with five options ranging from 'Very Easy' to 'Very Hard' was provided. Similarly with the number of sources measure, it has been found that confidence of response and perceived difficulty are significantly positively correlated to identification accuracy [Ballas, 1994; Marcell et al., 2000].

The survey was designed so that respondents could quit at any time, without losing any data they have provided until then. When all sounds have been presented or when respondents opted to quit, a thank you message and the opportunity to leave comments were given. Also, the final page displayed a description of each sound that was presented against the description provided by the respondent. This was done only to give respondents the satisfaction usually achieved at the end of a game, although no scores were given. The site was designed to use browser cookies so that the same individual could not repeat the survey (assuming that they did not delete the cookies on their computer or access the site from a different computer).

Results and discussion

There were 179 individuals initiating and (at least) partly completing Survey 1, 98 males and 81 females with mean age of 32.3 (SD=10.2). They were of 34 distinct nationalities, residing in 21 different countries. From all the responses, 32% came from Greeks and 31% from British nationals, but only 23% came from people residing in Greece and 54% from people residing in the UK. The rest of nationalities and countries of residence were of much smaller percentages. Thirty-five per cent had never lived in the UK but as many as 34% had been in the UK for at least 10 years. Since only few of the participants responded to all 47 sounds, we gathered 81.5 responses on average for each sound.

The free text descriptions of question 1 were assessed against the real sound producing events, resulting in a percentage of successful identification for each sound. On average, identification accuracy was at 75%, which is comparable to other studies (81% and 74% in [Marcell et al., 2000] and 78% in [Marcell et al., 2007]). Forty-five per cent of the sounds were identified with over 90% accuracy (compared to 52% in [Marcell et al., 2000] and 50% in [Marcell et al., 2007]), and only 11% of the sounds scored under 50% (compared to 15% in [Marcell et al., 2000, Marcell et al., 2007]). The identification accuracy for all 47 sounds appears on Table 5-1.

SERVICE	SOUND DESCRIPTION	ID (%)
.1. NEWS INFORMATION	Typing on an old-fashioned typewriter	99
	BBC news ident	77
	ITV news ident	74
	Channel 4 news ident	65
	Flicking through the pages of a newspaper	57
.2. SPORTS INFORMATION	Tennis game sounds + crowd applause	99
	People cheering and blowing a horn at sporting event	96
	Basketball game sounds	94
	Referee whistle	90
.3. HERE & NOW INFORMATION	Public announcement at an airport	97
	Restaurant ambient noise (discussions, cutlery clattering)	92
	Pub ambient noise (playing pool, discussions, music)	91
	Caricature of 'hello' sound (Worms game)	83
	Train platform attention chime + background train noise	68
	Service bell (e.g. at a hotel front desk)	55
	Town crier (of London)	41

.4. ENTERTAINMENT DOWNLOADS	Rock concert sounds (electric guitar, crowd cheering)	92
	Tuning into a radio station (analogue scanning)	92
	Clip played at the beginning of a movie (20th Century Fox)	91
	Beginning of 'EastEnders' intro audio clip (UK TV series)	51
	Data transfer (B-movie style)	10
.5. ENTERTAINMENT LIVE	Audience laughing and clapping	99
	Guys laughing	99
	Audience clapping/applauding	88
.6. INCOMING CALLS	Old-fashioned phone ringing	100
	Telephone ringing as heard through the telephone line	96
	Typical mobile phone ringtone (Nokia)	81
.7. INCOMING MESSAGES	Pigeon flying off	99
	Typing on computer keyboard	91
	Message transmitted in Morse code	84
	Texting on mobile phone keypad	72
	Typical Nokia incoming SMS sound	57
	Post arriving through door letterbox	4
.8. SELF REMINDERS	'Wolf' whistling (typically to draw attention)	96
	"Ahem" sound - clearing throat	94
	Typical alarm clock	83
	Generic audio notification on mobile devices (Windows Mobile)	64
	Clicking fingers + "aha" sound ("eureka moment")	58
	Human whistling	55
.9. BACKUP REMINDERS	Modem or fax connecting	100
	Truck/lorry reversing (backing up)	79
	Caricature of 'oh-oh' sound (ICQ)	41
	VCR tape being inserted and rewind	5
.10. OTHER SERVICES	Church bells (UK)	95
	Wind chimes	79
	'Huh?' - sounding surprised and confused	68
	Taking a picture with an old camera (shutter + film wind)	58

Table 5-1: Identification scores of sounds in Survey 1

The identification process was done following most of the guidelines suggested in [Marcell et al., 2000]. For example, descriptions that contained synonyms, extraneous information or obvious misspellings were considered correct as long as they captured the essence of the sound. For example, the description "when I tell someone in detail what my job is" although playful, it captures the essence of 'confusion' that the sound intended to convey. Responses that were literal (e.g. "someone saying 'huh'") or too generic (e.g. "electronic sound") were not considered to have captured the essence of the sound ('confusion' and 'alarm' respectively). However, we did not follow the guideline according to which a response that provides a description judged to be "acoustically precise alternative interpretation" should be regarded as correct. In particular, this guideline was not followed when these alternative interpretations did not reflect the metaphors that were (unknowingly to the survey participant) utilised to link the sounds to services. For example, some respondents described the 'service bell' sound as 'bicycle bell', which is an acoustically precise description of the sound, but it demonstrates that they do not perceive its potential property of 'helpdesk assistance'. On the other hand, the church bells were meant to link to 'Other Services' via the randomness of their sound (an admittedly poor metaphor), so responses describing any kind of bells were marked as correct. Finally, although extra care was taken to phrase question 1 so that respondents focused on the source of the sounds, some responses were still descriptions of the stimuli properties (e.g. "square wave lfo and then a vocoded voice"). These responses were marked as incorrect, since there is evidence that this kind of description means failure to identify the source of sound [Vanderveer, 1979].

There was no significant effect between identification accuracy of question 1 and number of alternative causes estimated by respondents in question 2 ($z=-.86$, ns, $r=.014$), or between number of alternative causes and self-reported difficulty of question 3 ($r=.045$, $p>.001$). It seems that the metric of question 2 was unable to produce any meaningful or reliable results in relation to identification accuracy. In fact, a few respondents commented on the difficulty they had with answering this question (and challenged its usefulness). However, there was a significant effect between the self-reported difficulty and identification accuracy ($\chi^2=552.12$, $p<.001$), which is in agreement with the accuracy and confidence correlation confidence of responses [Ballas & Howard, 1987, as reported by Ballas, 1994]. Finally, there was an effect on identification accuracy by nationality and country of residence across all responses. Both British nationals and UK residents answered significantly more correctly than Greeks nationals ($\chi^2=23.15$, $p<.001$) and Greece residents ($\chi^2=19.2$, $p<.001$). However, as expected, sounds culturally specific to the UK (related to UK TV news or shows) were mostly responsible for this significance (chi-squared test repeated without these sounds failed to reach significance of $p=.001$).

One of the major limitations of Survey 1 is that, due to resource constraints, the identification accuracy was judged only by the author. Although relevant guidelines were considered and applied were appropriate, it is common for similar identification studies to employ two or more independent judges [e.g. Ballas, 1994; Marcell et al., 2000]. This does not only reduce subjectivity in accuracy estimations, but more importantly encourages discussions when large discrepancies occur. In hindsight, we have noticed one example were multiple judges could possible have contributed towards more reliable accuracy results. Descriptions for the sound of "a pigeon flying off" were considered correct even if they were not specific to the type of bird. This obviously was not capturing the essence of the sound, which was related to the traditional concept of homing pigeons. If this stricter filter were used, the identification score for this sound would drop from 99% to 42%.

5.1.2 Representativeness of Auditory Icons

Background

Although the biggest advantage of auditory icons is the meaningful associations between sounds and interface functions or events, there has been very little systematic empirical research to this end. When users are not taught the associations chosen by the researchers, they often fail to understand them. Evidence for this comes from our experiment (Section 4.1) and other observational field studies [e.g. Cohen, 1994]. However, since auditory icons are most commonly taught before tested, it is difficult to evaluate the appropriateness or success of the associations. We are aware of only two research streams that end-users have been reported to participate in the association process of auditory icons, both in a military aircraft/rotorcraft cockpit context [Gardner & Simpson, 1992; Leung et al., 1997; Simpson & Gardner, 1998, as reported by Simpson, 2007]. Both of the research teams report that pilots were asked to choose the most appropriate sounds for alerts from a pool of candidate sounds, but only the Gardner and Simpson work reports a systematic approach. According to their "soundimagery" methodology, untrained representative end-users (in this case pilots) were also asked to rate the appropriateness of the sounds they chose for each alert. The associations that were selected more frequently and/or associations that achieved appropriateness rating higher than the overall average were suggested to be promoted for commercial use or further testing.

Sounds Design and Acquisition

The most identifiable sounds from Survey 1 were selected for the representativeness evaluation. This way, we aimed to eliminate poor associability scores of the final auditory icons due to users mishearing the sound (rather than just misunderstanding the metaphor). Instead of choosing a threshold of identifiability below which sounds of Survey 1 would be deemed inappropriate, we selected the three most identifiable sounds that represented each service. This way, the least identifiable sounds for each service were eliminated, and the number of alternatives for each service was controlled. Exceptions to the rule were applied when two or more sounds for a particular category were highly identified and utilised the same metaphor (e.g. TV news ident by different channels). In these cases, the next most identifiable response of a different metaphor was chosen instead. Another exception was due to an overlook and the 'Windows Mobile notification' was selected although it achieved the fourth highest score in its category. This method of selection ensured certain variety in identification scores (and our oversight contributed in that direction even more), which gave us the opportunity to test the auditory icon design process. We were interested to see if there would be any correlation between the identifiability scores of Survey 1, the representativeness (measured in Survey 2), and the efficiency of auditory icons (presented in the next set of studies in Chapter 6). The least identifiable sound that was carried forward to the next survey was successfully identified by 64% of the respondents (the sounds selected appear in Table 5-2).

Survey 2: "Mobiles beyond 'beeps'!"

An online survey was chosen again as the method of testing because of its cost-effective nature, flexibility of stimulus presentation, and ability to access across-culture audience. The same recruiting process was followed through emails and social networking tools, plus an advertisement on the University of Bath homepage. There was no effort to avoid reaching the same respondents with Survey 1. People responding to both surveys would have no advantage over others since Survey 1 respondents were not aware of the goals of our research and were not informed of the services.

Design and Implementation

The survey was built in HTML and PHP. On the first page, basic information was given about the task to follow, the estimated time of completion (approximately 10 minutes) and respondents' ability to quit at any time. The high level goals of the research were described as "not to create more notifications and advertising, but merely to make audio notifications more meaningful for those who want them". The same demographic information as in Survey 1 was also gathered on the first page (see Appendix V.1 for screenshots of Survey 2).

Each one of the following 10 pages displayed the same instructions on their upper part, and one of the 10 clusters of services from our classification (Table 4-5), along with the cluster description and instances of services of that type (see Appendix V.2). Pages appeared in a random order for each respondent, and following feedback from Survey 1, a progress indication was added. On each page, the three most identifiable sounds of each service (results from Survey 1) were embedded in three Flash players, which were displayed simultaneously, but in a randomised order every time. Next to each player we gave the description of the corresponding sound. Having ensured the sounds' identifiability was relatively good, we provided the descriptions to avoid errors in the association data by misheard sounds. Furthermore, this ensured that there was no advantage for the respondents who had been exposed to the sounds during Survey 1. Respondents were asked to initiate each sound (and repeat if needed), and respond to the same three questions on each page:

1. "If you had to choose one, which of the sounds below do you think best represents this category of services?"
2. "For the sound you chose above, how representative do you think it is for this category of services?"
3. "Any comments so far?"

For question 1, three radio buttons were provided next to the Flash players. Question 2 provided with five radio button options laid out on a spectrum from 'Not Representative' to 'Representative'. For question 3, a 2-line text box was provided for free text entry. When all 10 services were presented, participants were thanked and given the opportunity to enter any overall comments. The same individuals would not be able to repeat the survey from the same computer, unless they deleted their browser cookies.

Results and discussion

There were 133 individuals initiating and (at least) partly completing Survey 2, 67 males and 66 females with mean age of 31.8 (SD=10.5). They were of 28 distinct nationalities, residing in 14 countries, making results of Survey 2 slightly less international than Survey 1. From all the responses 45% came from British and 13% from Greek nationals, but only 8% came from people residing in Greece and 65% from people residing in the UK. The rest of the nationalities and countries of residence were of smaller percentages. Thirty-three per cent had never lived in the UK but as many as 41% had been in the UK for at least 10 years. Since some of them did not complete all 10 pages, we collected 112.2 responses per service on average (with a minimum of 110 and a maximum of 115).

In terms of sound selection (question 1), there was a wide range of scores with some sounds being selected by as few as 1.7% of the respondents, and others by as many as 85.2%. It is worth noting that these numbers have only a relative value against the other two sounds presented for each category, rather than an absolute value against all 30 sounds presented in Survey 2. For example, 'old-fashioned phone ringing' was chosen by only 39.8% of the respondents because of the high competition it had from the very similar 'telephone ringing as heard through the telephone line'. Representativeness scores for all 30 sounds of Survey 2 are presented in Table 5-2, where numbers in bold indicate the final 10 sounds selected to represent the 10 services in the studies presented in Chapter 6.

In terms of representativeness ratings (question 2) of the sounds respondents chose, the means per sound ranged from 2.5 to 4.5 on the 1-5 Likert scale, with an average of 3.5 for all sounds (SD=0.6). This suggests that participants tended to rate the sounds as fairly representative once they had selected them as the most representative amongst the three sounds. There was no strong correlation between the frequency of chosen associations and the ratings of representativeness ($r=.14$, $p<.001$). In fact, four of the associations that won the within-cluster competition, failed to pass the average of the rating across all sounds, criterion suggested by [Simpson, 2007]. For example, 'wind chimes' had the lowest average in terms of representativeness for 'Other Services', although it had a frequency rate of 58% and the two other competing sounds had 25% and 14%. This indicates that all three of the suggested metaphors for this service were weak.

There were also many interesting comments from respondents. Some pointed out that none of the sound options given were really representative of the corresponding services, and they had to choose the "least irrelevant". They also mentioned how their favourite or most pleasant was not always the most representative and found difficultly in choosing between these two metrics. Therefore, some concluded they would prefer to be given the possibility to nominate their own sounds for each service, as their associations would make more sense to them. Pleasantness seemed to be connected to their acoustic preferences as well as their perception of annoyance notifications could cause to people around them. Some sounds were reported as "disturbing", "irritating", "embarrassing" or just "unpleasant". Some auditory icons were also found to be too long (another factor contributing to embarrassment when in public), and possibly difficult to differentiate from environmental sounds (especially for the auditory icons with background noise). Finally, a couple of the respondents pointed out that some auditory icons might be of limited use due to cultural bias.

Service	Description	ID %	REP %
.1. News Information	**BBC news ident**	**77**	**85**
	Typing on an old-fashioned typewriter	99	10
	Flicking through the pages of a newspaper	57	2
	Channel 4 news ident	65	--
	ITV news ident	74	--
.2. Sports Information	**People cheering and blowing a horn at sporting event**	**96**	**69**
	Tennis game sounds + crowd applause	99	23
	Basketball game sounds	94	5
	Referee whistle	90	--
.3. Here & Now Information	**Public announcement at an airport**	**97**	**67**
	Caricature of 'hello' sound (Worms game)	83	17
	Restaurant ambient noise (discussions, cutlery clattering)	92	13
	Pub ambient noise (playing pool, discussions, music)	91	--
	Service bell (e.g. at a hotel front desk)	55	--
	Town crier (of London)	41	--
	Train platform attention chime + background train noise	68	--
.4. Entertainment Downloads	**Clip played at the beginning of a movie (20th Century Fox)**	**91**	**58**
	Rock concert sounds (electric guitar, crowd cheering)	92	26
	Tuning into a radio station (analogue scanning)	92	14
	Data transfer (B-movie style)	10	--
	Beginning of 'EastEnders' intro audio clip (UK TV series)	51	--
.5. Entertainment Live	**Audience clapping/applauding**	**88**	**65**
	Audience laughing and clapping	99	21
	Guys laughing	99	11
.6. Incoming Calls	**Old-fashioned phone ringing**	**100**	**40**
	Telephone ringing as heard through the telephone line	96	40
	Typical mobile phone ringtone (Nokia)	81	16
.7. Incoming Messages	**Message transmitted in Morse code**	**84**	**61**
	Pigeon flying off	99	28
	Typing on computer keyboard	91	9
	Post arriving through door letterbox	4	--
	Texting on mobile phone keypad	72	--
	Typical Nokia incoming SMS sound	57	--

.8. Self Reminders	**Generic audio notification on mobile devices (Windows Mobile)**	**64**	**45**
	'Wolf' whistling (typically to draw attention)	96	30
	"Ahem" sound - clearing throat	94	23
	Typical alarm clock	83	--
	Clicking fingers + "aha" sound ("eureka moment")	58	--
	Human whistling	55	--
.9. Backup Reminders	**Truck/lorry reversing (backing up)**	**79**	**36**
	Caricature of 'oh-oh' sound (ICQ)	41	32
	Modem or fax connecting	100	30
	VCR tape being inserted and rewind	5	--
.10. Other Services	**Wind chimes**	**79**	**58**
	'Huh?' - sounding surprised and confused	68	25
	Church bells (UK)	95	14
	Taking a picture with an old camera (shutter + film wind)	58	--

Table 5-2: Identification scores of sounds in Survey 1, and representativeness scores of sounds in Survey 2. Bold font indicates the final 10 sounds selected to represent the 10 services.

5.1.3 Discussion of Auditory Icon Design Process

We are aware of two auditory icon design processes suggested in the literature. The first [Mynatt, 1994], lists both evaluation steps of sound identification and sound-referent mapping. The second [Ulfvengren, 2003], suggests a more detailed methodology and breaks down the sound-referent mapping process in two steps. A combination of soundimagery and associability studies is suggested, but the process lacks a sound identification step. The process presented here, constitutes the first attempt to unify the two literature processes into one detailed methodology that produces auditory icons that are both easily identifiable by most people, and they have meaningful mapping to their referents. Survey 1 ensured sound identifiability (as suggested by [Mynatt, 1994]) and Survey 2 identified, amongst multiple sounds, those that are more directly associated with the interface functions or events (equivalent of [Ulfvengren, 2003]'s soundimagery step). The third step of this process will be to test the auditory icons selected here, to ensure that the associations between them and the functions are easy to identify and learn (learnability test would be equivalent to [Ulfvengren, 2003]'s associability step). This step is completed in Chapter 6, where the intuitiveness, learnability, memorability and longitudinal user preference of the auditory icons is measured and compared to the earcons', which are designed in the next section (Section 5.2).

Because our auditory icon design procedure had a focus on mobile services, we decided to make adjustments to the identification (Survey 1) and representativeness (Survey 2) methods from what is suggested in the literature. Everyday sound identification processes can be established by asking participants what they hear and then judging the accuracy of their responses. Some researchers ask participants to describe what produced the sound [e.g. Ballas, 1994] while others just ask them to describe what they hear [e.g. Marcell et al., 2000]. We opted for the former approach, as the events that produced the sounds are significant in auditory icon design. It is usually the meaning of these events that relate to the services, rather than the properties of the stimuli. For example, "electronic sound" is not capturing the alarming property of the sound. If users were not prompted by the sound to think of the event behind it, we would not know if their descriptions were lacking the conceptual reference because they did not 'hear' it or because they misunderstood the task. Furthermore, another way accuracy of sound identifiability can be judged is by classifying participants' responses into conceptually distinct categories and then quantifying the causal uncertainty [e.g. Ballas, 1994]. This approach is adapted from picture identification studies and does not necessary reflect the needs of auditory icons design. The membership of the responses in the categories created by the judges are not the best indication of whether the participant has identified the property of the sound-producing event that will be later on used as a metaphor to relate it to a service. For example, descriptions that include terms of "confusion", "surprise", "ignorance" or phrases as "when I tell someone in detail what my job is" are likely to be assigned in different categories, although they all capture the metaphor that relates this sound to 'Other Services'. This is why in our identification scoring process we followed guidelines that regard as correct descriptions "that accurately captured the meaning of the sound source or the conceptual nature of the sound" [Marcell et al., 2000]. However, we regarded this essence as correct, only if it was related to the metaphor utilised for the particular auditory icon. Therefore, the guideline that accepts descriptions that were "judged as an acoustically precise alternative interpretation" [ibid] was not followed.

Our approach for measuring representativeness of sounds for services through votes (Survey 2 – question 1) was also changed from the (sparse) suggestions in literature. In soundimagery studies (as reported by [Simpson, 2007]) participants listened to sounds sequentially and they were then asked to assign them to the alerts that were displayed to them. The research by Gardner, Simpson and Ulfevngren that has used soundimage studies focuses on military aircraft cockpit alerts or other warning sounds, which tend to have simple meanings (e.g. 'missile launch' or 'low on fuel'). Therefore, displaying the list of all warning labels throughout the experiment is effective for choosing the most meaningful associations. However, the concepts of the mobile service clusters are not as simple, and presenting their labels all at once could be confusing. Therefore, we decided to present each one of the services independently and give participants the chance to familiarise themselves with their content through descriptions and examples. Then, they were given the option to choose the most representative of the three sounds presented to them. The disadvantage of our method is that it restricts participants to choose only amongst the sounds that have already been chosen as possible candidates for the particular service. It is possible that one of the sounds that was created as a candidate for a particular service, participants could choose as the most representative for a different service. The upside of this weakness is that the voting procedure is streamlined, as participants are required to listen only to a small set of sounds that experimenters have already pre-selected as potentially good candidates. In any case, researchers would need to do some kind of filtering of the potentially infinite number of everyday sounds before the associations are empirically chosen, and it is preferable that this was done by a specifically designed brainstorming session.

The second measure of sounds' representativeness was through subjective rating (Survey 2 – question 2), and was similar to the one found in the literature [Gardner & Simpson, 1992, as reported by Simpson, 2007]. There, participants were asked to rate "the strength of the sound-meaning association" on a scale from 0 to 10 (what the authors call "degree of soundimagery"). We chose to present a spectrum from "representative" to "not representative", as we considered this wording as more immediate and easy to understand by non specialists. On the other hand, in studies where this strength is tested, many authors opt for the subjective measure of "appropriateness". We did not choose to use this wording, as appropriateness can be affected by many factors, such as context of use or pleasantness to the ear. The aim of this study was to produce auditory icons with strong relationships to services, so that they can be evaluated in metrics such as learnability, memorability, appropriateness and pleasantness (see Chapter 6). However, we identified two possible problems with the metric we used. First, from the free-form comment responses we know that some participants did not like the fact they could not rate appropriateness or pleasantness, and were split between rating these metrics and representativeness. Perhaps providing both metrics would reduce the possibility that representativeness measure was affected by low preference ratings. Second, the rating results were possibly biased towards high representativeness, since participants had already chosen the sound they were rating. This is also supported by the overall average of ratings being above neutral (3.5 out of 5) despite comments that voting was sometimes made on the basis of which sound was "least relevant".

5.2 Designing Earcons

Earcons are fundamentally different from auditory icons in the fact that the associations between the sounds and their referents are in essence arbitrary (or loosely metaphorical). However, they can be constructed so that they formulate a grammar (compound earcons) or represent hierarchical structures, such as menus (hierarchical earcons). Although compound earcons have been utilised to successfully represent a deep hierarchy, they do so by introducing further indirectness in the signal-referent relationship (sounds need to be interpreted in numbers, and then numbers into nodes' contents) [Brewster, 1998]. This is expected to introduce higher workload in the sound interpretation, which is particularly undesired in what should be simple, everyday technology. On the other hand, hierarchical earcons have been proved efficient in representing small hierarchies with little training [e.g. Brewster el al., 1993]. Therefore, this type of earcon was chosen to represent our 10-node service hierarchy (Table 4-5).

The challenges of designing hierarchical (or any other type of) earcons as mobile service notifications are significantly different than the requirements of auditory icons. A set of guidelines has been produced as a result of iterative design and empirical evaluation [Brewster et al., 1993]. Amongst the most important suggestions is to use musical instrument timbres that are easy to tell apart, and that complex intra-earcon variations in pitch and rhythm should be applied to make them distinguishable from one another. It was also found that complex timbres were the most effective for differentiation, and pitch alone was the least effective property for differentiation (unless very large variations of two or three octaves are used).

We constructed 10 earcons to represent the 10 services of our classification, following the literature guidelines to the best of our ability. However, sound composition will always be a subjective process, and the effectiveness of our earcons could be affected by our design. Therefore, as with auditory icons, we empirically tested the designed sounds to see how distinguishable they are, and how well they represented the service hierarchy. Next, we present the experimental comparison of our earcons against a set of simple control tones (Experiment 1). After obtaining feedback from the participants, we improved the earcons with the help of a professional musician. The new set of earcons was retested for distinguishability, their ability to represent the classification structure, and their pleasantness (Experiment 2).

5.2.1 Experiment 1: Earcons vs Simple Tones

Method

A repeated measures experimental design was followed. Audio type was the independent variable (earcons and simple tones) and the dependent variables were the number of absolute errors, the number of category errors and reaction times. It was hypothesised that earcons will be significantly more accurate (in absolute and category level accuracy), and significantly quicker than simple tones.

Participants

Twenty-four participants took part in the experiment, 15 of whom were males. Their mean age was 28.1 (SD=5.5) with a range from 23 to 41. None reported hearing difficulties, 79% of them played a musical instrument or sung in a choir, while 2 of them noted that they were professional musicians. Their overall musical ability was self-rated at 3.3 in a 5-point Likert scale, with 1 being very low and 5 very high. They were all familiar with using a PC, and tea and biscuits were offered as incentive for their participation in the experiment.

Instruments and Measures

The lab experiment was designed and ran using the MediaLab[1] software package, on an IBM desktop PC with a standard 15" display. Logitech headphones were attached to the computer and user input was recorded by MediaLab through the use of the keyboard and the mouse of the PC.

Ten earcons and 10 control sounds were constructed to represent the 10 services. In order for the earcons to represent the hierarchy of services, each one of the four super-categories (A-D in Table 4-5) was represented with a different musical instru-

[1] http://www.empirisoft.com (last accessed 30/06/2009)

ment/timbre. The sub-categorisation of World-to-User services to ‘Information Services’ and ‘Entertainment Services’ was not used at this stage. Within each category a variety in pitch and rhythm complexity (number of notes, polyphony and duration) was utilised to make the earcons as distinct as possible. The 10 control sounds were simple tones (timbre of the musical instrument koto) and varied in pitch 3 semitones apart one another, within a range of just over one octave. They were randomly assigned to the 10 services so that participants’ learning was not aided by an obvious pattern (e.g. rising pitch to rising service numbers). Table 5-3 shows the 10 service categories along with the timbre and the musical notation of the earcons, while the sound files can be found in the folders “Chapter5.Control_Tones” and “Chapter5.Earcons” on the thesis’ accompanying CD.

Procedure

The procedure was divided in training and testing phases, which participants completed individually while seated in a quiet HCI lab. During the training phase, participants were given instructions about the testing task and the service hierarchy was briefly explained to them. The 10 services appeared on screen including the four categories, which were labelled “Services from the world”, “Services from other users”, “Services from myself” and “Other services”. This screen was displayed throughout the training and testing phases. Then, participants were given up to 5 minutes to learn the associations in a form of active learning: sounds were initiated by clicking on the corresponding service. During the testing phase the 10 stimuli were presented in a random order and participants were expected to indicate the correct service by pressing the corresponding function key (F1-F10). Each sound played for up to three times or until the participant made a selection. After all 10 sounds were presented, feedback on their responses was provided, including the overall success rate (e.g. 7/10) and a breakdown with the correct and incorrect responses. The testing cycle was repeated three times before training for the other sound type commenced. The order of sound type was counterbalanced between participants (earcons were referred to as “musical sounds” and control sounds as “simple tones”). A short questionnaire gathering user preference data was completed after each sound type, and a comparative questionnaire at the end of the experiment. The consent form, instructions and questionnaires can be found in Appendix VI.

Results

Accuracy of responses was measured on two levels: absolute and category level. Absolute accuracy was measured by the number of errors when assigning a sound to its service, and category level accuracy by the number of errors while assigning a sound to the correct of the four categories. The first metric indicates how easily distinguishable and learnable are the sounds of each category and the second metric indicates how well sounds represent the service hierarchy. For example, when ‘earcon 1’ was played (piano sound representing ‘News Information’), any response from services 1 to 5 (World-to-User services represented by piano timbre) would be considered as correct for the category level accuracy. Any response from services 6 to 10 (belong-

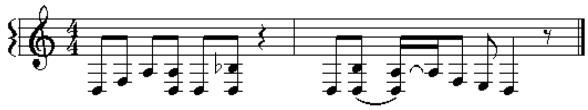
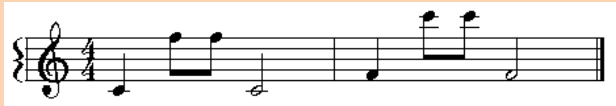
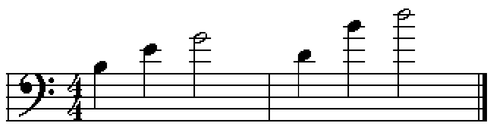
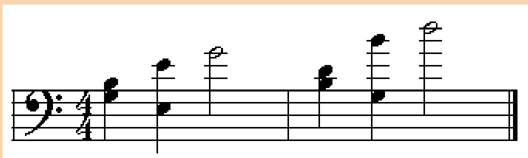
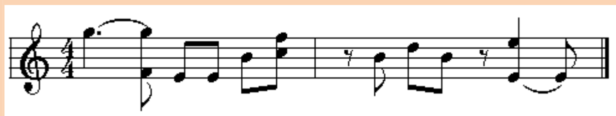
Service	Instrument	Musical Notation
1	Piano	
2		
3		
4		
5		
6	Pan Flute	
7		
8	Vibraphone	
9		
10	Violin	

Table 5-3: The 10 service categories along with the timbre and the musical notation of the earcons

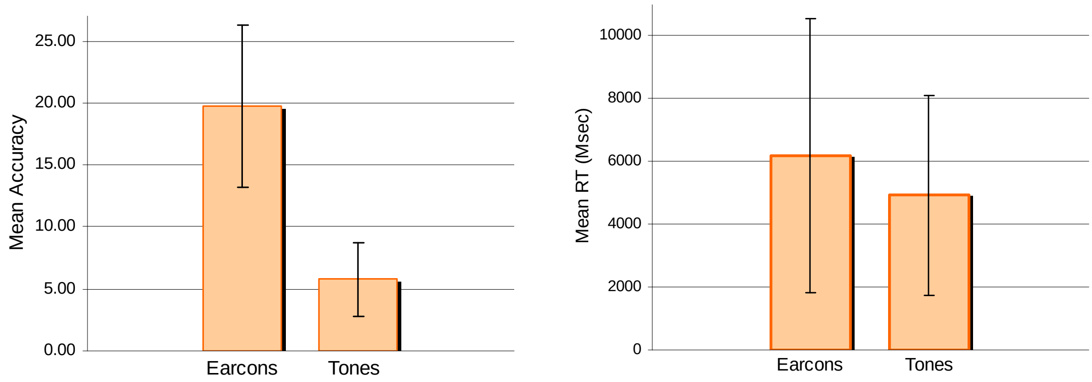
ing to different categories and represented by different timbres) would be considered as incorrect.

Absolute errors and response times across all three testing sessions were calculated for each sound type. A Shapiro-Wilk test found responses for earcons to be normally distributed across participants (D(24)=.934, p>.05) but tone responses significantly different from normal distribution ($D(24)=.906$, $p<.05$). On the other hand, category level errors were normally distributed for tone responses ($D(24)=.152$, $p>.05$) but not for earcon responses ($D(24)=.217$, $p<.005$). Earcon response times were also not normally distributed ($D(24)=.747$, $p<.001$), and remained so even after an outlier was removed (almost four times the standard deviation above the mean). Logarithmic transformations were attempted but affected the distribution of responses that were already normalised. Therefore, the non-parametric Wilcoxon signed-rank test was performed for both accuracy levels and response times.

Both accuracy hypotheses were confirmed. Absolute accuracy was significantly higher for earcons (Mdn=19) than tones (Mdn=5.5), $z=-4.29$, $p<.001$, $r=.62$, and category level accuracy was significant higher for earcons (Mdn=27.5) than tones (Mdn=12), $z=-4.29$, $p<.001$, $r=.62$. The average accuracy success for earcons was 66% (SD=.47) and 20% (SD=.39) for the control sounds (and category level accuracy 80% and 40% respectively). These results indicate that earcons were much easier to learn than simple tones, and they were also much better in representing the structure of services. However, earcons (Mdn=5.56sec) produced significantly slower responses than tones (Mdn=4.76sec), $z=-3.07$, $p<.002$, $r=.44$. Figure 5-1 shows the comparative absolute accuracy and response time means across sound type conditions (30 trials for each sound in total).

Furthermore, our observations and participants' comments indicated three more interesting results, which we tested for significance. First, tones for services 2 and 9 produced significantly more accurate results than the rest of the tones, $z=-4.15$, $p<.001$, $r=.85$. These were the sounds with highest and lowest pitch respectively. Second, earcon 10 produced significantly more accurate responses than the rest of the earcons, $z=-3.75$, $p<.001$, $r=.77$. This was the only earcon that did not share its timbre with any others. Finally, the 5 piano earcons (representing World-to-User services) were more difficult to learn than the rest of the earcons (three categories with two services each). Indeed, a Wilcoxon signed-rank test showed significantly better absolute accuracy results for earcons 6-10 (Mdn=11.5) than earcons 1-5 (Mdn=8.5), $z=-2.19$, $p<.05$, $r=.45$. This indicates that the piano earcons were more difficult to distinguish than the others. Therefore, a professional musician was asked to improve their distinguishability and the new sounds were tested in the experiment reported in the next subsection.

Subjective data were in agreement with the accuracy results, with earcons scoring much higher in user preference. In the self-assessment question regarding learnability, earcons were found slightly difficult at 2.8 on average on a 5-point Likert scale (with 1 being very difficult and 5 being very easy). Tones were perceived as very difficult to learn at 0.2 on average. In the perceived difficulty for category level accuracy, earcons were found again relatively easy at 3.2 (against 0.3 for tones). In a direct question whether they would like to have the sound types as audio notifications on their mobile phones, 71% of the participants said they would want to have at least some of the earcons, and 8% would want to have at least some of the tones. In a direct comparison question, 42% of the participants opted for earcons over the tones,



(a) (b)

Figure 5-1: Comparative absolute accuracy (a) and response time (b) means for earcons and simple tones (30 trials in total)

but 38% of them opted for the choice of "a different set of sounds" (not heard during the experiment). Finally, when given the choice between speech and non-speech mobile service notification, 50% opted for non-speech only and none for speech only. However, 38% said they would prefer a combination of speech and non-speech sounds, depending on the service they were notified about.

5.2.2 Experiment 2: Evaluating the Improved Earcons

Method

Between subjects experimental design was used in this experiment. The independent variable was the set of earcons (old vs new), and the dependent variables were the number of absolute errors, the number of category errors, and response time (as in Experiment 1). Thirteen new participants took part in this experiment and only listened to the new earcons. The responses from the 12 participants who were in the first condition of Experiment 1 (earcons-first condition) were used as the comparison data for the old earcons.

Participants

The 12 old participants (from the earcon-first condition of Experiment 1) had an average age of 28.5 (SD=5.2) and 7 of them were males. Of them, 17% were professional musicians, 83% played a musical instrument or sung, and they rated their musical ability at 3.3 (1 being very low, 5 being very high). The 13 new participants had an average age of 29.7 (SD=5.5) and 6 of them were males. Only 8% of them were professional musicians, 62% played a musical instrument or sung, and they rated their musical ability at 2.5. None of the 25 participants reported hearing difficulties, they were all familiar with using a PC, and tea and biscuits were offered as incentive for their participation in the experiment.

Instruments and Measures

The same equipment and set up with Experiment 1 was used. The new earcons were improved by a professional musician in order to be made more distinguishable and

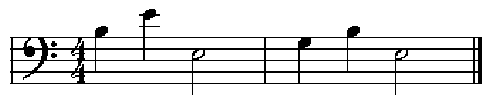
Service	Instrument	Polyphony	Key	Musical Notation
1	Piano	Monophonic	D Minor	
2			A Minor	
3			C Minor	
4		Polyphonic	G Major	
5			C Major	
6	Pan Flute	Monophonic	G Major	
7			E Minor	
8	Vibraphone	Monophonic	G Major	
9		Polyphonic	E Minor	
10	Violin	Polyphonic	D Minor	

Table 5-4: The new earcons, along with their characteristics and their musical notations

more aesthetically pleasant. First, the dimension of polyphony was systematically utilised and controlled to make them more distinguishable and increase their ability to represent the hierarchy of services. The subcategory of 'Information Services' was represented with monophonic piano earcons and the 'Entertainment Services' with polyphonic piano earcons. Similarly, the two services in the category 'services from myself' were further distinguished by making only one of them polyphonic and one monophonic. The "Services from others" category was represented by a pan flute, which could only be monophonic. However, the two services in this category were distinguished by manipulating the length of the notes (short against long) and having the same motive repeating for 'Incoming Calls' (as ringtones usually do).

Another improvement for the new earcons was the manipulation of the musical key of the pieces. Each earcon was designed to be in only one key, while major and minor keys were utilised to further distinguish the categories, similar to the polyphony dimension. ‘Information Services’ were in minor keys, ‘Entertainment Services’ in major, and the pairs of services in the other two categories alternating between G major and E minor. Finally, the melodies were design to be distinctively ascending, descending, repeating or jumping up and down. This was done to improve both distinguishability and aesthetics. The new earcons, along with their characteristics and their musical notations appear in Table 5-4, while the sound files can be found in the folder “Chapter5.New_Earcons” on the thesis’ accompanying CD.

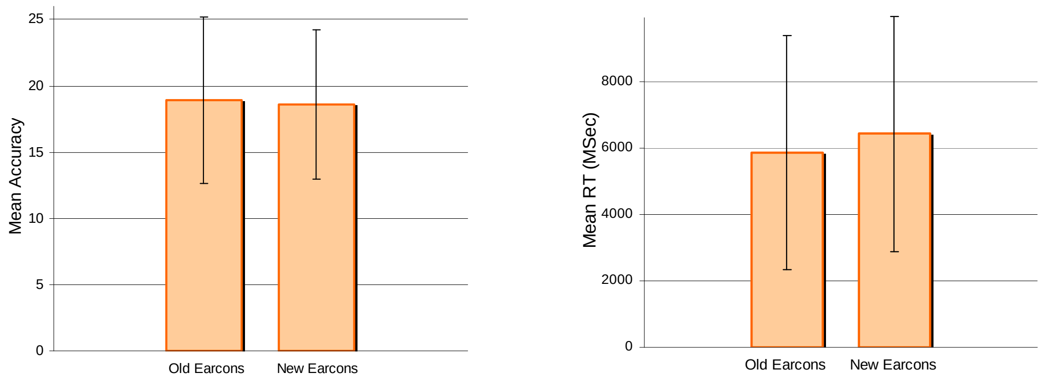
Procedure

Similar setup and procedure to Experiment 1 was followed, with the only difference being that only one condition was presented to participants (new earcons).

Results

Absolute errors and response times across all three testing sessions were calculated for the old and new earcons. Distributions were again not normally distributed and a Mann-Whitney test was performed. The null hypothesis was accepted for both accuracy levels. The new earcons (Mdn=21.5) performed better in absolute accuracy than the old earcons (Mdn=17.5), but no significant level of difference was reached U=71.5, z=-.029, ns, r=.06. Also, no significant difference between then new (Mdn=25) and old earcons (Mdn=27.5) was observed for the category level accuracy U=65.5, z=-.686, ns, r=.14. Furthermore, no significant difference was reached in response times U=50, z=-1.523, ns, r=0.3 (old earcons Mdn=5.49sec, new earcons Mdn=6.33sec). Figure 5-2 shows the comparative absolute accuracy and response times means across sound type conditions (30 trials for each sound in total).

In terms of subjective data, 75% of the participants would like to have at least some of the new earcons as notifications on their mobile phones, when 58% had answered the same for the old earcons. In the self-assessment question with regard to learnability, the new earcons scored 2.9 on average (with 1 being very difficult and 5 being very easy), against 1.5 of the old earcons. Finally, in the ability of earcons to represent the hierarchy there was a slight drop in subjective data. Old earcons scored at 1.8 and new earcons scored at 2.1 (1 being very easy and 5 very difficult).



(a) (b)

Figure 5-2: Comparative absolute accuracy (a) and response time (b) means for new and old earcons (30 trials in total)

5.2.3 Discussion of Earcon Design Process

The main requirement in earcons design is to make them distinguishable enough so that it is easy to learn and remember the (arbitrary or loosely metaphorical) associations to their referents. Furthermore, hierarchical earcons need to be designed so that they convey the structure of the concepts they represent, such as a hierarchy of mobile services, as this helps with the association memorising. However, a more subjective design requirement is in relation to their perceived pleasantness. Although this is also a requirement for other types of audio notifications (such as speech or auditory icons), the musical nature of earcons shifts their design process more towards the art of music composition. The possibility they might sound unfamiliar to users or incompatible with their preferences needs to be more carefully considered.

To address these challenges, expert consultation and/or iterative design and empirical evaluation are suggested [Blattner et al., 1989; Brewster et al., 1993]. Through these means we have designed, tested and attempted to improve a set of earcons. In the first experiment we conducted, we found earcons to be significantly more learnable and able to represent the (simplified) mobile service structure than simple tones. These results indicate that we successfully applied the earcon design guidelines, and that our earcons at least partially met the respective requirements. Our results are in agreement with those of [Brewster et al., 1993] with timbre being the most effective parameter for distinguishability, as the lone violin earcon was significantly more learnable, and the five piano earcons least learnable. Also in agreement with the literature, we found that simple timbre sounds that vary only in pitch (especially within small range) are not easily distinguishable or learnable. Furthermore, we found that the most distinguishable amongst them were the upper and lower boundaries of the set, which were just over one octave apart. This confirms the requirement for large pitch variations, and reflects the way participants attempted to learn the tones associations (also supported by their comments): using the pitch boundaries of the set as anchors, they tried (and mostly failed) to work their way though the octave. Although we do not have enough data for statistical analysis, there is also an indication that musicians were more effective in differentiating the simple tones. One of the two musicians in our experiment performed more than twice the standard deviation over the mean (achieving 47% overall accuracy), and the other musician achieved the second best score (also shared by others). However, their performance does not stand out from the other earcon responses. Although there are some inconclusive results in the literature with musicians' ability to perform better at similar tasks, their relative advantage seems to diminish with complex timbres (see Section 3.2.1).

As established in the literature, we followed iterative design and improved our earcons based on the empirical findings. We sought advice from a musician and focused on improving the distinguishability, especially for the five piano earcons. Unfortunately, the second experiment did not show any significant improvement between the old and the new earcons. There are two main reasons we believe performance did not improve. First, we did not explain the structure of the sounds to the participants during the training phase. Care was taken for the new earcons to represent the hierarchy better (e.g. by manipulating polyphony), but this improvement would not reflect on performance if participants did not perceive these mappings. Although the type of active training we applied in our experiment has been identified as essential part of earcon training, it probably was not enough to reveal to participants the relationships between earcon attributes and services. Second, some participants opted to cut short the five minutes of their training, as they felt confident they had learned the associations. In hindsight some admitted that they should have spent more time training before the testing rounds. Although we had explained the testing task before the training commenced, it appears it was not enough to keep participants motivated enough to fully exploit the training session. Perhaps motivation should be infused in a form of a contest in future studies.

Despite these limitations, our earcons reached 65% overall accuracy, which is comparable to findings of similar studies in the literature. For example, absolute accuracy of nine earcons representing a menu (by manipulating timbre and rhythm) was found to be just over 60% [Brewster et al., 1993]. Although in other studies earcons have reached as much as 90% for 18 earcons [Hoggan & Brewster, 2007], they are difficult to compare as the dimension of spatial sound presentation was manipulated. Nevertheless, there are two ways we expect our earcons scores could improve. First, as explained above, learnability should improve if the structure of earcons is made clear to the users. Second, the overall feedback we provided at the end of each testing round did not seem to facilitate learning. Accuracy did not change considerably between the three attempts in either of the two experiments (in fact, it slightly dropped with each attempt). This means that the experiment was essentially measuring participants' ability to learn the associations after 5 minutes of active learning but did not truly give them the potential to learn them further. Third, further fine-tuning of the earcon properties could improve their distinguishability. For example, we could exaggerate the register differences, or explore multiple timbres for each earcon as done elsewhere [Brewster et al., 1993]. Time constraints stopped us from re-iterating the design process and measure the effect of the discussed experimental design changes on learnability performance.

5.3 Chapter Summary

In this chapter, we have presented the methodologies we designed for creating a set of auditory icons and a set of earcons to represent 10 mobile services. These methodologies were based on the relevant domains of the literature. They were distinctively different for the two types of sounds, since their nature and design requirements are different. Auditory icons need to utilise everyday sounds that are easy to identify, and which are linked to the services they represent through a meaningful metaphor. We therefore produced a number of ideas for such metaphors for each service through a brainstorming session, and then empirically tested the sounds we collected through two online surveys. The first survey measured the identifiability of the sounds, and the second measured the level of their representativeness for the respective services. The least identifiable sounds were excluded, and of the remaining sounds, the most representative for each service were chosen as auditory icons.

On the other hand, earcons need to be easy to distinguish and pleasant to the ear, while they maintain their ability to represent a menu structure (if needed). We therefore designed a set of 10 earcons with these requirements in mind by following the relevant established guidelines of the literature. We then empirically tested the earcons against 10 control sounds (simple tones). Although earcons were found significantly better in all three metrics (distinguishability, pleasantness and ability to represent the service hierarchy), we followed iterative design and attempted to improve them with the assistance of a professional musician. In a second experiment however, they were not found significantly better than the previous set.

In creating or adapting those methodologies from the literature, we ensured that the produced notifications are reasonably good representatives of their respective sound types. We are therefore one step closer in answering our research question: "How can we design for non-intrusive yet informative auditory mobile service notifications?" By comparing the effectiveness of these two sets of notifications (in the next chapter), we will be able to draw conclusions not only in the relative effectiveness between auditory icons and earcons, but also assess the efficiency of the methodologies presented here.

Chapter 6

Evaluating the Notifications and the Design Process

In previous chapters we have described three discrete steps as part of the analytical and empirical design process for effective auditory icons and earcons as mobile service notifications. Having instantiated those steps, we have obtained a meaningful and manageable mobile service classification, and a set of carefully designed and empirically validated auditory icons and earcons. In this chapter, we perform an extensive evaluation on the effectiveness of these sets of sounds as mobile service notifications. The aims of this study are twofold. The first is to compare the effectiveness of notifications between the sound types we have obtained. Since efforts have been made to meet the particular design requirements of each sound type, this comparison should bear more reliable results than previous studies in the literature, where the sounds were not produced in a systematic way. The second goal is to assess the stimuli design methodologies we followed in Chapter 5, and derive a set design guidelines. In other words, to which extent have these design methodologies provided us with effective mobile service notifications, and how can they be modified to produce better notifications?

For both of these research goals a broad and extensive evaluation of the notifications' effectiveness is needed. The experience gained and the lessons learned in the studies of previous chapters informed the experimental design of this study. Our first investigation in auditory mobile service notifications (Section 4.1) has given us insights that inform the intuitiveness evaluation of the two newly acquired sets of sounds. In addition, the learnability experiments we conducted in Section 5.2 (and the literature review in Chapter 3) have equipped us with knowledge and experience to improve and extend the learnability procedures. All of our previous experimental investigations have also extended our understanding of accessing subjective preferences of appropriateness and pleasantness of the notifications. Furthermore, in this study we investigate memorability and gather user preference data about the two sound types over periods of time (thus significantly extending current literature). Remembering the meaning of service notifications is particularly important for services which are suggested to us infrequently. On the other hand, their pleasantness becomes more important for services that present themselves often and therefore have a greater potential of becoming annoying. We have argued that the metrics of intuitiveness, learnability, memorability and user preference are a complementary set in accessing the effectiveness of auditory icons and earcons as mobile service notifications.

In the next section we present a 4-stage comparison study that contributes to a deeper understanding of the nature of auditory icons and earcons, especially as mobile service notifications. We first present an overview of the study with the reasoning behind our design decisions, and then the method, procedure and results. By comparing the findings along the four metrics, we conclude in a number of specific suggestions for auditory mobile service notifications. In Section 6.2, we use the study outcomes to assess the design methodologies of the notifications. By comparing and relating the results to the methodologies described in Chapter 5, we are able to draw conclusions and produce guidelines towards successful mobile service auditory notification design.

6.1 Comparison Study: Intuitiveness, Learnability, Memorability and User Preference

In our experiment in Section 4.1 we measured intuitiveness of notifications by testing how untrained participants assigned the notifications to services. Similar to the soundimagery studies [as reported by Simpson, 2007] we are interested to see if there are inherent meanings in our experimental sounds that would associate them with any particular services. The more accurately and quickly a naïve listener recognises a sound and relates it to a concept or event, the more intuitive is the sound's meaning. By definition, we would therefore expect auditory icons to perform better than earcons, since efforts have been made to design them with strong associative links. This was also empirically validated in our previous experiment. In the study presented here we measured intuitiveness with two aims in mind. The first aim was to evaluate the strength of the metaphors of the auditory icons we used, in order to produce requirements for the notifications and their respective design guidelines. The second goal was to set a baseline for the learnability measurements that follow in the next stages of the study.

Learnability of associations between auditory stimuli and interface functions or events is traditionally measured in short laboratory sessions. This imitates the learning procedure skilled operators have to follow to get qualified for their line of work (e.g. flying an aircraft). Participants usually go through training and testing phases, which usually include a combination of the following steps:

- Training: presentation of the experimental stimuli for familiarisation; tutorial explaining the associations; a few minutes of active learning, where participants click on-screen labels to hear the sounds; continuous training during the testing phase.

- Testing: random presentation of sounds that participants have to associate to the correct function (displayed in a list on screen, or from memory); trial-by-trial feedback given on response, and/or overall feedback given at the end of each round; repetition until criterion is met, which can be the number of rounds performed, the level of accuracy achieved, or the time spent; measurements to determine learnability can be response times for accurate responses, number of overall errors, number of rounds or time required to reach criterion.

However, these steps are not necessarily representative of the methods we normally employ in order to learn the meaning of things (such as mobile notifications) in real life. We usually manage to learn such associations without tutorials or training sessions. The learnability of mobile notifications should therefore be tested under real life conditions too. To this end, the first part of learnability testing process is designed to imitate the casual type of training that takes place in everyday life. We designed a weeklong field study, during which participants received a number of notifications and were asked to identify the service they represented. Although these notifications did not signal the availability of an actual service, feedback was given after participants' selections in order to enable learning. Also, information about their social context and current activity was gathered in order to explore relationships between context of use and preference data.

Nevertheless, this kind of field study can have many confounding variables affecting the measurement of the dependent variables. We therefore designed two more measures of learnability. The first was a controlled laboratory session that measured how well participants had learned the associations during the field study (by comparing against the pre-training baseline set in the intuitiveness session). The final measure of learnability was similar to the laboratory sessions commonly found in the literature. It was carried out in order to produce results comparable to the literature, and set a baseline for the memorability tests that followed. The training/testing procedure we followed in this learnability experiment was informed from the relevant literature (Sections 3.2 and 3.4) and our own experience on mobile service notification testing (Sections 4.1 and 5.2). A combination of tutorial and active learning was applied, which has been established as the most effective training method [Brewster, 1998]. In addition, the exposure participants had to the experimental stimuli during the two preceding tests and the field study can be seen as part of the familiarisation stage introduced by some researchers. However, we expected that this kind of casual training would not be enough for participants to learn the associations, and thus feedback was provided during the testing phase. Based on our experience from the earcon experiments we have conducted (Section 5.2), the overall feedback at the end of the testing rounds was inadequate to improve performance. On the other hand, in the experiments where on-response feedback is provided, the learning process is intrinsically integrated with the testing session, from which the accuracy data are extracted. We are aware of only one study that separated the two procedures [Williams & Beatty, 2005]. In that study, active training (with on-screen visual aides) took place until participants decided to test themselves without the visual aides. If the learning criterion was not met (70% accuracy), they returned to the active learning process. In order to avoid interaction between learning and performing, we presented the two processes separately. After the tutorial, training was provided by presenting the sounds in a random order and requesting from participants to guess the respective service until their selection was successful. In other words, a new sound was presented only after they had selected the correct service for the previous one (which was repeating). We expected that this persistent on-response feedback would maximise learning. Conversely, the testing phase was not interrupted by on-response feedback, but instead overall feedback was produced at the end of each round. This had a detailed breakdown of the correct and incorrect responses to further facilitate learning. The on-response training was repeated until a 100% accuracy response was produced in the testing phase. With this experimental setup, we expected that six training/testing rounds would be adequate for participants to learn the 10 auditory icons/earcons, and it was therefore set as a cut off criterion too.

The experimental results of learnability had also the function of a baseline for the following memorability tests. Similar tests have been performed in the literature up to one or two weeks after the learnability tests, without any further training or exposure to the stimuli in between. However, the results always indicate a similar or improved recognition rate in the follow-up session (for each sound type individually). Perhaps a ceiling effect is formed by a combination of factors, such as the number of sounds learned, their ease of learning and the amount of time between the two sessions. In order to see if there is any significant difference in memorability of auditory icons and earcons, we tested the same participants one and four weeks after their training sessions. To ensure no further learning took place, there was no exposure to the experimental sounds between sessions, and only overall feedback (without detailed breakdown) was produced at the end of the memorability tests (one attempt per sound on each of the two testing days).

Finally, data on user preference for each individual notification were gathered throughout all stages of the study, including the weeklong field study. We believe it is particularly important to capture this information in the real context of mobile use. Literature studies commonly measure user reactions in laboratory studies that typically do not last more than an hour. It is possible that people's aesthetic preference and acceptance would change if they received the notifications on their mobile phones throughout a period of time. The cognitive and attentional demands, the social context and the frequency of the notifications are only some of the factors that could affect user preference. Furthermore, subjective data gathered for user preference in our first mobile service notification experiment (Section 4.1) were inconclusive. Responses on how much participants liked sound types were contradicting responses on how much they would like to have them on their mobile phones. In this study, we chose to phrase the question with regard to willingness to own, as it is more directly related to the aims of our research. We were interested to see which notifications types were going to be more effective as mobile services, and users' willingness to adopt them is a very crucial metric for this.

Next, we report the details of the 4-stage evaluation study we conducted to compare our auditory icons and earcons along the dimensions of intuitiveness, learnability, memorability and user preference. We present and discuss the results on the notifications' performance, leading to suggestions for design requirements.

6.1.1 Method

A repeated measures experimental design was followed in the 4-stage study. The independent variables were sound type, time and training, and the dependent variables were response accuracy and response time. In the first stage (Lab 1), the intuitiveness of the notifications was measured by the number of correct guesses by untrained listeners. In the second stage (Field Study), participants started learning the sound-service associations in the natural context of mobile phone usage for the duration of one week. They received audio notifications on their phones at different times of the day, and were asked to guess the services they represented. Learnability was enabled by immediate textual feedback on their guesses, and was monitored from the number of correct responses. In the third stage (Lab 2) participants repeated the Lab 1 process so that learnability during the field study could be measured accurately. Then, and since they had not reached 100% accuracy during the field study training, they underwent rigorous lab training and were again tested in how well they learned the associations. This provided us with further learnability data, such as the number of trials required to reach 100% accuracy, the number of errors made in the process, and the response time to each stimulus. Stage 4 (Web 1 and Web 2) consisted of 2 online experiments testing memorability after 1 and 4 weeks of Lab 2. Participants had no further training or exposure to any of the sounds between the experiments and data recorded were the number of correct/incorrect responses. User preference data were collected throughout all 4 stages, in the form of Likert-scale responses in questions with regard to their willingness to have the notifications on their mobile phones.

Hypotheses

Since only auditory icons are designed to have strong semantic relationships with their signified entities, it was hypothesised that *auditory icons will be significantly more intuitive notifications than earcons* (H1), as untrained listeners will have no basis on which to guess the association between earcons and services.

Previous research suggests that auditory icons are more learnable than earcons during lab training sessions (see Table 3-2) and there is no evidence to suggest that field training would be dissimilar. Furthermore, it has been found that users perform poorly in learning earcon associations in the field, even after months of using a prototype instant messaging system [Isaacs et al., 2002]. Therefore, we hypothesised that *auditory icons will be significantly more learnable notifications during the field study than earcons* (H2). Also, we predicted that *auditory icons would be quicker and easier to learn during the subsequent lab training* (H3).

Furthermore, our training session in Lab 2 was comparable to similar studies in the literature (see Sections 3.2.1, 3.2.2 and 3.4). In agreement with these results, we predicted that *both sound types would perform significantly better after lab training* (H4). Finally, we expected that the directness of the sound-service associations in auditory icons would have a positive effect on memorability (as it has been found to have on learnability). Therefore, in the web-based experiments, 1 and 4 weeks after the lab training (Web 1 and Web 2 respectively), *we predicted that the forgetting rate of auditory icons' would be slower than earcons'* (H5).

The experimental design in the four stages of the study is summarised in Table 6-1.

Stage	Method	Training	Day	Measurement	Hypotheses
1	**Lab 1**	No training	1	Intuitiveness	H1
2	**Field Study**	Field training	2 to 8	Learnability	H2
3	**Lab 2**	Lab training	8	Learnability	H3-H4
4	**Web 1 & Web 2**	No training	15 & 36	Memorability	H5

Table 6-1: The experimental design of the study

Participants

Sixteen participants took part in all four stages of the evaluation process, with the exception of two participants who could not complete the web-based experiments (stage 4). There were 11 males and 5 females and their average age was 39.2 with a large range, from 14 to 71 years old. All participants were part of a cohort of the Cityware[1] project and had been given mobile phones and free cellular airtime for 3 years in exchange for their participation in studies such as the one described here. As extra incentive 3 Amazon vouchers (of £50, £20 and £10) were offered for the 3 most responsive participants during the field study. All of them were familiar with using a computer, had been using a mobile phone for at least 3 years and were familiar with their current phone for at least 1 year.

Instruments and measures

The lab studies were designed and run using the MediaLab software, on IBM desktop PCs with a standard 15” display. Logitech headphones were attached to each computer and user input was recorded by MediaLab through the use of the keyboard and the mouse of each PC. For the field study, a custom-made Java application was installed on participants’ Nokia N95 phones, which communicated with a server initiating the notifications and recording the responses. For the last stage, a web-based questionnaire was developed in HTML and PHP, which was accessed via participants’ home or work computers.

The 10 auditory icons and 10 earcons produced by the processes described in Chapter 5 were used throughout the four stages of the evaluation process. To avoid terminology confusion, they were introduced to the participants as everyday and musical sounds respectively. Each sound corresponded to one of the mobile services of the classification produced in Chapter 4 (see Table 4-5). The detailed relationships between the sounds and the services are shown in Table 5-2 for auditory icons and Table 5-4 for earcons. Table 6-2 summarises the associations between the 10 services and the sounds. All sounds were sampled at 44.1 KHz at 128kbps bit rate and were normalised at -90 dB volume. However, during the field study their quality had to be reduced to 8 KHz sampling due to the memory limitations of the mobile phones.

---

[1] “Cityware: urban design and pervasive systems” - www.cityware.org (last accessed 30/06/2009) UK EPSRC, EP/C547683/1 grant.

Procedures

Stage 1: Lab 1 - Intuitiveness

Up to eight participants at a time performed the experiment at adjacent computers in a quiet lab and wearing headphones. Participants were presented with the list of the 10 services on their screens and the 4-level classification rationale was explained (Table 4-5). Two extra 'dummy' services were added to the original classification that had no sounds corresponding to them. They were 'Incoming MMS message' in the 'User-to-User' category, and 'Generic Alarm' in the 'User-to-Self' category. This was done to disrupt the one-to-one relationship between sounds and services, and hence hinder users from deducing associations based on the process of elimination. A similar technique has been applied elsewhere; for example, three dummy buttons were introduced amongst the labelled response buttons while testing learnability of melodic alarms for anaesthetists [Williams & Beatty, 2005].

Participants repeated the same procedure for earcons and for auditory icons with a short break in between in order to avoid fatigue effects. Also, the order of presentation was counterbalanced in order to avoid practice effects. For each sound type, all 10 sounds were presented in a random order twice, but the second presentation of any sound took place only after all 10 sounds had been played once. In the previous intuitiveness experiment we conducted (Section 4.1), we observed that the presentation of 81 sounds in one session caused restlessness and lack of motivation to some partici-

	Services	Auditory Icons	Earcons
1	News Information	BBC News ident	Piano – monophonic going up
2	Sports Information	Stadium crowd	Piano – monophonic going down
3	"Here and Now" Information	Public announcement at an airport	Piano – monophonic jumps up & down
4	Entertainment - Downloads	20th Century Fox	Piano – polyphonic going down
5	Entertainment - Live	Audience applauding (e.g. in a theatre)	Piano – polyphonic going up
6	Incoming Calls	Old-fashioned phone ringing	Flute – monophonic short notes repeating
7	Incoming SMS Messages	Message transmitted in Morse code	Flute – monophonic – long notes going down
8	Self Reminders	Windows Mobile reminder	Vibraphone – monophonic going down
9	Backup Reminders	Truck/lorry reversing	Vibraphone – polyphonic going up
10	Other Services	Wind chimes	Violin – varying pitch chords and single notes

Table 6-2: The 10 services and the corresponding sound descriptions

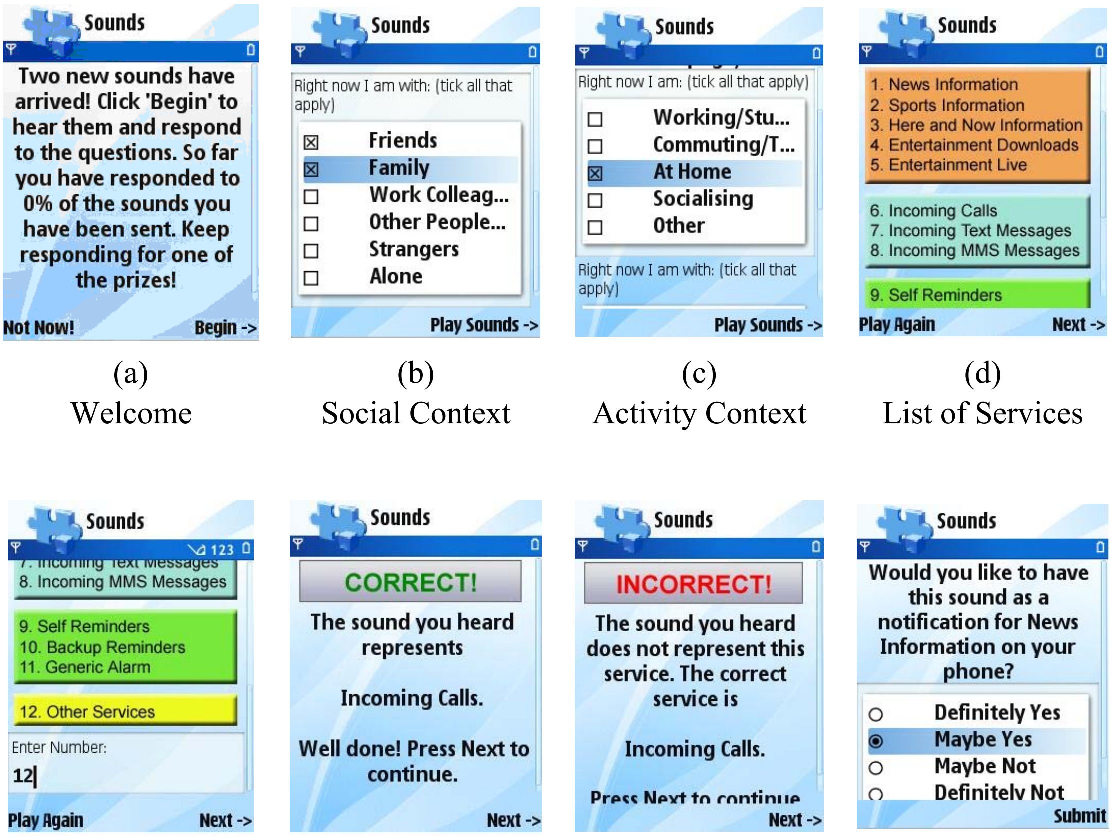
Figure 6-1: The screen presented during the Lab 1 session

pants. By reducing the consecutive presentations to 20 in this experiment, we expected that participants would stay interested while we still gathered enough data to draw conclusions.

All 12 services (10 with sounds assigned to them and 2 additional dummy services) were continuously displayed on screen throughout the testing process, including the 4-level categorisation rationale. The 12 function keys of the keyboard (F1-F12) were assigned to the services, and the associations were displayed in an ascending order next to the corresponding service labels (Figure 6-1). Each time a sound was heard, participants were required to indicate (as "quickly and accurately" as they could) the service they thought it corresponded to, by pressing a function key. Response time and correct/incorrect responses were recorded. Each sound was repeated up to 3 times at 1.5 second intervals or until a selection was made. Each trial was followed by a short questionnaire capturing user preferences with regard to each individual sound. Finally, a third questionnaire was presented, asking them to answer a few comparative questions between the two sound types. All questions were presented on screen and responses were captured using MediaLab. The relevant material (consent forms, instructions, questionnaires and screenshots) can be found in Appendix VII.

Stage 2: Field Study - Learnability

The field study commenced on the morning after Lab 1 was completed. The application on the participants' mobile phone played the 20 sounds at random times from 10am to 9pm every day for 1 week. In order to draw participants' attention to the task an SMS-type sound was played, not similar to the experimental sounds and not identical to current SMS notifications on mobile phones. This sound was presented along with a welcoming screen (Figure 6-2-a), where participants were given the option to dismiss the task until the next random time interval had elapsed. If they chose to begin the task, they were asked 2 questions with regard to their current context. The first question was with regard to their social context, and they were asked to tick the appropriate boxes for current co-presence to ‘Friends’, ‘Family’, ‘Work Colleagues’, ‘Other People I know’, ‘Strangers’ or ‘Alone’ (Figure 6-2-b). The second question was with regard to their current activity, and they were given the multiple choice options of ‘Working/Studying’, ‘Commuting/Travelling’, ‘At Home’, ‘Socialising’ or ‘Other’ (Figure 6-2-c). This information was gathered in order to explore whether context would affect users’ preferences about the notifications. Once this information was submitted, one of the experimental sounds was played and the list of 12 services was displayed on the phone’s screen (Figure 6-2-d and 6-2-e). Their task was to guess the correct correlation between sounds and services (similar to the Lab 1 experiment), by typing each time the corresponding service number in a text box at the bottom of the list. An on-screen button provided the option to replay the sound as many times as required, and another one provided detailed descriptions of the services. After they submitted their response, they were given feedback indicating whether their choice was correct or incorrect (Figure 6-2-f and 6-2-g). The correct service for the particular notification was displayed either as correction (if their response was incorrect) or as confirmation (if their response was correct). In the final



(a) Welcome | (b) Social Context | (c) Activity Context | (d) List of Services

(e) List of Services Response Entry | (f) Correct Response Feedback | (g) Incorrect Response Feedback | (h) Preference Data Entry

Figure 6-2: Screenshots from the (emulator of) the field study application

screen, participants were asked to rate how much they would like to have that particular notification on their mobile phone for that service (Figure 6-2-h).

The notifications were presented in pseudo-random order and in pairs, one of each sound type (not necessarily for the same service). This was done for two reasons. First, it ensured that participants were responding to both types of notifications in the same context each time. As the participants' activities, level of concentration, distractions etc. were expected to be changing throughout the day, we wanted to make sure that neither sound type was favoured by being responded to in more 'comfortable' situations. Secondly, the perceived annoyance or preference for the different sound types could also be affected by the context (especially the social context) within which the sounds were received. The presentation order of earcon and auditory icon within each pair was swapped every time.

Sounds that were responded to were not repeated in the same day. Pairs of sounds arrived approximately every one hour and each participant was expected to respond to all 20 sounds every day (i.e. 10 pairs of sounds per day). If they failed to respond to some of the sounds, either because they missed the welcoming notification or because they opted to skip the task, the frequency of the notifications was increased. This was done to increase the chances of responding to all the sounds within the day, but the minimum time allowed between two presentations was half hour in order to avoid causing annoyance to the participants. The interaction process of the mobile application is graphically modelled in Figure 6-3.

Stage 3: Lab 2 - Learnability

The Lab 2 experiment took place one week after Lab 1 (immediately after the end of the field study). The first part of the Lab 2 experiment was identical to the intuitiveness evaluation process of the Lab 1 experiment. After participants made their (now informed) guesses of corresponding services for both types of notifications, they were informed about the dummy services, which were then removed from the list.

A 4-step training/testing process then took place for each notification type (order of

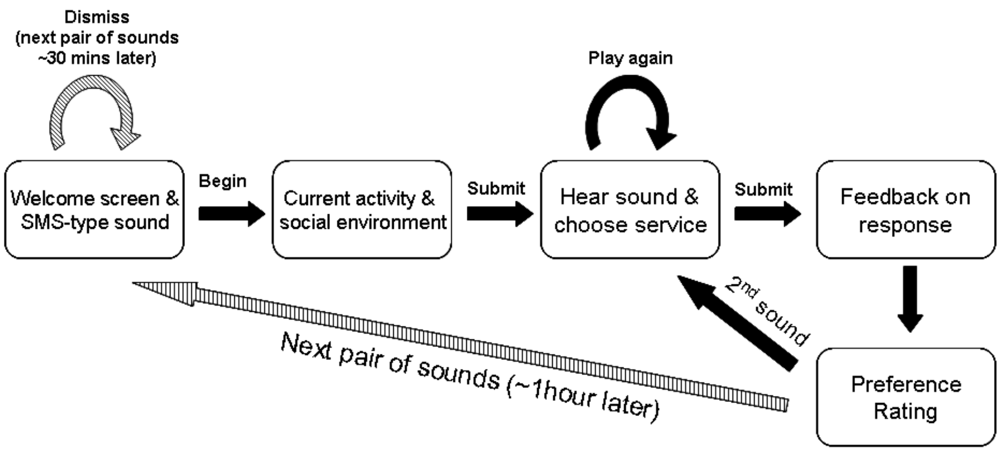


Figure 6-3: Mobile application interaction sequence

presentation counterbalanced amongst participants). In the first step, each sound was presented individually, along with its corresponding service and their associative reference. For the auditory icons, the events that produced the sounds and the metaphors associating them to the services were described. For the earcons, the instrument and some properties of the musical pieces were described (see Table 6-2), along with any metaphorical associations utilised. For example screenshots see Appendix VII.7.

In the second step of training, participants were given 4 minutes to attempt to learn and memorise the associations. The sound descriptions were displayed next to the services, and the sounds were played every time the participant clicked on the corresponding service buttons.

The third step of the training consisted of an absolute identification paradigm with trial-by-trial feedback. The notifications were played in a random order and participants had to choose the correct service, similar to the previous experiments. However, this time they were not allowed to continue to the next sound until they made the correct choice. If their choice was incorrect, they were prompted to retry. Sounds were repeated up to 3 times with 1.5sec intervals or until the correct choice was made.

The fourth and final step of the training was a similar test but with no feedback on response. Feedback was given only after all 10 sounds were heard, and was consisted of an overall percentage score for their correct responses in that round and a breakdown of which services they had answered correctly. Participants who did not reach 100% accuracy in the test repeated steps 3 and 4. The training process was stopped when they completed a test without errors, or had completed 6 rounds of the training/testing process. The number of rounds and the number of errors for each notification during testing (step 4) was recorded to indicate ease of learning. There was a 5-minute break between learnability testing of the two types of notifications. A comparative questionnaire completed the experimental procedure. After the end of the experiment, a short debriefing focus group took place, where participants were encouraged to describe their experiences during the laboratory and field sessions.

Stage 4: Web-based Experiments - Memorability

One and four weeks after the Lab 2 experiment, participants completed a web-based questionnaire. The procedure for the questionnaire was very similar to the fourth step of training in Lab 2. Participants listened to the 10 notifications of each type in a random order and indicated for each one the service they thought it represented. They had the chance to replay the sound and they were asked to indicate whether they "would like to receive this notification in the presence of others". The preference question was slightly rephrased from the previous stages, to explicitly include the aspect of social context. At the end of each sound type they were provided with a percentage score for their correct responses, but no detailed breakdown was given to prevent any further learning. For example screenshots of the Web interface see Appendix VII.7.

6.1.2 Results

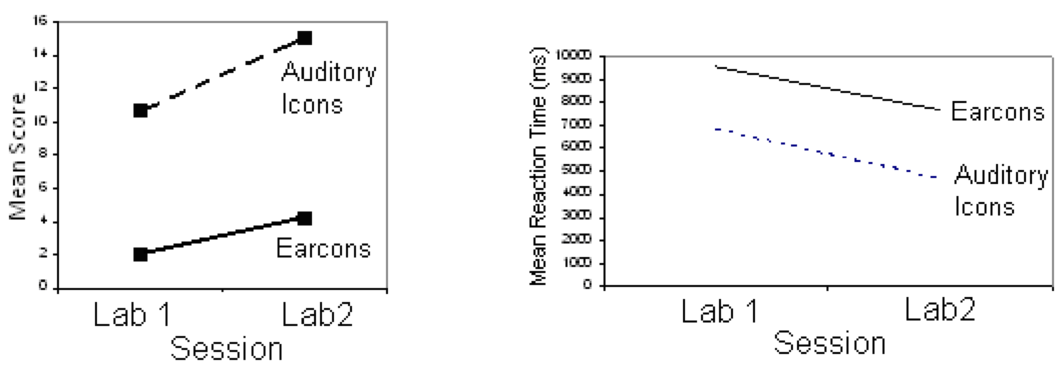
The results of the study are described in three parts. First, we present an analysis of participants' performance in terms of intuitiveness and learnability (in the two lab sessions and the field study). Secondly, we examine the memorability performance one and four weeks after the Lab 2 session. Thirdly, we explore factors contributing to variance amongst participants' preference, and differences amongst notifications within the sound types.

Intuitiveness and Learnability

This section describes the analysis of participants' ability to correctly identify the meaning of notifications in Lab 1, plus the degree to which identification improved in Lab 2, after the field study. Then, the results of the training in Lab 2 are analysed.

The numbers of errors in responses were calculated for each sound type across the 2 experiments. A 2x2 ANOVA was conducted, with two repeated measures variables: notification-type (auditory icon or earcon) and session (Lab 1 and Lab 2). There was a significant main effect of notification-type, $F=178.7$, $p<0.001$. This indicates that auditory icons were more easily associated with the correct service than were earcons regardless of training (i.e. before or after the field study). A paired-samples t-test shows a significant difference between scores for the two notification types in Lab 1 ($t(16)=11.007$, $p<0.001$). Thus the argument that auditory icons are more intuitive than earcons is supported by the empirical evidence. The guesses in Lab 1 were 53% correct for auditory icons and 10% for earcons (when chance rate was 8.8%).

There was also a significant main effect of session, $F=15.4$, $p=0.002$. This indicates that participants were more successful in correctly identifying services associated with both notification types in the Lab 2 session (after the field study training) than they were in Lab 1. There was a significant interaction between notification-type and session, $F=9.7$, $p=0.008$. This indicates that auditory icon associations were more easily learned during the week between lab sessions than were earcon associations. Accurate responses after the field study training climbed at 75% for auditory icons and 21% for earcons. Accuracy means of the two sound types for both lab sessions are illustrated in Figure 6-4-a.



(a) (b)

Figure 6-4: Comparative accuracy (a) and response time (b) means for auditory icons and earcons between sessions Lab 1 and Lab 2

Reaction times showed similar characteristics to the number of correct identifications. In both lab sessions, reaction times were lower for the identification of auditory icons than for earcons, $F(14)=42.01$, $p<0.001$. Reaction times were also lower in the second lab session than the first, for both types of notification, $F(14)=49.05$, $p<0.001$. However, there was no significant interaction between the two sound types, $F(14)=0.15$, $p=0.705$. Response time means of the two notification types for both lab sessions are illustrated in Figure 6-4-b.

For the Lab 2 training, a t-test was conducted in order to compare the number of trials required to reach a 100% correct identification rate for each type of notification (see Table 6-3). There was a significant effect of notification type, $t=5.5$, $p<0.001$. This again indicates that it was easier for participants to learn associations between auditory icons and services than between earcons and services. In the first testing trial during the training process in Lab 2, earcons reached 68% accuracy on average, while auditory icons reached 99%, with only one participant making one error. The breakdown of the accuracy results per service in this first training trial is illustrated in Figure 6-5. After the 6 trials, earcons reached accuracy of 89% on average.

	N	Mean	SD
Auditory Icons	15	1.1333	.51640
Earcons	15	3.7333	1.94447

Table 6-3: Number of trials required in lab 2 training to achieve 100% success

Memorability

Performance was analysed one and four weeks after the Lab 2 session in order to determine whether one or other set of notifications was more quickly forgotten. A 3x2 ANOVA was conducted, with two repeated measures variables: notification-type (auditory icon and earcon) and session (Lab 2, Web 1 and Web 2). There was a significant main effect of time, $F(8)=14.13$, $p=0.002$. This indicates that associations between sound and meaning were forgotten over time, for both earcons and auditory icons. There was also a significant main effect of type, $F(9)=37.5$, $p<0.001$. This indicates that auditory icons scored higher than earcons in all 3 stages (Lab 2, Web 1 and Web 2). Finally, there was a significant interaction between time and type, $F(8)=5.07$, $p=0.38$. This shows that earcon associations were forgotten more quickly than were auditory icons after one and four weeks. In particular, auditory icons were remembered with 97% accuracy and 91% accuracy in Web 1 and Web 2 respectively, while earcons' success in retention dropped to 62% and 43% accordingly.

Figure 6-6 illustrates the overall change in performance of participants for each of the sessions described: Lab 1, Field Study, Lab 2 before training, Lab 2 during training, Web 1 and Web 2. Note that the X-axis does not use a linear time scale throughout the stages, but the appropriate unit within each stage instead. Sessions have been labelled according to the measurement they related to (intuitiveness, learnability or memorability).

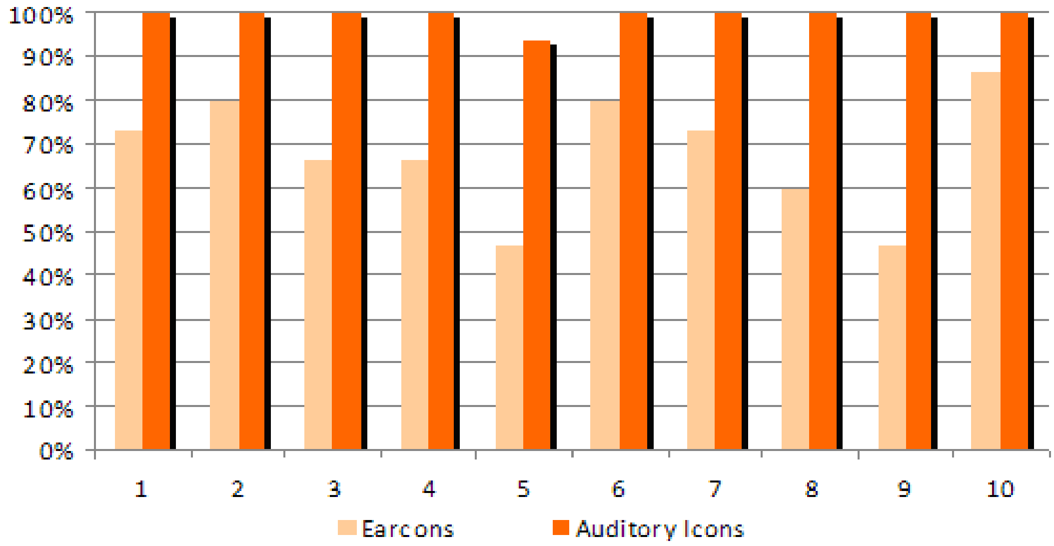


Figure 6-5: Accuracy results in the first testing trial of the training process in Lab 2

Preference and its effects amongst notifications

Looking at the four sessions together, there was a significant main effect of notification type on preference, $F(18)=19.72$, $p<0.001$, whereby participants preferred the auditory icons over the earcons. There was no significant main effect of session, $F(3)=1.96$, $p=0.13$, but there was a significant interaction effect (the difference in preference rating was largest in the Lab 2 session), $F(3)=11.00$, $p<0.001$.

Correlational analyses were used to compare average preference scores from Lab 1 session with performance in matching notifications with services in both Lab 1 and Lab 2 sessions. There was a strong, positive correlation between preference and successful identification of notifications in both Lab 1 ($r=0.527$, $p=0.02$) and Lab 2 sessions ($r=0.626$, $p=0.003$). This indicates that the most intuitive notifications were preferred more, and later on more successfully learned.

Finally, there were no significant effects of context (activity and co-presence) to the overall preference scores of all notifications during the field study. However, there was a significant drop in preference scores for earcon notifications when received while in the presence of others than while alone, $z=-2.66$, $p<.05$, $r=.12$. The context data logged are summarised in Table 6-4 (note that the percentages are calculated out of the number of times sounds were heard and do not need to add up to 100%, since participants could give no or multiple answers).

Activity (%)		Copresence (%)	
Working/Studying	26	Friends	2
Commuting/Travelling	14	Family	31
At Home	48	Work Colleagues	10
Socialising	7	Other People I know	5
Other	7	Strangers	2
		Alone	28

Table 6-4: Context data logged during the field study

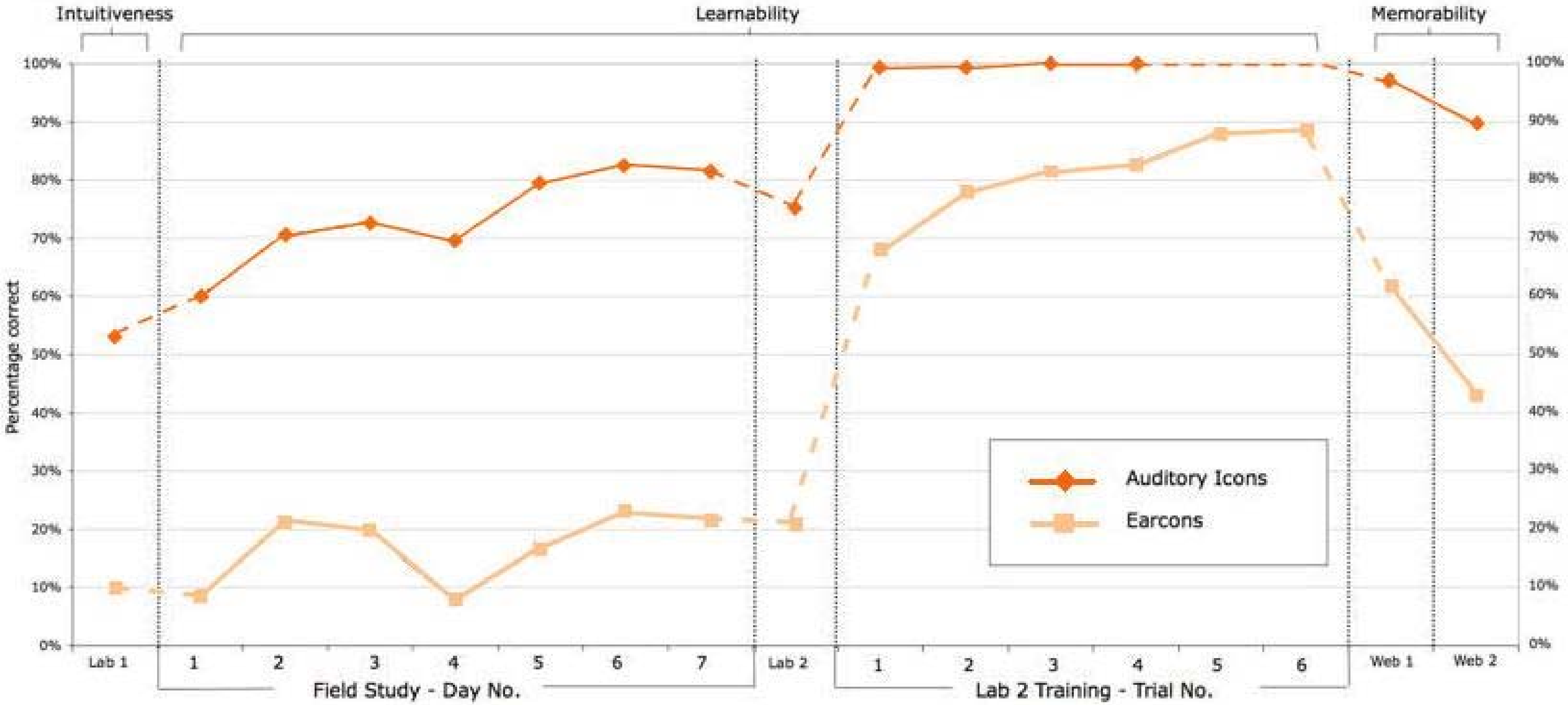


Figure 6-6: Percentage of correct identification in Lab 1, Lab 2 and Web

6.1.3 Discussion

Overall, auditory icons proved to be more effective as mobile service notifications than earcons in all four metrics we investigated. In particular:

1. Auditory icons were more intuitive to naïve listeners (H1),

2. Auditory icons were more learnable in casual everyday learning (H2) and rigorous laboratory learning (H3). Both sound types improved during lab training (H4),

3. Auditory icons were more memorable 1 and 4 weeks after they were learned (H5),

4. Auditory icons were consistently preferred over earcons as mobile service audio notifications across all stages of the study.

Each one of the metrics of effectiveness is discussed separately below, and any conclusions are translated in suggestions when selecting notification types for mobile services.

Intuitiveness

In terms of intuitiveness, it was not surprising that auditory icons performed significantly better than earcons (which performed approximately at chance rate), since they were designed with this goal in mind. In the relevant literature, any comparisons between sound types as notifications were conducted after an initial familiarisation with the experimental stimuli, which only indicates that auditory icons are much easier to learn. Our results indicate that auditory icons are more immediately recognisable because of the underlying metaphors they utilise and the directness of the signal-referent relationship. Consequently, mobile service designers (and other designers utilising similar auditory interfaces) are advised that ***underlying sound-service metaphors should be utilised (such as in auditory icons) for mobile service notifications to be immediately recognised***.

The metaphors utilised for our auditory icons were overall successful in achieving recognition, but not without exceptions. The auditory icon for ‘Self-reminders’ (service 8) performed the worst in its group, almost at chance rate (9%). The sound assigned for this service was a standard Windows Mobile reminder notification. Although not a natural sound, this is a synthetic sound that is gradually becoming part of our everyday soundscape, and therefore is moving from being an arbitrary earcon into the auditory icon space (see Figure 4.6). However, our results indicate that this sound was not recognised as a ‘reminder’ notification but as a generic ‘alert’ notification (which is a looser metaphor). This slight misinterpretation is indicated by the fact that 56% of the times it was assigned to the ‘Incoming SMS’ service, which is the most familiar mobile phone alert sound. In fact, the Windows Mobile notification was proven more representative sound for SMS than the ‘Morse code’ sound we had chosen (correctly assigned only 35% of times). In hindsight, this is not surprising. A reminder sound commonly used on mobile devices would be easily associated to the most common mobile service alert. This finding refines the previous requirement to: ***long-standing and widespread metaphors need to be utilised for sound-service links to be intuitive.***

The other two auditory icons that underperformed were the 'truck reversing' (service 9) and the 'wind chimes' (service 10) sounds, which did not exceed scores of 21% in Lab 1. They were both assigned to different services more frequently than the intended ones. 'truck reversing' was assigned 38% of times to 'Here & Now Information' and 'wind chimes' 27% of times to 'Self-reminders'. These results show that ***strong sound-service metaphors should be utilised to avoid users creating alternative semantic relationships to other services.***

Learnability

The results of the Lab 2 training are in agreement with the literature, as auditory icons were much easier and quicker to learn. In our study we followed a rigorous training process, and our participants learned all auditory icons almost perfectly in the first testing trial (with only one mistake amongst all participants). In studies where auditory icon training does not include active learning, participants need two to three trials to learn a similar number of notifications [e.g. Leung et al., 1997; Ulfevngren, 2003b]. On the other hand side, earcons reached only 68% accuracy after the same amount of training, a score that is still slightly lower than the accuracy auditory icons achieved before training. However, earcons had much more space for improvement (demonstrating steeper learning than for auditory icons) and reached 89% accuracy after 6 training/testing trials. This means that ***extensive training is absolutely necessary for arbitrary sound-service associations (such as in earcons) to be learned. Yet, there might still be a significant percentage of users who will be unable to learn all of them perfectly.*** However, many participants reported with relief that earcons made much more sense after they were explained and the rationales for memorising them were presented. Furthermore, we observed a variety of learning techniques in their efforts to learn the associations. Some participants tried to create semantic relationships between earcons and services, while others based their learning purely on the order of presentation (first service had a rising pitch, second had a falling pitch, third had a varying pitch etc). It is not uncommon for participants to develop their own semantic relationships (that designers did not intend) or follow different strategies for memorising associations [e.g. Bonebright & Nees, 2007; Brewster, 1998, Cohen, 1994].

However, one of the most interesting and valuable outcomes of this study is in relation to learnability during the field study. Although mobile phones currently employ earcon-like notifications, we found that they were extremely difficult to learn through every day usage. Our participants received all 10 earcons every day for one week and they managed to learn them only at 21% accuracy on average. On the other hand, auditory icons were correctly identified in over 80% of the times on the last two days of the field study. Even the least intuitive auditory icons (recognition rates 9%-21%) improved their scores to 31%-46% as a result of the field study learning. These results indicate that ***auditory icons should be preferred to earcons for users to gradually learn to recognise mobile service notifications through natural usage (even if the sound-service associations are somewhat poor)***.

We believe that everyday technologies for the general public, such as the mobile phone, should be designed to require as little training as possible. It is unreasonable to expect that users will spend the time and effort to go through explicit and intense learning of auditory notifications. For just 10 associations, participants failed to reach 90% accuracy with earcons, even after six rounds of rigorous training. With mobile services and widgets on the increase, it would be impractical, if not impossible, to follow such training procedures.

Memorability

In the literature, accuracy scores in identification of sound-referent associations seem to improve or remain the same in retention tests, performed three days to two weeks after training. This ceiling effect in the forgetting rate is probably a combination of the simplicity of the alerts learned and the number of days separating the two tests. We expected that notifications for mobile services (which represent more complex concepts than alerts such as 'low fuel' or 'missile launch') would be easier to forget over time. Indeed, both notification types we tested were partially forgotten one week after training, and even more after four weeks. However, similar to the intuitiveness and learnability results, earcons seem to be a less appropriate choice for mobile service notifications, since their meaning is easier and quicker to forget. Auditory icons were almost perfectly remembered (with 97% accuracy) after one week, and still performed marginally better after 4 weeks (91%) than earcons did immediately after the Lab 2 training (89%). On the contrary, earcons' success in retention one week after the training dropped to 62%, which was lower than the score achieved on the first trial of training. Four weeks later, performance dropped further to 43%, which was even lower than the auditory icons' initial intuitiveness score in Lab 1. In a real world scenario the number of services could be more than just the 10 that were tested here, but even more importantly, some of them could occur infrequently. Even if users undertook extensive learning to memorise the associations, if an earcon notification occurred just once a week, the likelihood is that the association would be easily forgotten. Therefore, ***auditory icons should be preferred to earcons for sound-service associations to be retained in memory (especially for notifications that do not occur many times per week).***

Preference

None of the sounds we presented in this study enthused our participants, as they all scored below average in the relevant 5-point Likert scale. This means that they would not want to have any of them as mobile service notifications, even in contexts that the audio notification modality was welcomed. There are at least three reasons we can identify that caused this negative trend. First, our sounds could have been aesthetically unpleasant to our participants. Although both sets of sounds were produced through processes that involved end-users, it is possible that further or better design was required to produce desirable sounds. However, most users today utilise very few notifications on their mobile phones, such as incoming calls, messages and self-reminders. It is therefore likely that the low ratings our notifications received were expressing a criticism towards (or a lack of appreciation of) the use of further notifications. Throughout the study we recorded numerous remarks that notifications for these kinds of services would not be welcomed. The third reason for the negative preference score could arise from the fact that there was no face value for the notifications participants received on their mobile phones during the field study. Even if some participants had a positive view towards receiving more service notifications and liked some of the sounds, they were constantly interrupted by them without the benefit of using the corresponding service (i.e. the service notification was by default irrelevant). This could have caused annoyance that was then reflected in the low preference scores. This is further supported in the case of earcons, as we observed a drop in their ratings during the field study and the Lab 2 experiment that followed.

Regardless of the overall negative ratings, earcons scored significantly lower than auditory icons in subjective preference, throughout the four stages of the study. This contradicts findings by [Sikora et al., 1995; Roberts & Sikora, 1997], where musical notifications were found as more appropriate than auditory icons in a business environment, and by [Jones & Furner, 1989, as reported by Graham, 1999, and Lucas, 1994]. In fact, in our study, earcons seemed to cause strong negative feelings and frustration to some participants. This was apparent from involuntary comments during the training, explicit responses in the questionnaires and comments during the debriefing focus group after the Lab 2 session. For example, earcons were characterised by some as “horribly discordant”, “ugly” or “miserable” and one even stated, “I loathe this sound” (earcon 10). It is worth noting that the same earcons were generally received in a positive way in our previous experiment (Section 5.2). A possible explanation for this (extra) negative attitude towards earcons could be the inefficiency of the learning process during the field study. As one participant put it “[I] came to hate them all because I got them wrong so they just irritated me, which was a kind of vicious circle!” However, this is not supported by the fact that earcons 8 and 9 were preferred more than the rest throughout the study but actually were found to perform slightly worst. On the other hand, we noticed more engagement and even a level of excitement towards auditory icons. Participants were happy to confirm to each other the meaning of the familiar everyday sounds, and explore the meaning of the less familiar ones. Therefore, we conclude that ***familiar sounds that ‘make sense’ should be used for auditory notifications to be liked and adopted by users***. There is evidence to indicate that sound familiarity had also a positive effect on earcons’ preference too. The earcons that were preferred the most (notifications for ‘User-to-Self’ services) were the only ones to use a vibraphone timbre, which is more similar to how mobile phones commonly sound today (in contrast for example to a piano). However, this preference could be attributed to other factors (such as the service category) and further studies need to be carried out to test this assumption.

Finally, from the context data gathered during the field study we found that participants consistently preferred auditory icons regardless if they were working, commuting, socialising or being at home. However, some anecdotal data describe auditory icons to be "too intrusive" and potentially "embarrassing to broadcast loudly". On the other hand, only earcons were actually found to be preferred less when participants were in the presence of other people. This social concern was also highlighted by a participant commenting (after the field study) that he would "think less of someone" if they had earcon notifications. We therefore rephrased the preference question from "would you like to have this notification on your mobile phone" in the laboratory and field studies to "would you like to receive this notification in the presence of others" for the web-based experiments. We found that the responses still favoured the auditory icons. We therefore can conclude that ***auditory icon mobile service notifications are more likely to be adopted by users over earcons in a variety of social contexts***.

Limitations and Future Directions

As with any other study in auditory interface design and evaluation, it is possible that the results from our comparison study have been influenced by the particular stimuli we used as representatives of the two sound types. However, we designed and employed analytical and empirical processes to ensure that we minimised any such bias. The effectiveness and improvement of these processes is discussed in the next section.

One of the limitations of our study is with regard to the memorability metric. Our intention was to compare how much notification associations were forgotten once they had been learned perfectly. However, the Lab 2 training ended with four participants never reaching 100% accuracy for the earcon notifications. Although perfectly retained notifications should score as well as they did at the end of the training, the associations that were not learned could have interfered with identifying some of the learned associations too. It is therefore possible that the steeper forgetting rate of earcons was partially due to the incompleteness of the learning. However, the difference between the sound types in memorability is so much larger than the difference in the final Lab 2 scores that we do not believe that our conclusions are largely compromised. Future studies could address this issue by extending the training process until earcons are perfectly learned too. Caution should be used though due to the fact that if participants are exposed to one type of notifications for significantly longer, it is likely that this extra familiarisation will have a positive bias on memorability performance. In our study we were not only limited in resources to continue training further, but we also observed great impatience growing in the participants who were having difficulties in learning the earcon notifications. Regular breaks could therefore be needed, which would put extra strain on the resources, and further defy the minimal training approach we aspire for mobile service notifications.

There were also some technical faults that have probably affected our observed data to a small degree. In the field study data, we observed a sudden drop in the performance of both notification types around day 4, and a slight drop on day 7 (Figure 6-6). This can be explained by a server problem that significantly reduced the number of notifications delivered on day 4 and slightly reduced the responses on day 7. With fewer questions asked on a particular day, each answer had more effect on the average of that day. As incorrect answers were more likely to occur at that early stage of learning, the effect was to depress the averages. If these faults had not appeared, we would expect the field learnability data to climb steadily and approximately in the gradient defined by the plot points of days 1 and 6. Some reduced responsiveness was also observed on a Sunday and less so on Saturday and on a public holiday. Perhaps participants felt less willing to participate during their leisure time. Nevertheless, the number of the responses did not seem to affect the data on those days.

Another limitation in the field data collection was caused by the conflicting requirements of gathering all 20 responses from each participant on any given day, and avoiding causing too much annoyance. If participants chose to ignore any pair of incoming notifications, the frequency of their appearance would have to increase in order to successfully get responses for all 20 sounds. Therefore, if someone consistently ignored the task, the effect would be for the requests to be condensed into a small part of that day. However, if notifications became too frequent, apart from causing annoyance (which could affect the quality of the responses), it would also defeat the goal of the field study. Since we wanted to imitate the everyday environment of learning through opportunistic use, continuous 'sessions' of responding and learning were undesirable. We decided to address this trade-off by restricting the frequency of notifications to a maximum of every 30 minutes. As a result, there were participants that could not respond to all sounds on a given day (the cut-off time was at 9pm). In some rare occasions this disrupted the buffer functionality of the application and some participants received the same sounds twice in one day. Nevertheless, reduction or repetition of notifications due to these problems were evenly distributed between the two notification types, and thus should not have affected the comparison results.

A final limitation of the field application was that the sound quality had to be significantly reduced due to the memory restriction on the phones. It was observed (both by us and some participants) that the acoustically more complex auditory icons were affected more than the earcons. Many participants reported that the low quality of the auditory icons made them annoying, in contrast to the high quality sounds during the lab sessions. Despite this quality bias in favour of earcons, participants still preferred the auditory icons during the field study.

This highlights the most interesting area for further research, as there is still a lot to explore in relation to user preference of auditory notifications. For example, why did our results contradict those of [Roberts & Sikora, 1997]? What is the effect of the participants' demographics, the nature of the functions/events to be notified about and the context of use? Furthermore, what would the user preference of speech notifications be like in longitudinal studies such as the one presented here? Unfortunately, limitations on resources, such as time, costs and participant sample size restrict us from answering these questions for the time being.

6.2 Effectiveness of the Notification Design Process

The results of the study presented in the previous section are valuable not only for producing specific conclusions on the effectiveness of non-speech mobile service notifications, but also for producing guidelines for their designing methodologies. This is achieved by identifying and comparing patterns in the responses of the study above

Service	Sound Description	ID (%)	REP Votes (%)	REP Rating (1-5)
1	BBC news ident	77	85	4.3
2	Stadium sounds	96	69	4.1
3	Public announcement at an airport	97	67	3.4
4	Beginning of a movie (20th Century Fox)	91	58	3.9
5	Audience clapping/applauding	88	65	3.0
6	Old-fashioned phone ringing	100	40	4.5
7	Message transmitted in Morse code	84	61	3.9
8	Mobile phone notification (Windows Mobile)	64	45	3.8
9	Truck/lorry reversing	79	36	3.5
10	Wind chimes	79	58	2.5

Table 6-5: Identification scores (Survey 1 – Section 5.1.1) and the two measures of representativeness (Survey 2 – Section 5.1.2) for the 10 auditory icons representing the 10 services

(Section 6.1) and the user reactions during the design methodologies (Chapter 5). This comparison process is essentially providing with an assessment of the methodologies, which is then translated in a set of design guidelines. Furthermore, these guidelines could also prove useful for auditory interface designers of domains similar to the mobile service context of use. The notification design methodologies for auditory icons and earcons are summarised and assessed separately below.

6.2.1 Guidelines for Auditory Icon Design

The process for creating the auditory icons consisted of four steps. First, a brainstorming session amongst researchers produced multiple ideas of everyday sounds that could represent each service cluster. Then, instances of sounds were collected or recorded, and shortlisted based on their overall quality. The two final steps involved end-users in the process, in the form of online surveys. Survey 1 assigned identifiability scores and filtered out the least identifiable sounds. The remaining sounds were put through Survey 2, which gathered votes and ratings in relation to their representativeness for the 10 services. The three measurements for each one of the final 10 auditory icons are summarised in Table 6-5.

There is evidence in the literature to suggest that the more direct this signal-referent is, the more learnable auditory icons become (see Section 3.4). This was confirmed in our study too, as auditory icons were found significantly easier to learn than earcons. However, even auditory icons that were only recognised at chance rate from naïve listeners were learned during Lab 2 training as easily as the more intuitive ones. This ceiling effect, however, was not present for all auditory icons in the less effective and casual field study learning. Studying the ease of learning for those auditory icons, we found no significant correlation between the initial intuitiveness score and the amount of learning that took place during the week. However, we observed that auditory icons ‘20th Century Fox clip’ and ‘audience clapping’ (services ‘Entertainment

Downloads' and 'Entertainment Live' respectively) were harder to learn than the average performance across auditory icons. In fact, the first one achieved a 2% lower recognition score after the field study (at 60% accuracy). This low score was almost exclusively due to its confusion with the other entertainment service. Similarly, the learnability of 'audience clapping'-'Entertainment Live' association was partially restricted due to its confusion with 'Entertainment Downloads'. This leads to our first guideline for designing auditory icons as mobile service notifications:

1. ***Avoid presenting clusters of services whose content is conceptually very similar.***

The main requirement, and often challenge, for auditory icons is to link them to the concept they represent in a meaningful way, so that they are intuitively recognised and easily learned as notifications. Apart from the exception mentioned above, auditory icon learnability and memorability results seemed to be otherwise unsurprising. Therefore, the rest of our analysis is focused on the intuitiveness results and delivers guidelines for designing auditory icons that are easily recognised.

Our auditory icons were correctly assigned to their services in Lab 1 at a 53% rate but with a wide range of success, from 9% to 97%. As it can be seen from the intuitiveness breakdown chart (Figure 6-7), there were three auditory icons that performed well above average (over 80%) and three that performed well below (at or under 20%). In particular, the generic Windows Mobile reminder sound for 'Self-reminders' (service 8) performed at almost chance rate at 9%, as it was mostly (and wrongly) assigned to the 'Incoming SMS' service instead (see discussion in 7.1.3). This failure could have been predicted from the results of the design process. First, looking at the identification survey results, this sound was identified by only 64% of the participants, which was the lowest score amongst the selected sounds for the representativeness survey. Furthermore, in revisiting the identification judgments for this sound, we observed that its score was actually unfairly high. Responses that identified it as an interface event, such as "computer sound, to attract the user's attention to

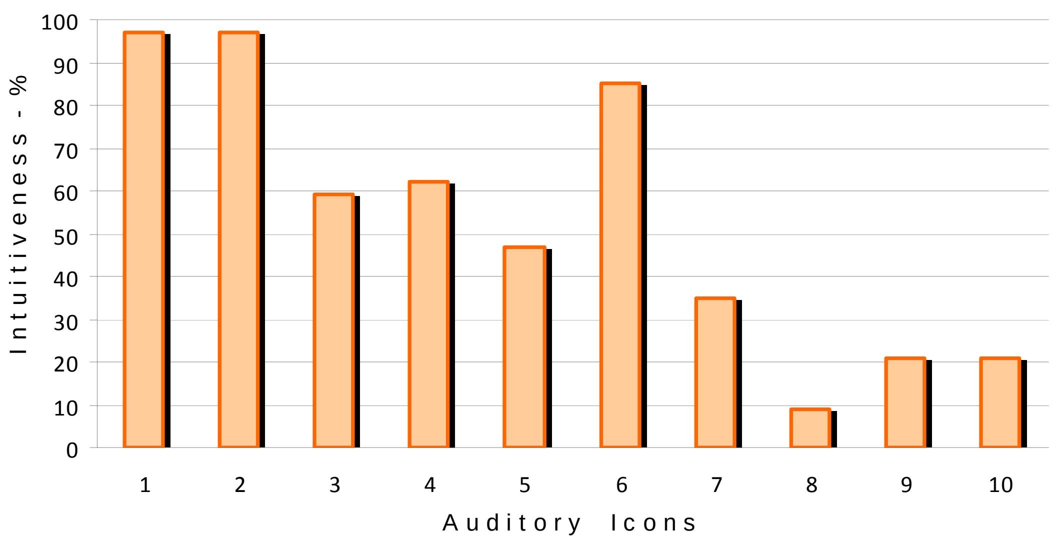


Figure 6-7: The intuitiveness scores achieved for each auditory icon during the Lab 1 experiment

a dialog or message", were marked as correct identifications. This is an obvious misalignment from our adaptation of the literature guidelines. Our intention was to regard as incorrect any acoustically precise alternative interpretations that failed to describe the elements that were utilised by the underlying metaphors. The Windows Mobile reminder sound was suggested during the brainstorming session from a Windows Mobile user, who was aware that this sound is used specifically for reminders. If it is perceived by others as any interface event, its particular semantic relationship is lost in translation. In fact, 21% of the identification responses were specific to email or message notifications (which we wrongly counted as correct), and 56% of Lab 1 guesses were assigned to 'Incoming SMS' service. A similar pattern was observed for the 'audience clapping' sound, where about 10% of the identifications were related to sporting events (we rightly counted as incorrect), and 41% of the guesses were assigned to 'Sports Information' (but the correct service still gathered more correct guesses at 47%). These observations underscore two important guidelines for the identification process:

2. ***Consider the whole set of services and suggest auditory icons that cannot be easily associated with more than one service.***

3. ***In everyday sound identification studies, judge responses on their success to describe the metaphors that auditory icons will utilise to link the sound to the service.***

The correlation between low identifiability (in Survey 1) and poor intuitiveness (in Lab 1) also holds for the next three least identifiable auditory icons, with one exception. Both 'truck reversing' ('Backup Reminders') and 'wind chimes' ('Other Services') had 79% in identification score and 21% in intuitiveness score. Also, the 'morse code message' ('Incoming SMS Messages') was the fourth least intuitive auditory icon (35%), with the next higher score in identifiability (84%). However, the 'BBC ident' (News Information) was identified at 77% of the times but still achieved the best representativeness and intuitiveness scores. This seeming contradiction is probably attributed to the demographic differences observed between the population samples assigning the three scores. The identification survey was more ethnically diverse with most responses coming from Greeks (31%). However, they failed to identify British-specific sounds (such as the news idents) significantly more frequently than British nationals. In particular, the 'BBC ident' was recognised by all British respondents (or UK residents). Furthermore, there were more British nationals and UK residents in Survey 2 (representativeness score) than in Survey 1, and all of the participants in the evaluation study (Section 6.1). This highlights the importance of soundscape cultural influences to auditory icon design. Also (excluding the 'BBC ident' sound for the reasons mentioned) the least intuitive auditory icons were indeed the least identifiable sounds in Survey 1. Therefore, we can suggest two more important guidelines with regard to the identification process:

4. ***Auditory icons are culture-specific, and should be designed with this in mind.***

5. ***In auditory icon design, select everyday sounds that are identified by at least 80% of a random sample of the intended end-users.***

However, weak identifiability cannot account alone for the poor intuitiveness performance of some auditory icons, as the correlation breaks down for highly identifiable sounds (approximately over 80%). Even though we attempted to mark responses of Survey 1 as correct if the underlying signal-referent metaphor was (unknowingly to the respondent) identified, not all auditory icons were immediately recognised by all Lab 1 participants. For example, the 'public announcement' sound ('Here & Now Information') was identified by 97% of the respondents but achieved only 60% intuitiveness (far from the 97% of other auditory icons). This probably means that the metaphor linking the sound to the 'Here & Now Information' services was somewhat poorer. Our intention was to measure the strength of these metaphors by the representativeness scores of Survey 2: number of votes and ratings. As three sounds were suggested for each service, the number of votes for every sound was the amount of times participants chose that sound to represent the service. The ratings were obtained by a 5-point Likert scale, where participants indicated the level of representativeness for the sounds they selected for each service.

However, the representative votes measurement proved to be an unreliable intuitiveness predictor, due to a design compromise. We decided to divert from the soundimagery process that displays all alert functions on the screen simultaneously [as reported by Simpson, 2007], and instead displayed one service at a time. This was done because the services represent concepts that are more complex than straightforward alerts. In our design, respondents could concentrate on the description and examples of each service cluster before they heard the candidate sounds. However, the fact that the number of votes for each sound was based on comparisons means that the obtained scores are not very useful as absolute numbers. Sounds with strong within-service competition are bound to score lower in representativeness even if they utilise a relatively strong metaphor. For example, 'old-fashion phone ringing' achieved over 80% in intuitiveness but gathered only 40% of the votes, as it was competing to the very similar 'telephone ringing as heard through the telephone line'. The lesson we learned from these contradictive findings can be translated to a straightforward guideline for the representativeness process:

6. ***Avoid suggesting alternative auditory icons that utilise very similar metaphors.***

Apart from good metaphors scoring low in votes, sounds with poor metaphors but even poorer competitors could still score high in number of votes. We intended to capture such inaccuracies through the representativeness ratings. Since these ratings were not given in relation to the other suggested sounds, they should provide a more reliable value as a metric for absolute representativeness (although it is a possibility that exposure to the other sounds influenced respondents' ratings). Indeed, we observed a relatively strong correlation between intuitiveness in Lab 1 and representativeness ratings in Survey 2 ($r=.7$, $p<.05$).

However, there is again a relative compromise in the absolute value of this measurement, as we observe the correlation to weaken for auditory icons with lower intuitiveness scores. For example, the least intuitive sound ('Windows Mobile reminder') still achieved 3.8 out of 5 in the rating scores, marking it as a good representation for the 'Self-reminders' service. Similarly, auditory icons 7 and 9 had intuitiveness scores of 35% and 21% respectively, but still were rated as fairly representative (3.9 and 3.5 respectively). One possible explanation for this unpredictably higher rating is that the underlying link between sounds and services may have become apparent to respondents only after (or because) the sound was suggested for the particular service. Furthermore, there seems to be a positive bias in the ratings, as the overall score was above average and respondents rarely rated the auditory icons they selected as not representative. This is further supported by the fact that auditory icons with high ratings received relatively lower number of votes and had poor intuitiveness. In other words, if a metaphor is not very good, it is not going to be popular, but people will judge it as a good representation of the service once they select it amongst the other options.

Therefore, it seems that none of our three measurements (identification, representativeness votes and representativeness ratings) can adequately predict auditory icon intuitiveness when tested by naïve listeners. However, the strengths of each measure are of somewhat complementary nature. Low identification score is a good predictor for poor intuitiveness, but higher identification scores need to be moderated by representativeness ratings to successfully predict intuitiveness. Similarly, high ratings need to be corrected by the number of votes, provided that guideline 6 is applied and no similar metaphors compete for the same service.

Since our three measurements are complementary in their ability to predict intuitiveness, we should expect to obtain a more objective measure for auditory icons' potential success in immediate recognition, if we multiplied the three scores. Indeed, this new total score (which we name P(I): Predictor of Intuitiveness) is very strongly correlated to intuitiveness scores from Lab 1 ($r=.87$, $p<.001$). This is also obvious by the graphical comparison of P(I) with the intuitiveness scores (multiplied by 300 in order to match the magnitude of P(I)) per auditory icon (Figure 6-8). Therefore, we can suggest:

7. ***A combination of scores on everyday sound identification, sound-service representativeness votes and ratings (as obtained by the methodology described in Section 5.1), can be used as a good predictor of auditory icon intuitiveness.***

The 3 auditory icons P(I) seems to be less successful are those for services 6, 7 and 8. As already discussed, the Windows Mobile sound (service 8) goes against our suggested guidelines 2 and 3. We expect that P(I) would be more successful in this case, if it did not compete with another service, and if its identification score was calculated correctly. Furthermore, the 'old fashioned phone ringing' sound (service 6) defies guideline 6, as it was competing against the 'telephone ringing as heard through the telephone line' sound, which essentially utilises the same metaphor to represent 'In-

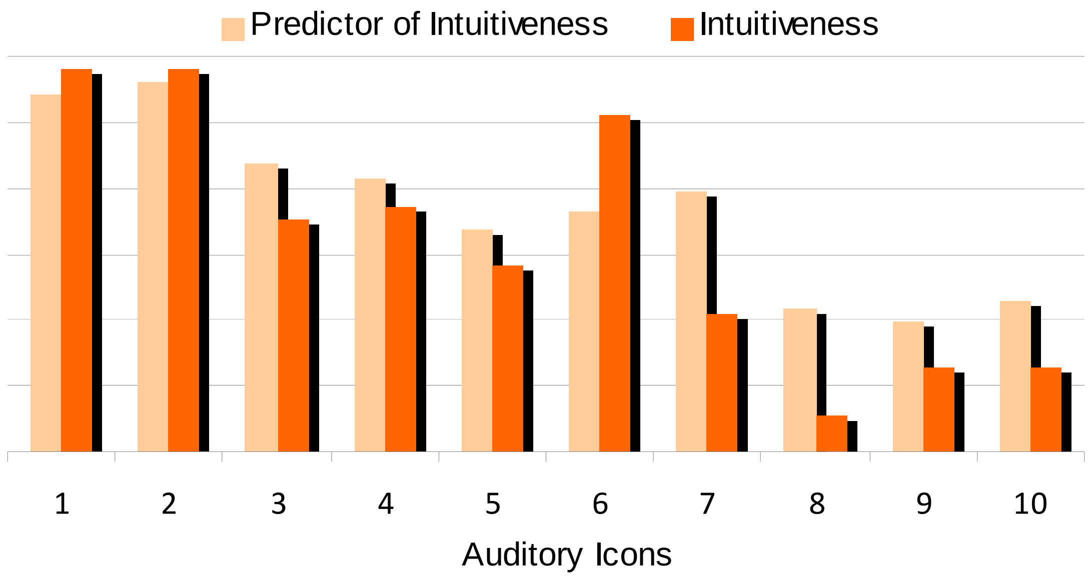


Figure 6-8: Predictor of Intuitiveness (P(I)=ID x Votes x Ratings) compared with Intuitiveness scores obtained from Lab 1 (multiplied by 300).

coming Calls'. However, P(I) failed to predict intuitiveness for 'Morse code' auditory icon (service 7) for no apparent reason. In all 3 measurements, it achieved scores within a 4% margin from the average of all auditory icons, yet its intuitiveness was almost half than what predicted. It was observed from the participants' responses in Lab 1 that 'Windows Mobile notification' sound was consistently assigned to 'Incoming SMS', perhaps forcing 'Morse code' to be assigned elsewhere. In fact, we found it was often assigned to 'News Information', perhaps because the sound was associated to the method news used to be transmitted at the days of the telegraph. However, unlike 'Windows Mobile notification' and 'audience clapping', this was not indicated in the identification survey responses.

This highlights again the limitation of the representativeness voting procedure. If in Survey 2 all services were candidates for all the sounds, we would expect the measure of representativeness and therefore intuitiveness to be stronger. First, sounds would not be suggested for particular services and thus reveal the designers' choice of metaphors. Upon these suggestions, respondents may find the association reasonable but end-users could have difficulty recognising them later on. Second, if a sound is a relatively good representation for more than one service, this will become immediately noticed by the number of votes it is given for the different services. However, this presenting all services at once would take the focus away from understanding the particular concept of each service, and respondents could misunderstand their content. Alternatively, representativeness ratings could have been collected for all 3 sounds suggested for each service, and thus increase the accuracy and reliability of the measurement. Both of these alternative designs need to be tested and balanced against the extra resource costs they will impose. Nevertheless, we can cautiously suggest:

8. ***In auditory icon associability studies, consider presenting all services as candidates for all auditory icons, depending on the conceptual complexity of the services (and the associated resource constraints).***

9. ***Depending on resource constraints, consider retrieving representativeness ratings for all possible (or reasonable) associations***

6.2.2 Guidelines for Earcon Design

We did not explicitly attempt to adapt the earcon design process from what has been suggested and followed in the relevant literature [e.g. Brewster et al., 1993]. It is a systematic methodology that has already been established, tested and refined. Following the relevant guidelines, we focused our efforts in producing earcons that were highly distinguishable, structured in a hierarchical way to represent the categories of the service classification, and also be pleasant. Furthermore, we empirically evaluated them and attempted to improve them through iterative design and with the assistance of a professional musician. Finally, some metaphorical association was attempted when possible but eventually we were able to apply a semantic link to only 2 earcons: ‘Entertainment Downloads’ earcon had a melody that was distinctively dropping in pitch, and ‘Incoming Calls’ earcon had a short melody that was repeated twice.

Compared to the design process of auditory icons, it is more difficult to draw clear conclusions and guidelines for the equivalent earcon process. First, there was not great or unexpected variance within earcons’ results in intuitiveness, learnability, memorability and user preference. The score of each earcon for each stage of the study is illustrated in Figures 6-9 and 6-10. As it can be seen from these figures, earcons performed fairly uniformly throughout the study. Second, for the small discrepancies occurred for some earcons, it is more difficult to understand and explain the reasons behind these differences. In other words, we acknowledge that we do not have the musical tools and experience to interpret the results in terms of musical composition. This actually highlights one of earcons’ disadvantages, as musical expertise is required for their design and in-depth interpretation of their evaluation. In fact, the improvement during iterative design and through the assistance of the professional musician failed to reach significance.

One of the observations we can make is that earcon 10 performed consistently better than the rest. Service 10 was the only one that was represented with a unique timbre (violin), and this probably set it apart and made it easier to remember. This is in line with literature findings, which indicate that timbre is the most effective earcon parameter for distinguishability [e.g. Brewster et al., 1993]. A further interesting finding with regard to earcon 10 is that it was not learned throughout the field study. In fact, there was a drop in performance between Lab 1 and Lab 2 measurements. On the contrary, once the associations and structures were explained, earcon 10 was learned and remembered better than the rest. Similar trends was observed for earcons 3 and 7, which had a minimal improvement of less than 8% during the field learning, but performed over 90% and 80% respectively in the second trial of Lab 2 training. This, once again highlights the importance and necessity of formal earcon training. Furthermore, the difference in type of training between the earcon experiments in Section 5.2 and the comparison study in Section 6.1 made a difference in learnability. After the three testing rounds of the designing experiments earcons reached 61% recognition accuracy, while after the first three trials of the Lab 2 training earcons had reached 81% accuracy. This stresses the importance of the high-intensity training needed for learning earcons. The above points can be summarised in our first two guidelines for earcon design:

10. ***Use iterative design for earcon notifications to improve their distinguishability.***

11. ***A combination of tutorial, active learning and trial-by-trial feedback should be used in the iterative evaluations/training of earcons.***

Even after extensive training, however, earcon 9 and (to a smaller extend) earcon 8 performed consistently worse than the rest. Although earcon 8 had the biggest improvement during field learning, it did not reach beyond 80% during the training and it also demonstrated the steepest forgetting curve. Earcon 9 had the worst by far performance in Lab 2 and remained below all earcons' scores during the training, failing to reach over 73%. Since they shared the same timbre and therefore were more alike, one possible explanation could be that they were confused for one another. This was partially supported from our data, with earcon 9 being assigned to service 8 about 1/3 of the times. This is somewhat surprising because the two sounds were different in many dimensions, such as monophony/polyphony, musical key, number of notes and their duration. Furthermore, one featured a melody with continuously ascending tones, while the other had a descending melody (see notations in Table 5-4). Another pair that was frequently confused was earcons 4 and 5, although they were different between them in all dimensions apart from timbre and polyphony. Anecdotal data confirmed that both of these pairs were confused because of the difficulty participants had to differentiate between the ascending and the descending melodies, although the difference from lowest to highest note (and vice versa) was about two octaves within all four earcons in question. Therefore, it is suggested that:

12. ***When ascending and descending melodies are utilised for earcon distinguishability, the difference in pitch between the highest and lowest notes should be steep and more than two octaves.***

Assigning earcons to the correct categories was done fairly successfully after the timbre-category associations were explained. In the first trial of Lab 2 participants assigned earcons in the correct category 89% of the times, but still did not exceed 93% by the end of trial 6. Earcons 6 to 9 were confused the most and also demonstrated a steeper forgetting curve with regard to their category membership. At the same time, earcons in the designing experiment of Section 5.2 reached no more than 85% category level accuracy.

In terms of user preference, findings were overwhelmingly against earcons, even though they were positively received in the designing experiments. In particular, 75% of the participants in Experiment 2 (Section 5.2.2) expressed the willingness to have at least some of the sounds (where ‘some’ was the middle point in a 5-point Likert scale from ‘all’ to ‘none’). However, the same response to this question was given by only 30% of participants in Lab 1 and 13% in Lab 2 (and remained roughly the same for Web 1 and Web 2 sessions). Furthermore, when participants asked specifically if they would like to have each individual notification, they responded positively in 25% of the times in Lab 1 and 9% in Lab 2. The drop from the Experiment 2 to the comparison study could be attributed to three reasons. First, participants of the comparison study had a more difficult, and thus frustrating, task to accomplish in Lab 1 and field study, where no or casual training was provided. This difficulty possibly led them to have a more negative position towards earcons. Second, the significant drop during and after the field study indicates that people became more annoyed by earcons when receiving them on a daily basis. It is worth noting again that there was no drop in user preference data for auditory icons. Third, it may be that preference to certain notifications is judged relative to the ones they are compared to each time. The simple tones of the designing experiment were extremely difficult to learn, as they had no melody in them. Under this light, earcons must have sounded like music to the ears of our participants! However, when compared to the familiar everyday sounds, the musical notifications were judged more negatively. Frequently participants mentioned that they found the associations counter-intuitive and would have liked to reassign the earcons or change their melodies (e.g. some were found “too serious” or “too sad” for the services they represented). Therefore, we suggest:

> ***13. When possible, involve users in the design of earcon notifications and their assignment to services.***

Finally, we had previously found that preference questions with regard to potential real usage of audio types were inconclusive and unreliable when asked during a laboratory setup (see experiment in Section 4.1). Responses on how much notifications were “liked” where different from responses on the “would you like to have them” question. However, in the current study we found that participants were consistent in their preferences throughout the laboratory and field studies. Therefore, our final guideline is formed:

> ***14. Trust can be placed in users’ initial responses in willingness to use different types of notifications, as they are unlikely to change through usage.***

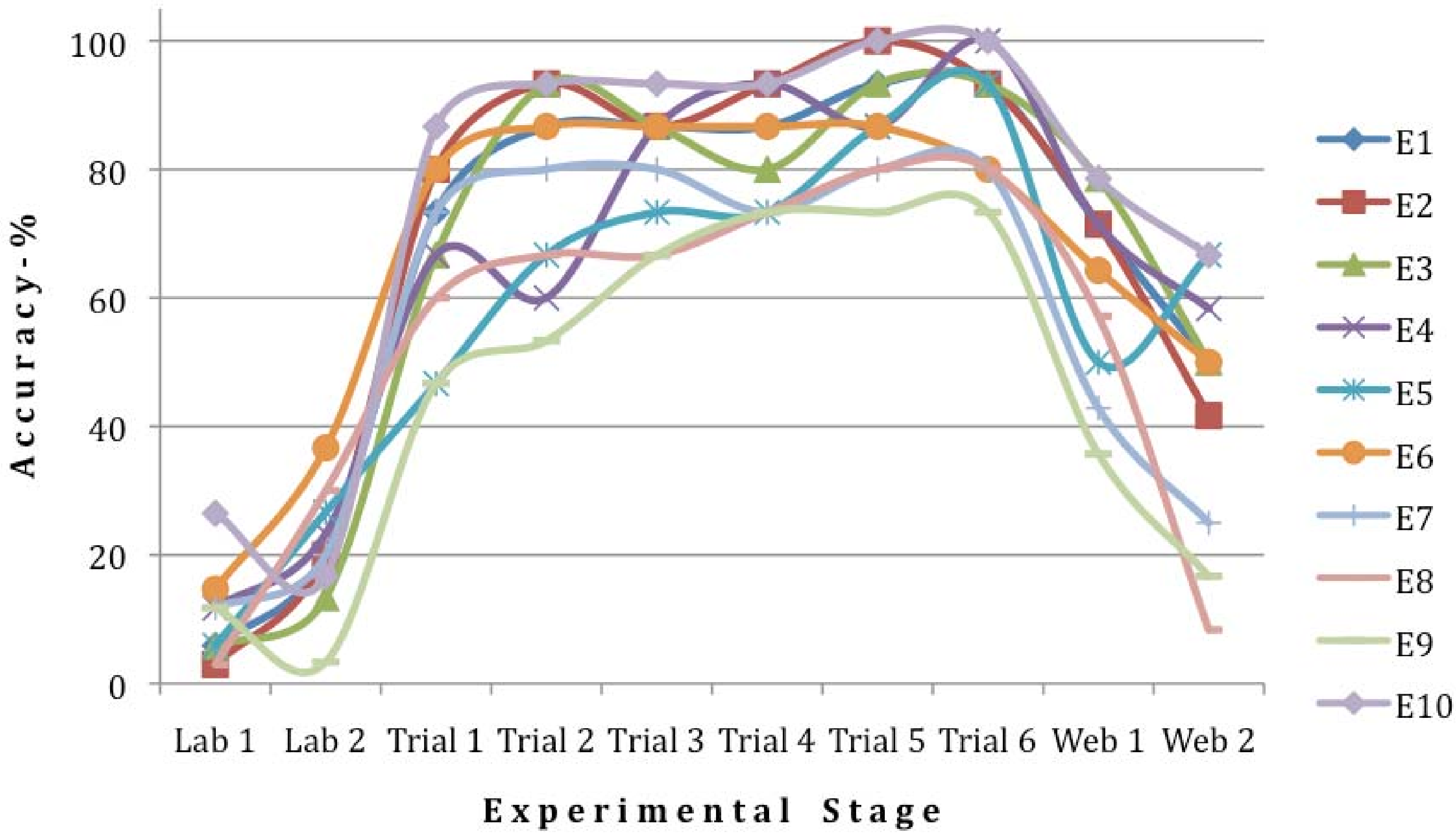


Figure 6-9: Performance of each earcon across the Lab and Web studies (X-axis). Trials refer to the Lab 2 training stage.

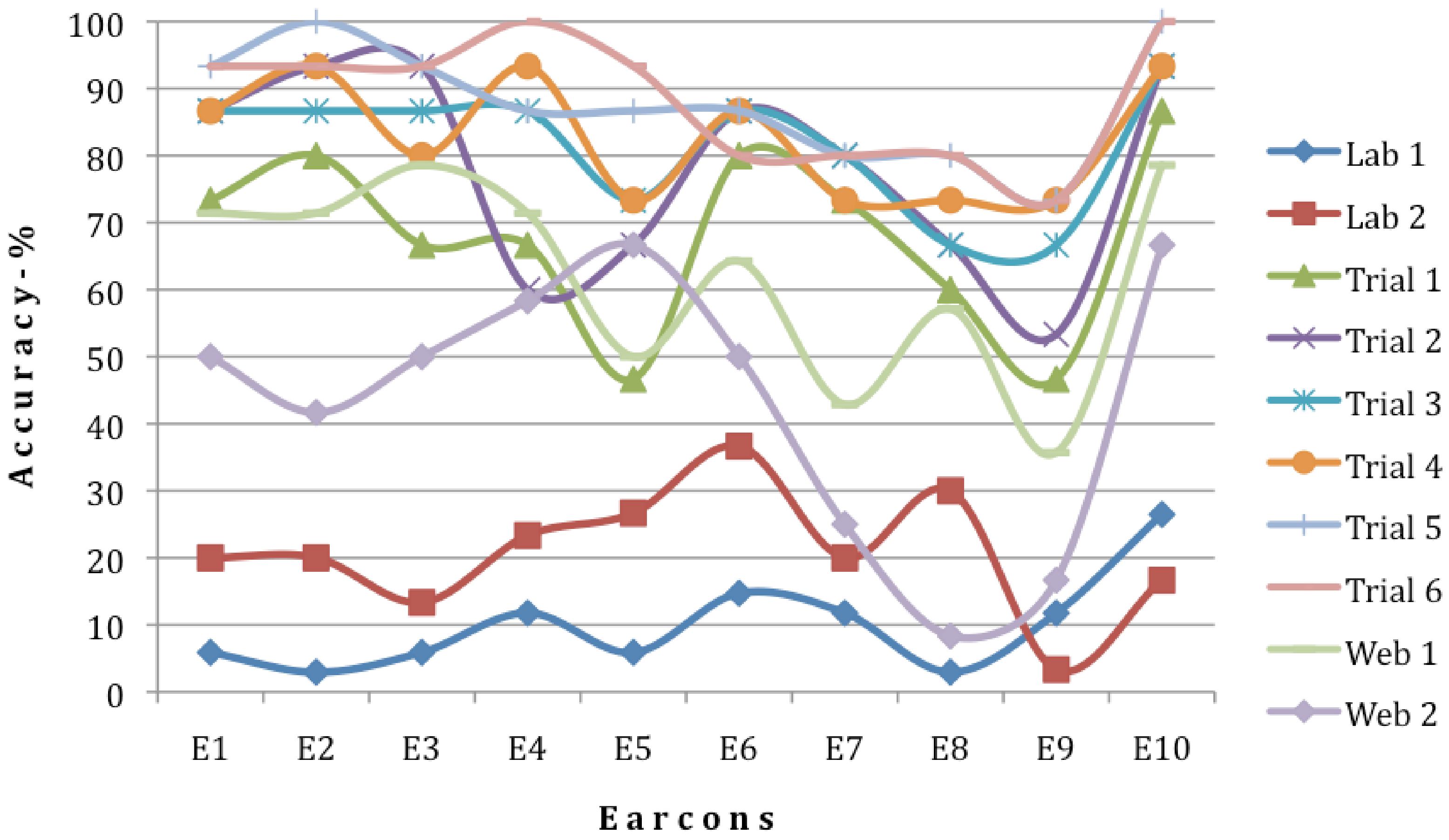


Figure 6-10: Performance of each earcon (X-axis) across the Lab and Web studies. Trials refer to the Lab 2 training stage.

6.3 Chapter Summary

In this chapter we have concluded our journey in mobile service awareness, and presented the outcomes of our research on mobile service auditory notifications. In the first part of the chapter, we presented an extensive evaluation study comparing the effectiveness of auditory icons and earcons as mobile service notifications. Design decisions to maximise the efficiency of the study were made after reviewing the relevant literature, and based on the experiences gained and lessons learned in our previous experiments. As a result, four metrics of notification effectiveness were identified and investigated in a 4-stage longitudinal study. In particular, intuitiveness of the notification types was measured in a laboratory study with naïve listeners; learnability was monitored in the natural context of mobile notifications use during a weeklong filed study, and in a subsequent laboratory experiment; memorability was determined through two web-based experiments one and four weeks after training; and finally user preference data were gathered implicitly or explicitly throughout all stages of the study.

Two distinct contributions were made from the analysis of the results of this study. First, the effectiveness of auditory icon and earcon notification was compared, and conclusions were drawn for the two auditory notification types for mobile services. Auditory icons were found to perform significantly better in all four metrics, and therefore were more effective as notifications. The results were translated in a set of specific suggestions for mobile service auditory notifications, which can also be useful in any auditory interface design in contexts similar to mobile phone usage. In particular, we concluded to the following suggestions:

1. ***Underlying sound-service metaphors should be utilised (such as in auditory icons) for mobile service notifications to be immediately recognised.***

2. ***Long-standing and widespread metaphors need to be utilised for sound-service links to be intuitive.***

3. ***Strong sound-service metaphors should be utilised to avoid users creating alternative semantic relationships to other services.***

4. ***Extensive training is absolutely necessary for arbitrary sound-service associations (such as in earcons) to be learned. Yet, there might still be a significant percentage of users who will be unable to learn all of them perfectly.***

5. ***Auditory icons should be preferred to earcons for users to gradually learn to recognise mobile service notifications through natural usage (even if the sound-service associations are somewhat poor).***

6. ***Auditory icons should be preferred to earcons for sound-service associations to be retained in memory (especially for notifications that do not occur many times per week).***

7. ***Familiar sounds that 'make sense' should be used for auditory notifications to be liked and adopted by users***

8. ***Auditory icon mobile service notifications are more likely to be adopted by users over earcons in a variety of social contexts.***

We are not aware of other studies that have systematically investigated the effectiveness of non-speech auditory notifications to the extent presented here, or any studies that have derived specific suggestions for mobile service auditory notifications.

The second contribution presented in this chapter was in relation to the notifications design processes. By comparing the effectiveness of the specific notifications we designed, we can assess the methodologies that produced them. By identifying the strengths and weaknesses of these processes for auditory icons and earcons (Sections 5.1 and 5.2 respectively), we have derived a set of design guidelines for mobile service notifications:

1. ***Avoid presenting clusters of services whose content is conceptually very similar.***

2. ***Consider the whole set of services and suggest auditory icons that cannot be easily associated with more than one service.***

3. ***In everyday sound identification studies, judge responses on their success to describe the metaphors that auditory icons will utilise to link the sound to the service.***

4. ***Auditory icons are culture-specific, and should be designed with this in mind.***

5. ***In auditory icon design, select everyday sounds that are identified by at least 80% of a random sample of the intended end-users.***

6. ***Avoid suggesting alternative auditory icons that utilise very similar metaphors.***

7. ***A combination of scores on everyday sound identification, sound-service representativeness votes and ratings (as obtained by the methodology described in Section 5.1), can be used as a good predictor of auditory icon intuitiveness.***

8. ***In auditory icon associability studies, consider presenting all services as candidates for all auditory icons, depending on the conceptual complexity of the services (and the associated resource constraints).***

9. ***Depending on resource constraints, consider retrieving representativeness ratings for all possible (or reasonable) associations.***

10. ***Use iterative design for earcon notifications to improve their distinguishability.***

11. ***A combination of tutorial, active learning and trial-by-trial feedback should be used in the iterative evaluations/training of earcons.***

12. ***When ascending and descending melodies are utilised for earcon distinguishability, the difference in pitch between the highest and lowest notes should be steep and more than two octaves.***

13. ***When possible, involve users in the design of earcon notifications and their assignment to services.***

14. ***Trust can be placed in users' initial responses in willingness to use different types of notifications, as they are unlikely to change through usage.***

Although these guidelines were drawn with mobile service notifications in mind, they are not necessarily restricted to this domain, as the design methodologies and comparison study could have been applied for any other type of interface functions or alerts. Perhaps limitation to their generalisability arises from the fact that mobile services represent more complex concepts than simple alert functions commonly found in the auditory interface literature. This shaped our representativeness procedure, where we decided to present only three candidate auditory icons for each service, so that respondents could focus on the meaning of the services every time (see guideline 8). Therefore, the suggested guidelines can assist the design of any auditory interface with somewhat complex functions that are represented by non-speech auditory icons. For example, in-car notifications about malfunctions or driving aides could be designed according to our guidelines in order to be immediately recognised and easily retained. Finally, these guidelines could be applied across the multiple mobile, domestic or public devices that often compete for our attention.

Chapter 7

Conclusions and Future Directions

7.1 Thesis Summary

The work presented in this thesis was originally motivated by the observed reluctance of mobile users to adopt the various novel mobile services brought by the technological advances in the area. We introduced the concept 'mobile service awareness' as potential solution to this problem, and as a supporting tool for the users who were more keen on using a variety of mobile services (especially since the launch of the iPhone). We chose this approach as mobile service discovery interaction methods have been found particularly problematic for finding and initiating mobile services (see [Garzonis & O'Neill 2006], [O'Neill et al., 2007] and Appendix II). These limitations are unlikely to be lifted due to the inherent difficulties of interacting with small devices, and the introduction and gradual uptake of contextually available services.

Mobile service awareness was proposed via a context-aware notification system, as a complete solution to the problem of missing out relevant services and managing their initiation. Focusing only on the delivery mechanism of this system, we argued that auditory notifications are the most appropriate and promising medium to address the awareness-interruption tradeoff. For these notifications to naturally blend in the daily soundscape but still attract attention when relevant to the situation, they need to be meaningful. This can be established by applying meaningful sound-service associations and/or through their explicit learning.

Our first investigation in comparing speech, auditory icon and earcon mobile service notifications gave us insights on the particularities of each sound type and opened the way for more systematic, in-depth comparisons. Before we evaluated the effectiveness of the non-speech auditory notifications, we had three preparation steps. First, we followed analytical reasoning and empirical card-sorting studies to arrive to a novel, meaningful mobile service classification, necessary for mitigating the problem of potentially infinite number of notifications. The next two preparation steps were to design the two sets of notifications through the appropriate systematic methodologies, by taking into account the classification produced. This would ensure that the stimuli to be compared were reasonably good representatives of the sound types. As a final step we designed a 4-stage longitudinal study to compare the effectiveness of the notifications we had produced. By investigating intuitiveness, learnability, memorability and user preference, we offer multiple novel and valuable contributions, including a set of specific conclusions with regard to non-speech mobile service notifications and guidelines for their respective design methodologies (summarised see Section 7.2).

7.2 Thesis Contributions

Our research question was scoped from ***"How can we improve mobile service usage?"*** to ***"How can we design for non-intrusive yet informative auditory mobile service notifications?"***, and a series of literature reviews, analytical and empirical studies were conducted to address it. Through these research methods, we gained knowledge and built the experience necessary to make informed and reliable in-depth comparisons on the effectiveness of auditory icons and earcons as mobile service notifications. As a result we provided two distinct contributions that address the research question on two levels. First, a set of specific conclusions on the effectiveness of auditory icons and earcons as mobile service notifications was drawn. Because of the systematic methodologies followed to produce the stimuli tested, these conclusions can be translated in suggestions that designers should have in mind when selecting notifications for auditory interfaces. Second, a set of guidelines for their design was derived by assessing the methodologies that produced the experimental stimuli (Chapter 5). These provide a unique and novel guidance for designing effective mobile service notifications, and significantly extend the literature suggestions on auditory interface design.

7.2.1 Conclusions on the Effectiveness of Auditory Icon and Earcon Notifications

The first major contribution of this thesis stems from the results of the comparison study investigating the effectiveness of auditory icons and earcons. Previous studies in the literature have focused only on learnability, short-term retention or collecting user preference data in short laboratory studies. In contrast, we have extended our comparison to also investigate intuitiveness of the notifications, learnability and user preference in longitudinal studies in the natural context and the laboratory, and long-term memorability.

Furthermore, most comparisons in the relevant literature are not only limited by the breadth (metrics) and depth (time span, context) of the comparisons, but also by the fact that the compared stimuli are not empirically validated prior to the comparison. On the contrary, we performed our in-depth comparison on auditory icons and earcons that have been produced through systematic, analytical and empirical methodologies. This underpins our findings reliability and generalisability, as we ensured that the sound instances were good representatives of the sound types and not biased by our abilities and personal preferences. Therefore, we are in a position to present our conclusions as suggestions that could apply to other types of auditory interfaces that utilise auditory icon or earcon notifications.

The results of the comparison study found auditory icons to be significantly more intuitive, easier to learn in casual everyday learning and laboratory training, and easier to retain in memory after one and four weeks. They were also preferred significantly more than earcons, which were found particularly unattractive to have in an active social context. The intuitiveness results are in agreement with the definition of the sound types and the laboratory learnability has been confirmed in similar studies.

Furthermore, this auditory icon superiority was also present in the casual learning and the memorability tests, making them a more appropriate choice for mobile service notifications. Although direction of user preference was in agreement with the other metrics, it contradicts some of the literature findings, where musical sounds were preferred to everyday sounds [Jones & Furner, 1989, as reported by Graham, 1999, and Lucas, 1994; Roberts & Sikora, 1997; Sikora et al., 1995].

The specific set of suggestions deriving from the analysis of the results is:

1. ***Underlying sound-service metaphors should be utilised (such as in auditory icons) for mobile service notifications to be immediately recognised.***

2. ***Long-standing and widespread metaphors need to be utilised for sound-service links to be intuitive.***

3. ***Strong sound-service metaphors should be utilised to avoid users creating alternative semantic relationships to other services.***

4. ***Extensive training is absolutely necessary for arbitrary sound-service associations (such as in earcons) to be learned. Yet, there might still be a significant percentage of users who will be unable to learn all of them perfectly.***

5. ***Auditory icons should be preferred to earcons for users to gradually learn to recognise mobile service notifications through natural usage (even if the sound-service associations are somewhat poor).***

6. ***Auditory icons should be preferred to earcons for sound-service associations to be retained in memory (especially for notifications that do not occur many times per week).***

7. ***Familiar sounds that 'make sense' should be used for auditory notifications to be liked and adopted by users.***

8. ***Auditory icon mobile service notifications are more likely to be adopted by users over earcons in a variety of social contexts.***

7.2.2 Design Guidelines for Non-speech Auditory Notifications

The second major contribution of the thesis is a set of design guidelines for auditory mobile service notifications, which was produced by identifying and comparing patterns in the participants' reactions between the stimuli-producing procedures (Chapter 5) and the effectiveness comparison study (Section 6.1).

The design process leading to the comparison study started with the classification of currently available services. This was done in a manner that new services could fit in the rationale and the existing clusters of the classification. Then, in two parallel streams, auditory icon and earcon notifications were designed by two distinct methodologies. A novel methodology for designing auditory icons was created, taking in to account and adapting sparse suggestions from everyday sound identification and warning alerts literature (Section 5.1). Earcons' design methodology followed more closely literature guidelines and procedures [Brewter et al., 1993] (Section 5.2). Both methodologies were designed with mobile service notifications in mind but are not considerably limited to this domain. One limitation to their generalisability arises from the fact that mobile services represent more complex concepts than simple alert functions commonly found in the auditory interface literature. This informed the representativeness measurements of our auditory icon design procedure, and is reflected in guideline 8 below. Therefore, our design guidelines can assist the design of any auditory interface that utilises non-speech notifications, especially if the interface functions or events are somewhat more complex than simple alerts. Most importantly, they could be applied across the plethora of auditory interfaces that often compete for our attention.

The design guidelines suggested are:

1. ***Avoid presenting clusters of services whose content is conceptually very similar.***

2. ***Consider the whole set of services and suggest auditory icons that cannot be easily associated with more than one service.***

3. ***In everyday sound identification studies, judge responses on their success to describe the metaphors that auditory icons will utilise to link the sound to the service.***

4. ***Auditory icons are culture-specific, and should be designed with this in mind.***

5. ***In auditory icon design, select everyday sounds that are identified by at least 80% of a random sample of the intended end-users.***

6. ***Avoid suggesting alternative auditory icons that utilise very similar metaphors.***

7. ***A combination of scores on everyday sound identification, sound-service representativeness votes and ratings (as obtained by the methodology described in Section 5.1), can be used as a good predictor of auditory icon intuitiveness.***

8. ***In auditory icon associability studies, consider presenting all services as candidates for all auditory icons, depending on the conceptual complexity of the services (and the associated resource constraints).***

9. ***Depending on resource constraints, consider retrieving representativeness ratings for all possible (or reasonable) associations.***

10. ***Use iterative design for earcon notifications to improve their distinguishability.***

11. ***A combination of tutorial, active learning and trial-by-trial feedback should be used in the iterative evaluations/training of earcons.***

12. ***When ascending and descending melodies are utilised for earcon distinguishability, the difference in pitch between the highest and lowest notes should be steep and more than two octaves.***

13. ***When possible, involve users in the design of earcon notifications and their assignment to services.***

14. ***Trust can be placed in users' initial responses in willingness to use different types of notifications, as they are unlikely to change through usage.***

7.2.3 Further Contributions

Secondary to the thesis' two major contributions, are two smaller yet significant contributions. First, in our extensive literature review on the different types of sounds utilised in auditory interface research, we have provided with a topology describing their relationship in a novel way (Section 3.3). Notification types are often classified according to the directness of the signal-referent relationship. However, there are also sparse references in the literature with regard to the sounds' ability to acquire meaning through exposure and social conventions. By combining these two dimensions in an orthogonal way, we provided with a 2-dimensional representation that adds and extends our understanding of how auditory notifications can be utilised. We also discussed one of the advantages of this new representation, which is its ability to demonstrate the fluidity of the sound type relationships amongst individuals and over time (Figure 3-6).

The other secondary contribution of this thesis is the novel mobile service classification described in Section 4.2, and used in the studies presented in Chapters 5 and 6. Our approach to provide three high-level categories describing the 'origin' of a service ('World-to-User', 'User-to-User' and 'User-to-Self') has been empirically validated in two card-sorting studies. With mobile widgets on the rise (especially popularised by the iPhone), users and designers alike may soon find that a meaningful classification is needed to organise and manage them. Therefore, the application of distinct families of notifications (as demonstrated in the experiments presented in Section 5.2) is only one of the beneficial applications of our classification.

7.3 Future Directions

The research presented in this thesis introduced the concept of mobile service awareness, and made a significant contribution in this direction. By doing so, it has also opened way for further research, which we present in two sections: the potential improvements and extensions in our work in auditory mobile service notifications, and the potential contributions in mobile service awareness beyond the auditory interface.

7.3.1 Further Mobile Service Awareness via Auditory Notifications

Research in the domain of auditory interface is of challenging nature, as sound design always requires a level of artistic input, which hinders the application of purely objective methods. This difficulty has been identified by many researchers in the literature, such as [Cohen, 1994], who, after his study on monitoring background activities through auditory icons, concludes:

> "One thing discovered in this process is just how difficult the art of sound design really is"

Therefore, the need to apply iterative, empirical design has been adopted by most researchers as necessary means to minimise the subjectivity impact of sound design. The two following quotations are indicative:

> "The lack of a complete theory of human auditory perception and the intrinsic variability between human listeners makes it impossible to predict the interpretation of any given acoustic signal on the basis of existing knowledge. This implies that the development of Auditory Display techniques is necessarily experimental, requiring validation through user evaluation" [Williams, 1992]

and

> "Whatever progress was made beyond these pitfalls was due to the design-and-test-methodology – participants were able to describe problems extremely accurately (even pointedly), and often they suggested the solutions" [Cohen, 1994]

Although we have adopted a systematic and empirical research approach throughout this thesis, our work remains open to the same criticisms applicable to all similar research: the design of the auditory stimuli we used may have influenced the direction and/or the magnitude of our results. Therefore, future research is needed to further validate and possibly improve the design methodologies and guidelines we suggest. Furthermore, commercial designers of notifications may benefit more from quicker and simpler design processes. Therefore attempting to streamline the methodologies presented here should also be investigated in the future.

Moreover, speech notifications were excluded from the final comparison study of Chapter 6. This was decided as they were expected to perform near perfection with regard to intuitiveness, learnability and memorability, and their presence in our study would significantly increase demand on limited resources, such as participants' availability and interruption tolerance (especially during the field study). However, it would be interesting to see how speech notifications would be received by participants, especially during a field study similar to the one we conducted. It is within our research interests to explore and compare user preferences between speech and non-speech notifications, and also to extend our understanding on the comparative willingness to use auditory icons, earcons and hybrid auditory interfaces.

Finally, throughout this thesis we have argued that meaningful notifications are more likely to be less intrusive. This has been based on arguments related to the nature of hearing and context of mobile use. However, it was not possible to measure the perceived intrusiveness of the notifications our participants received in order to compare it to the measures of meaningfulness. In future work, this relationship should be explored, by finding effective ways to control all the confounding variables present in naturalistic field studies (which is the ideal environment for testing).

7.3.2 Mobile Service Awareness beyond the Auditory Interface

The aim of the research presented here has been to provide awareness of mobile services where and when they are relevant to users' everyday activities. Due to resource limitations, our investigations were limited to the contributions the auditory interface can offer towards this goal. However, there are at least two more components that are necessary to provide with a complete mobile service awareness system.

First, multimodal or crossmodal notifications are needed for when the auditory modality is inappropriate or otherwise inaccessible. We have presented literature findings that indicate that most of the times mobile phones are tacked away in pockets and bags, or are left out of reach from their owners. This means that the visual and tactile notifications will often be difficult or impossible to perceive. Nevertheless, there will be contexts were one's mobile will be inaudible, either because of prevailing ambient noise or because the ringer is off (purposefully set or forgotten so). For at least these occasions, alternative modalities should be responsible for notifying about the relevant service. For this, future studies need to find the appropriate information encodings between stimuli and services, and balance them with the modalities' attentional demands. Furthermore, multimodal notifications could be utilised in a complementary manner, where different dimensions of the notification are encoded to different modalities [e.g. Hoggan & Brewster, 2007].

Finally, our work has been based on the assumption of a context-aware system that determines the relevancy of services in every situation. Our literature review on context-aware systems indicates that such a system is indeed feasible, although it is not expected to be error-free. However, the nature of the intuitive notifications designed and suggested here, are expected to partly alleviate such errors from the context engine. Future work to integrate a context-aware system with a notification engine would provide a complete mobile service awareness solution.

The contribution of this thesis is the first step towards a system with a set of cohesive and comprehensive notifications, which defensively support our daily activities. Hopefully, it will be improved and extended to be applied across the multiple mobile, domestic or public devices that often compete for our attention. Mobile phones, PDAs, health monitor devices, in-car notifications, PCs, TVs, tourist guides, ovens, elevators and ATMs do not have to relentlessly 'beep' at us.

REFERENCES

Adam, E. C., [1993], Fighter cockpits of the future, in Proceedings of the Digital Avionics Systems Conference, 1993. 12th DASC., AIAA/IEEE, pp.318-323.

Adams, M. J. and Pew, R. W., [1990], Situational awareness in the commercial aircraft cockpit: a cognitive perspective, in Proceedings of the Digital Avionics Systems Conference, 1990. Proceedings., IEEE/AIAA/NASA 9th, pp.519-524.

Albers, M. C., [1996], Auditory Cues for Browsing, Surfing and Navigating, in Proceedings of the ICAD '96.

Alexanderson, P., [2004], Peripheral awareness and smooth notification: the use of natural sounds in process control work. third Nordic conference on Human-computer interaction, pp.281-284.

Aoki, P. M., Grinter, R. E., Hurst, A., Szymanski, M. H., Thornton, J. D. and Woodruff, A., [2002], Sotto voce: exploring the interplay of conversation and mobile audio spaces. In (Eds.), SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, 431-438, ACM, Minneapolis, Minnesota, USA.

Baber, C. and Noyes, J. M., [1993], Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers, Taylor & Francis, Inc.

Baldauf, M., Dustdar, S. and Rosenberg, F., [2007], A survey on context-aware systems. International Journal of Ad Hoc and Ubiquitous Computing, pp.263-277.

Ballas, J. A. and Howard, J. A., Jr., [1987], Interpreting the language of environmental sound. Environment and Behavior, 19, pp.91-114.

Ballas, J. A., [1994], Delivery of Information through Sound. In Kramer, G. (Eds.), Display, Sonification, Audification and Auditory Interfaces, 79-94.

Barfield, W., Rosenberg, C. and Levasseur, G., [1991], The use of icons, earcons and commands in the design of an online hierarchical menu. IEEE transactions on professional communication, 34 (2), pp.101-108.

Barkhuus, L., Brown, B., Bell, M., Sherwood, S., Hall, M. and Chalmers, M., [2008], From awareness to repartee: sharing location within social groups. In (Eds.), Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, 497-506, ACM, Florence, Italy.

Barsalou, L. W., [1991], Deriving categories to achieve goals. In Bower, G. H. (Eds.), The psychology of learning and motivation, CA: Academic.

Bernstein, M., [2005], An Introduction to Music, Kessinger Publishing.

Bhatia, S. and McCrickard, S., [2006], Listening to your Inner Voices: Investigating Means for Voice Notifications, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M., [1989], Earcons and Icons: Their Structure and Common Design Principles. SIGCHI Bull, 21 (1), pp.123-124.

Blattner, M. M., Papp, A. L. I. and Glinert, E. P., [1994], (Sonic enhancement of two-dimensional graphic displays. In Kramer, G. (Eds.), Auditory Display, 447 - 470, Addison-Wesley, New York.

Block, J. F. E., Nuutinen, L. and Ballast, B., [1999], Optimization of Alarms: A Study on Alarm Limits, Alarm Sounds, and False Alarms, Intended to Reduce Annoyance. Journal of Clinical Monitoring and Computing, 15 pp.75-83.

Block, F. E., Rouse, J. D., Hakala, M. and Thompson, C. L., [2000], A proposed new set of alarm sounds which satisfy standards and rationale to encode source information. Journal of Clinical Monitoring and Computing, 16, pp.541-546.

Bly, S. A., Harrison, S. R. and Irwin, S., [1993], Media spaces: bringing people together in a video, audio, and computing environment. Communications of the ACM, 36 (1), pp.28-46.

Bonebright, T. L. and Nees, M. A., [2007], Memory for auditory icons and earcons with localization cues., in Proceedings of the ICAD 07, pp.419-422.

Brewster, S. A., Wright, P. C. and Edwards, A. D. N., [1993], An evaluation of earcons for use in auditory human-computer interfaces In (Eds.), Proceedings of the SIGCHI conference on Human factors in computing systems 222-227 ACM Press, Amsterdam, The Netherlands.

Brewster, S. A., [1998], Using non-speech sounds to provide navigation cues. ACM Transactions on Computer-Human Interaction (TOCHI), 5 (2), pp.224-259.

Brewster, S., [2002], Overcoming the lack of screen space on mobile computers. Personal and Ubiquitous Computing, 6 (3), pp.188-205.

Brewster, S. A., [2002], Chapter 12: Non-speech auditory output. In Jacko, J. a. S., A. (Eds.), The Human Computer Interaction Handbook, 220-239, Lawrence Erlbaum Associates, USA,

Brewster, S. and Brown, L. M., [2004], Tactons: Structured Tactile Messages for Non-Visual Information Display, in Proceedings of the fifth conference on Australasian user interface (AUIC), Australian Computer Society, pp.15 - 23.

Brown, B. and Randell, R., [2004], Building a Context Sensitive Telephone: Some Hopes and Pitfalls for Context Sensitive Computing http://dx.doi.org/10.1007/s10606-004-2806-4 Comput. Supported Coop. Work 13 (3-4), pp.329-345.

Brown, L. M., Brewster, S. A. and Purchase, H. C., [2005], A First Investigation into the Effectiveness of Tactons, in Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, IEEE Computer Society, pp.167-176.

Bussemakers, M. P. and De Haan, A., [2000], When it sounds like a duck and it looks like a dog... Auditory icons vs. earcons in multimedia environments., in Proceedings of the International Conference on Auditory Display, pp.184-189.

Chen, G., and Kotz, D., [2000], A Survey of Context-Aware Mobile Computing Research, Technical Report: TR2000-381, Department of Computer Science, Dartmouth College.

Cherry, C., [1953], Some Experiments on the Recognition of Speech, with One and with Two Ears. The Journal of the Acoustical Society of America, 25 (5), pp.975-979.

Chewar, C. M., Bachetti, E., McCrickard, D. S. and Booker, J. E., [2004], Automating a Design Reuse Facility with Critical Parameters: Lessons Learned in Developing the LINK-UP System, in Proceedings of the Conference on Computer-Aided Design of User Interfaces (CADUI), pp.233-244.

Chipchase, J., Persson, P., Piippo, P., Aarras, M. and Yamamoto, T., [2005], Mobile essentials: field study and concepting. conference on Designing for User eXperience, pp.57.

Clarkson, E., Clawson, J., Lyons, K. and Starner, T., [2005], An empirical study of typing rates on mini-QWERTY keyboards, in Proceedings of the CHI '05 extended abstracts on Human factors in computing systems, ACM Press, pp.1288-1291.

Cohen, J., [1994], Monitoring Background Activities. In Kramer, G. (Eds.), Auditory Display: Sonification, Audification and Auditory interfaces, 499-522.

Davies, N., Cheverst, K., Mitchell, K. and Efrat, A., [2001], Using and Determining Location in a Context-Sensitive Tour Guide. Computer, 34 (8), pp.35-41.

Deutsch, J. A. and Deutsch, D., [1963], Attention: Some theoretical considerations. Psychological Review, 70 (1), pp.51-60.

Dey, A. K. and Abowd, G. D., [2000], CybreMinder: A Context-Aware System for Supporting Reminders In (Eds.), Proceedings of the 2nd international symposium on Handheld and Ubiquitous Computing 172-186 Springer-Verlag, Bristol, UK.

Dingler, T., Lindsay, J. and Walker, B., [2008], Learnabiltiy of Sound Cues for Environmental Features: Auditory Icons, Earcons, Spearcons, and Speech. In (Eds.), Proceedings of the 14th International Conference on Auditory Display, Paris, France.

Dixon, N., F., [1981], Preconscious Processing, Wiley, NY.

Dourish, P. and Bly, S., [1992], Portholes: supporting awareness in a distributed work group. In (Eds.), SIGCHI conference on Human factors in computing systems, 541-547, ACM, Monterey, California, United States.

Edworthy, J. and Stanton, N., [1995], A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology. Ergonomics, 38 (11), pp.2262 - 2280.

Edworthy, J. and Hards, R., [1999], Learning Auditory Warnings: The Effects of Sound Type, Verbal Labelling and Imagery on the Identification of Alarm Sounds. International Journal of Industrial Ergonomics, 24 (6), pp.603-618.

Endsley, M. R., [1995], Toward a theory of situation awareness in dynamic systems. Human Factors, 37 (1), pp.32-64.

Entin, E. B., Entin, E. E., MacMillan, J. and Serfaty, D., [1995], Situation awareness and human performance in target recognition, in Proceedings of the Systems, Man and Cybernetics, 1995. Intelligent Systems for the 21st Century., IEEE International Conference on, pp.3833-3837 vol.4.

Filho, R. S. S., [2000], Awareness and Privacy in Mobile Wearable Computers. IPADS: Interpersonal Awareness Devices, http://awareness.ics.uci.edu/~rsilvafi/papers/VirtualColocationFinalPaper.pdf.

Fitch, W. T. and Kramer, G., [1994], Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex Multivariate System. In Kramer, G. (Eds.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, Addison-Wesley, New York.

Fogarty, J., Hudson, S. E., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C. and Yang, J., [2005], Predicting human interruptibility with sensors. ACM Transactions on Computer-Human Interaction (TOCHI), 12 (1), pp.119-146.

Gardner, D. and Simpson, C., [1992], Sound image development and testing for RAH-66 (Comanche) warning caution and advisory system - Final Report, under contract to Sikorsky Aircraft Division, Psycho-Linguistic Research Associates, Woodside, CA.

Garzonis, S. and O'Neill, E. (2006). Factors contributing to low usage of mobile data services: user requirements, service discovery and usability. People and Computers XX, Proceedings of HCI 2006: Engage, London, UK, September 2006, 221-236.

Gaver, W. W., [1986], Auditory Icons: Using Sound in Computer Interfaces. Human-Computer Interaction, 2 (2), pp.167-177.

Gaver, W. W., [1989], The SonicFinder: An Interface That Uses Auditory Icons. Human-Computer Interaction, 4 (1), pp.67-94.

Gaver, W. W., [1991], Sound support for collaboration In (Eds.), Proceedings of the second conference on European Conference on Computer-Supported Cooperative Work 293-308 Kluwer Academic Publishers, Amsterdam, The Netherlands.

Gaver, W. W., [1992], The affordances of media spaces for collaboration. In (Eds.), ACM conference on Computer-supported cooperative work, 17-24, ACM, Toronto, Ontario, Canada.

Gaver, W. W., [1997], Auditory Interfaces. In Helander, M. G., Landauer, T. K. and Prabhu, P. V. (Eds.), Handbook of Human-Computer Interaction, Elsevier Science.

Gaver, B., [2002], Provocative Awareness. Computer Supported Cooperative Work, 11 (3), pp.475-493.

Gellersen, H. W., Schmidt, A. and Beigl, M., [2002], Multi-Sensor Context-Awareness in Mobile Devices and Smart Artifacts, in Proceedings of the Mobile Networks and Applications (MONET).

Gershman, A. V., McCarthy, J. F. and Fano, A. E., [1999], Situated Computing: Bridging the Gap between Intention and Action, in Proceedings of the 3rd IEEE International Symposium on Wearable Computers, IEEE Computer Society, pp.3.

Geven, A., Sefelin, R. and Tscheligi, M., [2006], Depth and Breadth away from the desktop - Optimal Information Hierarchies for Mobile Use, in Proceedings of the Eighth International Symposium on Human Computer Interaction with Mobile Devices and Services, MobileHCI 2006, pp.157-164.

Gonzalez, C. and Wimisberg, J., [2007], Situation awareness in dynamic decision-making: Effects of practice and working memory. Journal of Cognitive Engineering and Decision Making, 1 (1), pp.56-74.

Graham, R., [1999], Use of Auditory Icons as Emergency Warnings: Evaluation within a Vehicle Collision Avoidance Application. Ergonomics, 42 (9), pp.1233-1248.

Gutwin, C. and Greenberg, S., [1995], Support for Group Awareness in Real-time Desktop Conferences, in Proceedings of the Second New Zealand Computer Science Research Students' Conference.

Guzman, E. S. D., Yau, M., Gagliano, A., Park, A. and Dey, A. K., [2004], Exploring the design and use of peripheral displays of awareness information. CHI '04 extended abstracts on Human factors in computing systems, pp.1247-1250.

Hansmann, U., Nicklous, M. S. and Stober, T., [2001], Pervasive Computing Handbook, Springer Verlag.

Harrison, B. L., Fishkin, K. P., Gujar, A., Mochon, C. and Want, R., [1998], Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. In (Eds.), SIGCHI conference on Human factors in computing systems, 17-24, ACM Press/Addison-Wesley Publishing Co., Los Angeles, California, United States.

Healey, J. and Picard, R. W., [1998], StartleCam: A Cybernetic Wearable Camera In (Eds.), Proceedings of the 2nd IEEE International Symposium on Wearable Computers 42 IEEE Computer Society.

Hickling, E. M., [1994], Ergonomics and engineering aspects of designing an alarm system for a modern nuclear power plant. In Stanton, N. (Eds.), Human Factors in Alarm Design, 165-178, Taylor & Francis, Bristol, PA.

Hoggan, E. and Brewster, S. A., [2007], Designing Audio and Tactile Crossmodal Icons for Mobile Devices, in Proceedings of the ACM International Conference on Multimodal Interfaces, ACM Press, pp.pp162-169.

Holmquist, L. E., Wigström, J. and Falk, J., [1998], The Hummingbird: Mobile Support for Group Awareness, in Proceedings of the CSCW 1998.

Horvitz, E., Kadie, C., Paek, T. and Hovel, D., [2003], Models of attention in computing and communication: from principles to applications. Commun. ACM, 46 (3), pp.52-59.

Hourizi, R. and Johnson, P., [2004], Designing to support awareness: a predictive, composite model. In (Eds.), Proceedings of the SIGCHI conference on Human factors in computing systems, 159-166, ACM, Vienna, Austria.

Hsu, J.-M., Wu, W.-J. and Chang, I.-R., [2007], Ubiquitous Multimedia Information Delivering Service for Smart Home, in Proceedings of the Multimedia and Ubiquitous Engineering, 2007. MUE '07. International Conference on, pp.341-346.

Hudson, S. E. and Smith, I., [1996], Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In (Eds.), Proceedings of the 1996 ACM conference on Computer supported cooperative work, 248-257, ACM, Boston, Massachusetts, United States.

Hudson, S. E. and Smith, I., [1996], Electronic mail previews using non-speech audio. In (Eds.), Conference companion on Human factors in computing systems: common ground, 237-238, ACM, Vancouver, British Columbia, Canada.

Hudson, J. M., Christensen, J., Kellogg, W. A. and Erickson, T., [2002], "I'd be overwhelmed, but it's just one more thing to do": availability and interruption in research management, in Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, ACM, pp.97-104.

Ichikawa, F., Chipchase, J. and Grignani, R., [2005], Where's The Phone? A Study of Mobile Phone Location in Public Spaces, in Proceedings of the Mobile Technology, Applications and Systems, 2005 2nd International Conference on, pp.1-8.

Isaacs, E., Walendowski, A. and Ranganathan, D., [2002], Hubbub: A Sound-Enhanced Mobile Instant Messenger that Supports Awareness and Opportunistic Interactions, in Proceedings of the Computer-Human Interaction (CHI), pp.179-186.

James, W., [1890], The Principles of Psychology, New York.

Johnston, W. A. and Heinz, S. P., [1978], Flexibility and capacity demands of attention. Journal of Experimental Psychology: General, 107 (4), pp.420-435.

Jones, S. D. and Furner, S. M., [1989], The construction of audio icons and information cues for human computer dialogues, in Proceedings of the Ergonomic Society's 1989 Annual Conference, Taylor & Francis, pp.436-441.

Kaenampornpan, M., [2009], PhD Thesis, A Context Model, Design Tool and Architecture for Context-Aware Systems Design, University of Bath, UK.

Keller, P. and Stevens, C., [2004], Meaning from Environmental Sounds: Types of Signal-Referent Relations and their Effect on Recognizing Auditory Icons. Journal of Experimental Psychology: Applied, 10 (1), pp.3-12.

Kern, N. and Schiele, B., [2006], Towards Personalized Mobile Interruptibility Estimation.

Kramer, G., [1994], Some Organizing Principles for Representing Data With Sound. In Kramer, G. (Eds.), SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII, 185-222, Addison-Wesley, New York.

Labov, W., [1973], The boundaries of words and their meanings. In Shury, C.-J. N. B. R. W. (Eds.), New ways of analyzing variations in English, DC: Georgetown U. Press.

Leplatre, G. and Brewster, S. A., [2000], Designing non-speech sounds to support navigation in mobile phone menus, in Proceedings of the 6th International Conference on Auditory Display (ICAD), pp.190-199.

Leung, Y. K., Smith, S., Parker, S. and Martin, R., [1997], Learning and Retention of Auditory Warnings, in Proceedings of the 3rd International Conference on Auditory Display (ICAD).

Lucas, P., [1994], An evaluation of the communicative ability of auditory icons and earcons, in Proceedings of the Second International Conference on Auditory Display (ICAD), pp.121-128.

Lukkari, J., Korhonen, J. and Ojala, T., [2004], SmartRestaurant: mobile payments in context-aware environment, in Proceedings of the 6th international conference on Electronic commerce, ACM, pp.575-582.

Luyten, K. and Coninx, K., [2004], ImogI: Take Control over a Context-Aware Electronic Mobile Guide for Museums, , in Proceedings of the Workshop on HCI in Mobile Guides, in conjunction with 6th International Conference on Human Computer Interaction with Mobile Devices and Services.

Lyytinen, K. and Yoo, Y., [2002], Issues and challenges in ubiquitous computing. Communications of ACM, 45 (12), pp.62-65.

Marcell, M. M., Borella, D., Greene, M., Kerr, E. and Rogers, S., [2000], Confrontation naming of environmental sounds. Journal of Clinical and Experimental Neuropsychology, 22 (6), pp.830-864.

Marcell, M., Malatanos, M., Leahy, C. and Comeaux, C., [2007], Identifying, rating, and remembering environmental sound events. Behavior Research Methods, 39 (3), pp.561-569.

McCarthy, J. F. and Anagnost, T. D., [2000], EVENTMANAGER: Support for the Peripheral Awareness of Events In (Eds.), Proceedings of the 2nd international symposium on Handheld and Ubiquitous Computing 227-235 Springer-Verlag, Bristol, UK.

McCrickard, D. S. and Chewar, C. M., [2003], Attuning notification design to user goals and attention costs Commun. ACM 46 (3), pp.67-72.

McGee-Lennon, M., M., W. and McBryan, T., [2007], Audio Reminders in the Home Environment, in Proceedings of the International Conference on Auditory Displays.

McKeown, D., [2005], Candidates for within-vehicle auditory displays, in Proceedings of the ICAD 05-Eleventh Meeting of the International Conference on Auditory Display.

Moore, B., [2003], An Introduction to the Psychology of Hearing, Academic Press.

Moray, N., [1959], Attention in dichotic listening: Affective cues and the influence of instructions. The Quarterly Journal of Experimental Psychology, 11 (1), pp.56 - 60.

Munoz, M. A., Rodriguez, M., Favela, J., Martinez-Garcia, A. I. and Gonzalez, V. M., [2003], Context-Aware Mobile Communication in Hospitals. Computer, 36 (9), pp.38-46.

Mynatt, E. D., [1994], Designing with auditory icons: how well do we identify auditory cues? , in Proceedings of the Conference companion on Human factors in computing systems, ACM, pp.269-270.

Mynatt, E. D., Back, M., Want, R., Baer, M. and Ellis, J. B., [1998], Designing audio aura, in Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co., pp.566-573.

Nass, C. and Gong, L., [2000], Speech interfaces from an evolutionary perspective. Commun. ACM, 43 (9), pp.36-43.

Nullmeyer, R. T., Stella, D., Montijo, G. A. and Harden, S. W., [2005], Human factors in Air Force flight mishaps: Implications for change, in Proceedings of the 27th Annual Interservice/Industry Training, Simulation, and Education Conference, National Training Systems Association.

O'Neill, E., Thompson, P., Garzonis, S. and Warr, A., [2007], Reach out and touch: using NFC and 2D barcodes for service discovery and interaction with mobile devices, in Proceedings of the Pervasive 2007, pp.19-36.

Patel, S., Kientz, J., Hayes, G., Bhat, S. and Abowd, G., [2006], Farther Than You May Think: An Empirical Investigation of the Proximity of Users to Their Mobile Phones, in Proceedings of the UbiComp 2006: Ubiquitous Computing, pp.123-140.

Patterson, R. D., [1982], Guidelines for Auditory Warning Systems on Civil Aircraft.

Patterson, R. D. and Mayfield, T. F., [1990], Auditory Warning Sounds in the Work Environment [and Discussion]. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 327 (1241), pp.485-492.

Pedersen, E. R. and Sokoler, T., [1997], AROMA: abstract representation of presence supporting mutual awareness. SIGCHI conference on Human factors in computing systems, pp.51-58.

Pirhonen, A., Murphy, E., McAllister, G. and Yu, W., [2006], Non-speech sounds as elements of a use scenario: A semiotic perspective, in Proceedings of the 12th International Conference on Auditory Display (ICAD2006), Department of Computer Science, Queen Mary, University of London, UK, pp.134-140.

Pramana, E. and Leung, Y. K., [1999], Exploring the Learnability of Structured Earcons, in Proceedings of the HCI International (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I - Volume I, L. Erlbaum Associates Inc., pp.555-559.

Rasmussen, J., [1986], Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering, Elsevier Science Inc.

Rigas, D. and Memery, D., [2002], Utilising Audio-Visual Stimuli in Interactive Information Systems: A Two Domain Investigation on Auditory Metaphors, in Proceedings of the International Conference on Information Technology: Coding and Computing, IEEE Computer Society, pp.190.

Rigas, D. and Alty, J., [2005], The rising pitch metaphor: an empirical study. Int. J. Hum.-Comput. Stud., 62 (1), pp.1-20.

Roberts, L. A. and Sikora, C. A., [1997], Optimising Feedback Signals for Multimedia Devices: Earcons vs. Auditory Icons vs. Speech, in Proceedings of the International Ergonomics Association (IEA'97).

Robertson, T., [2002], The Public Availability of Actions andArtefacts. Comput. Supported Coop. Work, 11 (3), pp.299-316.

Sanderson, P., Wee, A., Seah, E. and Lacherez, P., [2006], Auditory alarms, medical standards, and urgency, in Proceedings of the 12th International Conference on Auditory Display (ICAD2006), pp.24-27.

Sanderson, P. M., Wee, A. and Lacherez, P., [2006], Learnability and discriminability of melodic medical equipment alarms. Anaesthesia, 61 (2), pp.142-147.

Sawhney, N. and Schmandt, C., [2000], Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. ACM Transactions on Computer-Human Interaction (TOCHI), 7 (3), pp.353-383.

Schiaffino, S. and Amandi, A., [2004], User - interface agent interaction: personalization issues. International Journal of Human-Computer Studies (IJHCS), 60 (1), pp.129-148.

Schilit, B. and Theimer, M., [1994], Disseminating Active Map Information to Mobile Hosts. IEEE Network, 8 (5), pp.22-32.

Schmandt, C., [1999], Everywhere Messaging. In (Eds.), Handheld and Ubiquitous Computing, 22-27.

Schmidt, A., Aidoo, K. A., Takaluoma, A., Tuomela, U., Laerhoven, K. V. and Velde, W. V. d., [1999], Advanced Interaction in Context In (Eds.), Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing 89-101 Springer-Verlag, Karlsruhe, Germany.

Schmidt, A., [2000], Implicit Human Computer Interaction Through Context. Personal Technologies, 4 (2-3), pp.191-199.

Schmidt, K., [2002], The Problem with 'Awareness': Introductory Remarks on 'Awareness in CSCW'. Computer Supported Cooperative Work, 11 (3), pp.285-298.

Serenko, A., [2006], The use of interface agents for email notification in critical incidents. International Journal of Human-Computer Studies (IJHCS), 64 (11), pp.1084-1098.

Siewiorek, D., Smailagic, A., Furukawa, J., Krause, A., Moraveji, N., Reiger, K., Shaffer, J. and Wong, F. L., [2003], SenSay: A Context-Aware Mobile Phone In

(Eds.), Proceedings of the 7th IEEE International Symposium on Wearable Computers 248 IEEE Computer Society.

Sikora, C. A., Roberts, L. and Murray, L., [1995], Musical vs. real world feedback signals, in Proceedings of the CHI'95, pp.220-221.

Simpson, C. A., [2007], Doing Science on Auditory Display Design in the Cockpit: Merging Laboratory Rigor and the Aircraft Cockpit Environment, in Proceedings of the 3th International Conference on Auditory Display (ICAD2007), Schulich School of Music, McGill University, pp.139-142.

Sodnik, J., Dicke, C., Tomazic, S. and Billinghurst, M., [2008], A user study of auditory versus visual interfaces for use while driving. Int. J. Hum.-Comput. Stud., 66 (5), pp.318-332.

Sorkin, R. D., [1987], Design of auditory and tactile displays. In Salvendy, G. (Eds.), Handbook of human factors, 549-576, Wiley & Sons, New York.

Stevens, R., Brewster, S., Wright, P. and Edwards, A., [1994], Design and evaluation of an auditory glance at algebra for blind readers, in Proceedings of the Second International Conference on Auditory Display (ICAD), pp.21 - 30.

Stevens, C., Brennan, D. and Parker, S., [2004], Simultaneous Manipulation of Parameters of Auditory Icons to Convey Direction, Size, and Distance: Effects on Recognition and Interpretation, in Proceedings of the 10th International Conference on Auditory Display.

Stevens, C., Perry, N., Wiggins, M. and Howell, C., [2006], Design and evaluation of auditory icons as informative warning signals, Australian Transportation Safety Bureau.

Treisman, A. M., [1964], Verbal Cues, Language, and Meaning in Selective Attention. The American Journal of Psychology, 77 (2), pp.206-219.

Treisman, A., [1985], Preattentive processing in vision. Comput. Vision Graph. Image Process., 31 (2), pp.156-177.

Tzelgov, J., Srebro, R., Henik, A. and Kushelevsky, A., [1987], Radiation search and detection by ear and by eye. Human Factors, 29 (1), pp.87-95.

Ulfvengren, P., [2003], Design of Natural Warning Sounds in Human-Machine Systems, PhD Thesis, Stockholm, Sweden.

Ulfvengren, P., [2003], Associability: A comparison of sounds in a cognitive approach to auditory alert design. Human Factors and Aviation Safety an International Journal, 3 (4), pp.313-331.

van Dantzich, M., Robbins, D., Horvitz, E. and Czerwinski, M., [2002], Scope: providing awareness of multiple notifications at a glance.

Vanderveer, N. J., [1979], Ecological acoustics: Human perception of environmental sounds . PhD Thesis.

Velmans, M., [1999], When Perception Becomes Conscious. British Journal of Psychology, 90 (4), pp.543-566.

Vertegaal, R., Velichkovsky, B. and Veer, G. v. d., [1997], Catching the eye: management of joint attention in cooperative work. SIGCHI Bulletin, 29 (4), pp.87-92.

Vertegaal, R., [2003], Introduction. Communications ACM 46 (3), pp.30-33.

Wahid, S., Bhatia, S., Lee, J. C., Perez-Quinones, M. A. and McCrickard, D. S., [2006], CaBN: Disparate Information Management through Context-Aware Notifications, in Proceedings of the 16th International Conference on Computer Theory and Applications (ICCTA 2006), pp.46-49.

Walker, B. N. and Kramer, G., [2005], Mappings and Metaphors in Auditory Displays: An Experimental Assessment. ACM Transactions on Applied Perception, 2 (4), pp.407-412.

Want, R., Hopper, A., Falc\&\#227;o, V. and Gibbons, J., [1992], The active badge location system. ACM Transactions on Information Systems (TOIS), 10 (1), pp.91-102.

Weiser, M., [1991], The computer for the 21st century. Scientific American, 265(3):94--104, 265 (3), pp.93 - 104.

Weiser, M., [1993], Some computer science issues in ubiquitous computing Commun. ACM 36 (7), pp.75-84.

Weiser, M. and Brown, J., [1995], Designing Calm Technology.

Williams, S. M., [1992], Perceptual principles in sound grouping, Addison Wesley Longman.

Williams, S. and Beatty, P. C., [2005], Measuring the performance of audible alarms for anaesthesia. Physiological Measurement, 26 (4), pp.571-581.

Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B. and Yarin, P., [1998], Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information. In (Eds.), Cooperative Buildings: Integrating Information, Organization, and Architecture SE-Lecture Notes on Computer Science, 22-32.

Wittgenstein, L., [1953], Philosophical investigations, Blackwell, Oxford.

APPENDICES

APPENDIX I: USERS' NEEDS

I.1 – A Short History of Usage of Personalised Ringtones

Ringtone downloads were amongst the most popular services a few years ago. There are reports that present the UK download ringtone market to have grown as much as 500% (to £177million) between 2000 and 2005[1]. This showed a clear demand for personalisation in auditory notifications, even if it consisted of short reproductions of popular songs or movie/TV soundtracks. However, this growth was reversed in the following years. "Caslon Analytics" reported claims by "M:Metrics" that "the percentage of mobile phone subscribers in the UK, Germany, France, Spain and Italy recurrently buying a ring tone" fallen consistently in 2006 (to a low of 3.4% in the UK)[2]; and claims by "JupiterResearch" that ringtones accounted for 29% of the overall European mobile content market in 2007, down from 33% in 2006[2]. This drop in ringtone downloads came us mobile phones' multimedia capabilities grew to accommodate any sound file as ringtone. In fact, since 2005 there have been a number of tools to help users upload their own ringtones to their mobile devices (e.g. Magix[3] and Phonezoo[4]). This way, people today have the option to truly personalise the incoming call notifications to any everyday or synthetically produced sound. The question remaining is to see whether they will choose unique, creative sounds. An indication for public desire to increase ringtone diversity is evident by the expansion of "The Ringtone Society"[5], which became international in 2008. According to their mission statement they aim to "liberate the world of musically banal ringtones" by asking composers and musicians to create "original ringtones" and everyone to "acquire and use as many as [they] like", and thus protect society from "the digitally-dull ringtone". Interestingly, how others perceive our ringtones seems to matter a lot. According to a 2006 survey by Phonecontent.com[6], 80% of British people were "worried what others would think of their choice of ringtones", while 90% of them felt they "have been ridiculed by family and friends for selecting weird ringtones".

---

[1] "Ringtone boom draws to a close" - thesun.co.uk, 05 Oct 2006: http://www.thesun.co.uk/sol/homepage/news/66024/Ringtone-boom-draws-to-a-close.html (last accessed: 30/06/2009)

[2] "Caslon Analytics ringtones" – Dec 2007: http://www.caslon.com.au/ringtonesnote.htm (last accessed: 30/06/2009)

[3] http://www.magix.com/us/ringtone-maker/ (last accessed: 30/06/2009)

[4] http://www.phonezoo.com (last accessed: 30/06/2009)

[5] http://www.ringtonesociety.com/content/society (last accessed: 30/06/2009)

[6] "Britons Too Worried About Choosing Ringtones" - phonecontent.com, 02 May 2006: http://www.phonecontent.com/bm/news/1468.shtml (last accessed: 30/06/2009)

I.2 – Factors Affecting User Needs

One of the most easily predictable reasons to why mobile services are not heavily used is because they simply do not address users' needs. However, the word "needs" can be particularly difficult to define or differentiate from "desires" or "valued preferences". Although it is outside the scope of this thesis to discuss the philosophical differences between human needs and desires, the concept of value is integral to user requirements, services and their uptake.

Furthermore, this willingness to use and adopt services can be particularly difficult to measure accurately. There are several factors influencing the answer to the question users implicitly or explicitly ask themselves: "Is it worth using?" A first factor is striking a successful balance between the financial costs and the perceived usefulness of a service. We gained some insights with regard to this balance from a cooperative evaluation study of a mobile network platform [Garzonis & O'Neill, 2006]. Although in that study we did not aim to take into account the financial cost of accessing the services, users intuitively and explicitly weighed their overall satisfaction about their experience against the money they would have to pay to access the data services. Despite the fact that they were not personally being charged during the study, they were still reluctant to accept charges during the interaction (e.g. 25p for a map) and expressed their concern with comments like "It's wasting my time and my money… ", "Will it charge me again?", and even "… this would be costing me money, wouldn't it? So I'd be getting a bit pissed off at this stage".

A second factor affecting willingness to use services is how the ease of use is affecting services' perceived value. If frustration overweighs the benefits of using these services, people will eventually turn them down. This was also supported by the results of the same study, where experienced mobile users faced major difficulties in interacting with the services. They frequently made intense remarks of dissatisfaction during the evaluation. One of the strongest findings was that most participants explicitly stated during the evaluation that they would not use it in real life. Among other interesting comments were "I'd rather ask someone on the street" or "I'd wait until I got to my PC". There were even some who made very negative comments, such as "I hate it", "it's absolutely useless" or "can I swear on camera?" These comments indicate that even if a service were initially considered as desired, user interest could rapidly fall due to the dissatisfaction caused by difficulty of use.

Furthermore, the desire for using a service can greatly fluctuate over time, as it is affected by factors such as advertising or popularity of use. The short history of personalised ringtone usage demonstrates the way service desire can be affected by unpredictable changes in the market (see Appendix I.1).

Appendix II: Mobile Service Discovery

There are two obvious ways for mobile users to determine which services are available in any given context: explore and look for cues in the immediate physical environment, or explore the virtual world through their mobile device.

By scanning our immediate environment we make many conscious and unconscious assumptions about the services available in that environment, with varying levels of confidence. The service discovery procedure is completed by the initiation of the service, which normally will take place through the mobile device. This initiation sometimes can be facilitated by the use of virtually augmented objects, which usually provide both the physical cues of the availability of a service and a shortcut to initiate the service. The most widespread such objects in research projects and commercial applications are RFID (Radio Frequency Identification) tags, NFC (Near Field Communication) tags and 2D barcodes. When embedded in physical objects such as posters and signs, they become the service initiation shortcut while the object is announcing its availability. When people choose to interact with them (e.g. touching the NFC-enabled phone on the tags, or pointing the phone-camera on the 2D barcodes), extra information is given usually in the form of a web link or an SMS. The service may be completed immediately (e.g. requesting an SMS with information about the next bus arrival), or require further interactions (e.g. opening the mobile browser to the page of a concert website, where tickets can be purchased).

In the absence of any environmental cues, the process of service discovery will be solely executed through the mobile device. The simplest such example is placing a phone call: it involves looking at the mobile screen to find out if there is network coverage (and its strength), and initiating the call through the mobile interface. Apart from navigating the phone menu, one can also discover and initiate services by browsing the operators' web platform, the mobile Internet or calling directory enquiries. Table II-1 summarises the common cues and methods available for the in-environment and on-device mobile service discovery processes.

MOBILE SERVICE DISCOVERY	CUES and METHODS
ON-DEVICE	Navigate phone menus/icons
	Browse operator platform
	Browse mobile internet
	Call directory enquiries
IN-ENVIRONMENT	Buildings and establishments
	Signs, posters and displays
	Physical objects and artefacts
	Virtually enhanced objects (e.g. NFC, Barcodes)

Table II-1: Common cues and methods available for the two approaches of discovering mobile services.

II.1 – Issues with On-Device Mobile Service Discovery

From Table II.1, 'Calling directory enquiries' mainly involves human interaction and is therefore out of the scope of human computer interaction. All other options involve certain interaction with the mobile device, from looking at the screen and pressing the green button twice (to see if there is signal and call the last called number) to extensive and deep navigation in the device's menus or network's links. These interactions are bound to be affected by the particular characteristics and limitations of the mobile device (e.g. small screen, limited input methods).

On-device mobile service discovery proved to be exceptionally difficult in previous work [Garzonis & O'Neill, 2006]. The results of a cooperative evaluation study indicated that the struggle of choosing the appropriate service exceeded the overall struggle of completing the task. Participants visited 500% web pages more than needed to find the optimum service, when for the overall task the respective number was 213%. Also, they spent 1403% more time than was needed for service discovery and 499% more for task completion.

There were also strong indications that the structure of most of the services did not match the users' mental models. The optimum path was followed in only 3.6% of the tasks and only by two participants. In most cases, task completion was hindered by deeply nested options with ambiguous labelling. For example, "Nearest Bookmakers" was under "More Sports", while "Essential Services" was one level down from "Tattoo & Piercing". In another example, labels such as "Pinpoint me" and "Find me now" strongly suggested that the service would display the user's location. However, upon selecting them, users were asked manually to input their location in order to see the relevant map. Even then, they had to go through 6 more steps (including purchase confirmation and a terms and conditions page) to get the map.

Although one should be careful in generalising the results of any one study, the difficulty reported in [Garzonis & O'Neill, 2006] was observed even when user had to choose from a fixed set of 12 services. It is clear that on-device mobile service discovery is bound to be even more complicated if users were introduced to contextually available services.

II.2 – Issues with In-Environment Mobile Service Discovery

The first obvious obstacle for in-environment mobile service discovery is the confirmation of availability and the initiation process. In cases where assumptions about the availability of services are made, there is always a chance that these assumptions are wrong. For example, despite the presence of a wireless access point, the wireless network might be unavailable to me as it can have restricted access or even be disabled. In the case where a service is explicitly advertised (e.g. on a poster or public display), the initiation step can still pose difficulties. There might be instructions to follow (such as a link address or a telephone number) but this would require an extra interaction step, which takes effort and time (e.g. to find and open the mobile browser, navigate to the URL address etc.), and introduces further opportunity for an error (e.g. mistyping the link).

This extra step can be considerably shortened by the use of virtually enhanced objects. For example, by touching a mobile device on the NFC tag on a poster describing the service, one can avoid manually entering the link to the web service. Nevertheless, these posters will need to be visible and accessible enough for users to become aware of the services they are advertising. This in itself gives rise to a few issues. First, there is still a requirement for an extra interaction step that it is unclear how error-prone it can be. In fact, there is evidence that there are usability issues when interacting with NFC and 2D Barcodes, and users may require explicit training to use them efficiently [O'Neill, et al., 2007]. Second, there might be difficulty locating such posters in every context (especially in visually clattered environments), or accessing them in the presence of many other users (e.g. around museum exhibits). Third, the interaction with the object (e.g. poster) can be seen as a privacy breach as everyone around the user can see the service she is accessing (also suggested in a longitudinal NFC study in [O'Neill, et al., 2007]). Fourth, it is not always acceptable to place such signs in prominent and accessible sites. For example, in world heritage locations (like the city of Bath) strict rules apply with regard to the quality and quantity of the visual cues one can impose. Similar restrictions might apply in privately owned social spaces. On the other hand, if these cues become subtler, they run the risk of going completely unnoticed. This is particularly true in the dynamic and widely diverse environment of mobile use, where attention might be split between crossing the road, conversing and keeping an eye for an open grocery shop.

Finally, there is an issue with personalisation and contextual relevance, as all available services in a location will have to be equally visible to all people at all times. Realistically however, only some people are going to find some of these services useful and only at certain times (depending on their actions, intentions etc).

We do not claim that the arguments presented here are proving beyond doubt that in-environment mobile service discovery is not feasible. On the contrary, and in some cases, good design and sufficient user training could provide viable solutions. However, everyday technologies (such as mobile phones) should be designed with the aim of eliminating user training.

APPENDIX III: FIRST INVESTIGATION EXPERIMENT MATERIAL

III.1 - Stimuli Characteristics

	Speech			Auditory Icons			Earcons		
Service	Duration (sec)	Vol. (dB)	Bit Rate (Kbps)	Duration (sec)	Vol. (dB)	Bit Rate (Kbps)	Duration (sec)	Vol. (dB)	Bit Rate (Kbps)
Calendar	2.04	90.8	1411	1.94	83.7	88	1.87	102.9	1411
Call	1.79	88.2	1411	2.19	95.5	176	1.75	101.3	1411
Downloads	1.37	91.3	1411	4.21	91.2	705	1.12	98.3	1411
Synchronise	2.19	85.7	1411	4.6	95.2	176	2.87	99.8	1411
I.R.	2.21	93.8	1411	2.93	96.3	89	3	101.5	1411
I.M.	2.09	88.2	1411	0.5	93.5	176	1.25	100	1411
SMS	2.12	90.2	1411	3.6	86.1	705	2.37	100.4	1411
TV	0.73	96.7	1411	3.91	85.7	705	3	98.9	1411
To-Do	2.1	88.5	1411	1.64	102.5	176	1.87	100.3	1411

	Duration		Volume		Bitrate	
	Ave	SD	Ave	SD	Ave	SD
Speech	1.85	0.50	90.38	3.31	1411.00	0.00
Auditory Icons	2.84	1.36	92.19	6.09	332.89	281.36
Earcons	2.12	0.72	100.38	1.39	1411.00	0.00

III.2 – Consent Form for Experiment

INFORMED CONSENT TO PARTICIPATE IN

An Evaluation of Auditory Notifications

An Evaluation of Auditory Notification is being conducted by Stavros Garzonis (csmsg@bath.ac.uk), Andrew Shakespear (as309@bath.ac.uk), Francis Binns (fb219@bath.ac.uk), Daniel Goldstien (dag21@bath.ac.uk), Erxiong (Ivan) Xu (ex200@bath.ac.uk), Michael Crocker (mjc25@bath.ac.uk) & Iain Kingston (isk20@bath.ac.uk) in the Department of Computer Science at the University of Bath.

You are being asked to take part in this study by doing a test of recognizing different sounds for different mobile services. Your participation in the evaluation will take less than 30 minutes.

The experiment will be recorded for analysis purposes only. Your responses will be stored anonymously to protect your privacy. If you wish, we shall send you a copy of any subsequent publications that use any of the data from the study. Potential benefits associated with the study include a greater understanding of user needs associated with mobile services.

If you agree to participate in this experiment as described, and for any relevant responses to be used in publications anonymously, please indicate your agreement by writing your name, e-mail address, then sign and date below. Thank you for your participation in this research.

Name:…………………………………………………

E-Mail:…………………………………………………

Signed:…………………………………………………

Date:…………………………………………………

III.3 – Experiment Instructions

Dear participant,

Thank you for taking part in our experiment today. We estimate that we will need no more than 30 minutes of your time. Your contact details and responses will not be passed to any third party and your anonymity will be preserved. Please be sincere throughout the experiment and remember that we are not evaluating your responses but our system.

We investigate 3 different ways for providing auditory notifications for services delivered through a mobile phone. You will be presented with a list of 9 services, which will be described to you by the experimenter. Then, you will hear a series of sound notifications and you will be asked to match each sound to the service you think it represents best.

Please answer as quickly and accurately as possible.

Procedure

1. Explanation of 3 different types of sounds: "speech", "natural sounds" and "earcons".
2. Presentation of examples of the 3 types of sounds and familiarisation with interacting with the plasma screen.
3. Presentation and explanation of the 9 mobile services.
4. You will evaluate the sound notifications for these services.

At the end we will ask you to fill in a questionnaire.

III.4 – Post-Test Questionnaire

Please indicate your choice by circling your answer or writing on the line provided.

Name: …………………………………………………………………..

Gender: M F

Age: …………

Occupation / course (if student):
…………………………………………………………………..

Do you own a mobile phone: Y N

If yes, how long have you owned a phone: ………. years ……….. months

- Hearing the ‘speech’ notifications clearly was:

Very Easy *Easy* *Neutral* *Hard* *Very Hard*

- Understanding the ‘speech’ notifications was:

Very Easy *Easy* *Neutral* *Hard* *Very Hard*

- Remembering the ‘speech’ notifications was:

Very Easy *Easy* *Neutral* *Hard* *Very Hard*

- “I liked the ‘speech’ notification”:

Strongly Agree *Agree* *Neutral* *Disagree* *Strongly Disagree*

- How long would it take you to learn and remember ‘speech’ notifications if they were on your mobile device?

Under a week *1-2 weeks* *2-3 weeks* *3-4 weeks* *4+ weeks* *Never*

Why do you think this?

- “I would like to have ‘speech’ notifications on my mobile phone”:

Strongly Agree *Agree* *Neutral* *Disagree* *Strongly Disagree*

Why do you think this?

- Hearing the 'natural sound' notifications clearly was:

Very Easy	*Easy*	*Neutral*	*Hard*	*Very Hard*

- Understanding the 'natural sound' notifications was:

Very Easy	*Easy*	*Neutral*	*Hard*	*Very Hard*

- Remembering the 'natural sound' notifications was:

Very Easy	*Easy*	*Neutral*	*Hard*	*Very Hard*

- "I liked the 'natural sound' notifications":

Strongly Agree	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*

- How long would it take you to learn and remember 'natural sound' notification if they were on your mobile device?

Under a week	*1-2 weeks*	*2-3 weeks*	*3-4 weeks*	*4+ weeks*	*Never*

Why do you think this?

- "I would like to have 'natural sound' notifications on my mobile phone":

Strongly Agree	*Agree*	*Neutral*	*Disagree*	*Strongly Disagree*

Why do you think this?

- Hearing the ‘earcon’ notifications clearly was:

Very Easy | *Easy* | *Neutral* | *Hard* | *Very Hard*

- Understanding the ‘earcon’ notifications was:

Very Easy | *Easy* | *Neutral* | *Hard* | *Very Hard*

- Remembering the ‘earcon’ notifications was:

Very Easy | *Easy* | *Neutral* | *Hard* | *Very Hard*

- “I liked the ‘earcon’ notifications”:

Strongly Agree | *Agree* | *Neutral* | *Disagree* | *Strongly Disagree*

- How long would it take you to learn and remember ‘earcon’ notification if they were on your mobile device?

Under a week | *1-2 weeks* | *2-3 weeks* | *3-4 weeks* | *4+ weeks* | *Never*

Why do you think this?

- “I would like to have ‘earcon’ notifications on my mobile phone”:

Strongly Agree | *Agree* | *Neutral* | *Disagree* | *Strongly Disagree*

Why do you think this?

- Which type of notification did you find easiest to recognise?
  (Order 1->3, with 1 being the easiest).

Earcons ……….

Natural ……….

Speech ……….

- For which type of notification do you feel you responded quickest?
  (Order 1->3, with 1 being the quickest).

Earcons ……….

Natural ……….

Speech ……….

- Which type of notification did you prefer over all?
  (Order 1->3, with 1 being your favourite choice).

Earcons ……….

Natural ……….

Speech ……….

Why do you think this?

- Which type of notification would you want to use on you mobile device?
  (You may tick more than one answer).

Earcons ……… Natural Sounds ………. Speech ……….

Why do you think this?

- Can you see any of the notifications getting developed for future use on mobile devices? (Indicate Y or N).

Earcons ……… Natural Sounds ………. Speech ……….

Why do you think this?

Thank you for your time!

APPENDIX IV – CARD-SORTING MATERIAL

IV.1 – List of Services & Descriptions

	SERVICE NAME	SERVICE DESCRIPTION
1	Voice call	The service providing voice calls
2	Video call	The service providing video calls
3	Conference call	The service providing conference calls
4	Answer phone	The service providing voice messages on answer-phone
5	Answer fax	The service providing fax messages on answer-fax
6	Text messaging (SMS)	The service providing text messages
7	Multimedia messaging (MMS)	The service providing multimedia messages
8	Chat	The service providing real time messages on a chat conversation
9	Location messaging	The service providing messages that are left for you only at specific locations
10	Instant messaging (e.g. MSN)	The service providing online status, messages, pictures etc. from people on your contact list (e.g. MSN)
11	Friends' location	The service providing information on the current location of people on your contact list
12	Calendar meeting entry	The service providing meeting reminders that are on your calendar application
13	Calendar birthday / anniversary entry	The service providing birthday or anniversary reminders that are on your calendar application
14	Calendar memo entry	The service providing memo reminders that are on your calendar application
15	To-do	The service providing reminders that are related to your to-do list items
16	Synchronise with personal device	The service providing synchronisation of your mobile device with your PC, laptop etc.
17	Sports information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to sports
18	Football information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to football
19	Cricket information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to cricket
20	Rugby information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to rugby
21	Horse racing information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to horse racing

22	Motorsports information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to motorsports
23	Golf information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to golf
24	Tennis information	The service providing live latest scores, results, news, pictures, videos, fixtures, league tables and statistics related to tennis
25	Online betting	The service providing the ability to place a bet and find out betting odds on any sport
26	Ringtones	The service providing downloads of different ringtones for personalising incoming calls
27	Music	The service providing downloads of music songs and videos, news and reviews
28	Games	The service providing downloads of games, news and reviews
29	Wallpapers	The service providing downloads of pictures that you can set as wallpapers or send to your friends
30	Music TV	The service providing live image of music bands performing
31	Music radio	The service providing live music radio channels
32	Mobile TV	The service providing a variety of TV channels
33	Film	The service providing news, screensavers, trailers & ringtones related to films
34	Celebrity gossip	The service providing celebrity magazines
35	Horoscopes	The service providing tarot readings, personalised horoscope games, daily star-sign alerts, location of nearest clairvoyants
36	Postcards	The service providing print-out and post delivery of pictures and messages you have on your phone
37	Photography	The service providing the ability to create and share online photo albums of your mobile pictures
38	Magazines	The service providing online magazines to browse
39	Comedy	The service providing news, interviews and reviews of stand up comics or funny ringtones downloads
40	News	The service providing the latest news from news agencies, newspapers and magazines
41	Weather	The service providing the weather forecast
42	National lottery	The service providing the latest results from national lottery draws, break down of the prizes, news on the winners and the ability to buy lottery numbers
43	Business news	The service providing the latest business headlines, company news and stock markets
44	Showbiz news	The service providing the latest news from a showbiz news desk
45	Find & Seek	The service providing the ability to find essential and entertainment services, restaurants, clubs, cinemas etc. Also

		provides with maps and directions to any destination
46	Timetables and reports	The service providing information on train and airplane timetables, delay reports and road traffic reports
47	Directions	The service providing maps with your location and route planner with driving or walking directions to any destination
48	Holidays and breaks	The service providing access to a travel shop with a variety of travel agencies and companies
49	Directory enquiries	The service providing telephone numbers and postal addresses of individuals and companies
50	Adult	The service providing access to services with adult content
51	Food and drink	The service providing a guide for pubs, restaurants, bars and clubs
52	Online account	The service providing space on a web server to save your personal content for backup

IV.2 – Consent Form for Card-Sorting

INFORMED CONSENT TO PARTICIPATE IN

Card Sorting Exercise

A study on Mobile Services Categorisation is being conducted by Stavros Garzonis (S.Garzonis@bath.ac.uk) in the Department of Computer Science at the University of Bath.

You are being asked to take part in this study by sorting cards in specified categories. Your participation in the evaluation should take less than 30 minutes.

The study will be recorded for analysis purposes only. Your responses will be stored anonymously to protect your privacy. If you wish, we shall send you a copy of any subsequent publications that use any of the data from the study. Potential benefits associated with the study include a greater understanding of and design for audio notifications of mobile services.

If you agree to participate in this experiment as described, and for any relevant responses to be used in publications anonymously, please indicate your agreement by writing your name, e-mail address, then sign and date below. Thank you for your participation in this research.

Name: …………………………………………………

E-Mail: …………………………………………………

Signed: ……………………………   Date: …………………………………

IV.3 – Open Card-Sorting Instructions

There are many ways to organise mobile data services in categories depending on the reasons behind the taxonomy and individual preference. A mobile network provider has hired you to create a new taxonomy of their services, having in mind that there will be a different **notification** for the availability of each category or subcategory of services.

For this task, you will be applying a common technique in HCI, called "card sorting". There are 52 "service" cards handed to you in a random order and a few blank "category cards" (as many as you need). Each service card has a service name and a short description of the service; all the provider's services are represented by one card each.

- Go carefully through all the services and group them in a way you think is most appropriate for similar notifications. All the services in the same category (or sub-category) will share the same alert notification when they are required to draw users' attention.
- When you are happy with the categories you have created, take one blank category card and give a name and a short description to the category matching each pile of services. Give a short justification on the reason(s) you have clustered these services together.
- You can also create sub-categories in each category or super-categories of two or more categories. Don't forget to name, describe and justify each new sub- or super-category by filling a blank category card. Create as many categories with as much nesting as you deem necessary.
- You can reconsider and reshuffle the services at any point, as many times as you want.
- When you have reached your final decision, number all category cards according to the level of depth of the hierarchy (e.g. 1.1, 1.2.1 or 2.3).
- Staple together each of the piles of cards (level 1 categories).
- Complete individually the questionnaires provided at the end of the task (one for each member).
- Put all piles in the envelope provided, along with the completed questionnaires.

IV.4 – Open Card-Sorting Post-Test Questionnaire

Name: ……………………………………………… Group: ......

Age: ……..

A. Mobile Usage

1. How long have you been using a mobile phone (approx.)?

   Years: Months:

2. How many times do you use your phone for calls?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 per month	Less than 1 per month

3. How many times do you use your phone for sms or mms?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 per month	Less than 1 per month

4. How many times do you use your phone for downloading pictures, music, ring-tones or games?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

5. How many times do you use your phone for taking photographs?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

6. How many times do you use your phone's to-do list

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

7. How many times do you use your phone for calendar entries/reminders

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

8. How many times do you use your phone to seek information about sports, entertainment etc?

More than 5 per day	2 to 4 Per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

<u>B. Card Sorting</u>

9. "I found the task of categorisation to be…"

Very Difficult	Difficult	Neutral	Easy	Very Easy	No Opinion

Comments:

10. "I had difficulty understanding what each service does"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

11. "The services fitted well in the categories I created"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

12. "There are services that could be under more that one category"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

13. "I found it useful to be able to create sub-categories"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

C. Group Collaboration

14. "Overall the collaboration of the group was…"

Very Bad	Bad	Neutral	Good	Very Good	No Opinion

Comments:

15. "All members of the group contributed equally to the task"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

16. "We had disagreements on the end result"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

17. "The end result is better than if I had to do it alone"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

Any other comments:

Thank you for your time.

IV.5 – Closed Card-Sorting Instructions

Mobile Services Categorisation for Notification Purposes

There are many ways to categorise the services offered by mobile network providers. We are interested in creating a categorisation so that services in the same categories share the same audio notification (if the user desires one). So, how should the services be categorised so that they share similar audio notifications?

In front of you there are 10 categories organised in 4 branches (pink cards). You will be given a pile of 52 green cards (in random order), each one of which represents a mobile service. Please read the description of the service on each card and place it under the category you think it fits best. Also, please tick on each card the appropriate 'fit' option (Perfect, Good, Fair) according to how well you think the card fits in the category you chose. You can change your mind and re-allocate the cards as many times as you wish throughout the exercise.

You are encouraged to think aloud throughout the exercise.

Please note that there is no 'right' way of completing this exercise.

IV.6 – Closed Card-Sorting Post-Test Questionnaire

Name: ………………………………………………

Age: ……..

Gender: ……

A. Mobile Usage

1. How long have you been using a mobile phone (approx.)?

   Years: Months:

2. How many times do you use your phone for calls?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 per month	Less than 1 per month

3. How many times do you use your phone for sms or mms?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 per month	Less than 1 per month

4. How many times do you use your phone for downloading pictures, music, ring-tones or games?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

5. How many times do you use your phone for taking photographs?

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

6. How many times do you use your phone's to-do list

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

7. How many times do you use your phone for calendar entries/reminders

More than 5 per day	2 to 4 per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

8. How many times do you use your phone to seek information about sports, enter-tainment etc?

More than 5 per day	2 to 4 Per day	Once per day	1-6 per week	1-4 Per month	Less than 1 per month

B. Card Sorting

9. "I found the task of categorisation to be…"

Very Difficult	Difficult	Neutral	Easy	Very Easy	No Opinion

Comments:

10. "I had difficulty understanding what each service does"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

11. "The services fitted well in the categories given"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

12. "There are services that could fit under more that one category"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

Comments:

13. "I would like to have created my own categories"

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	No Opinion

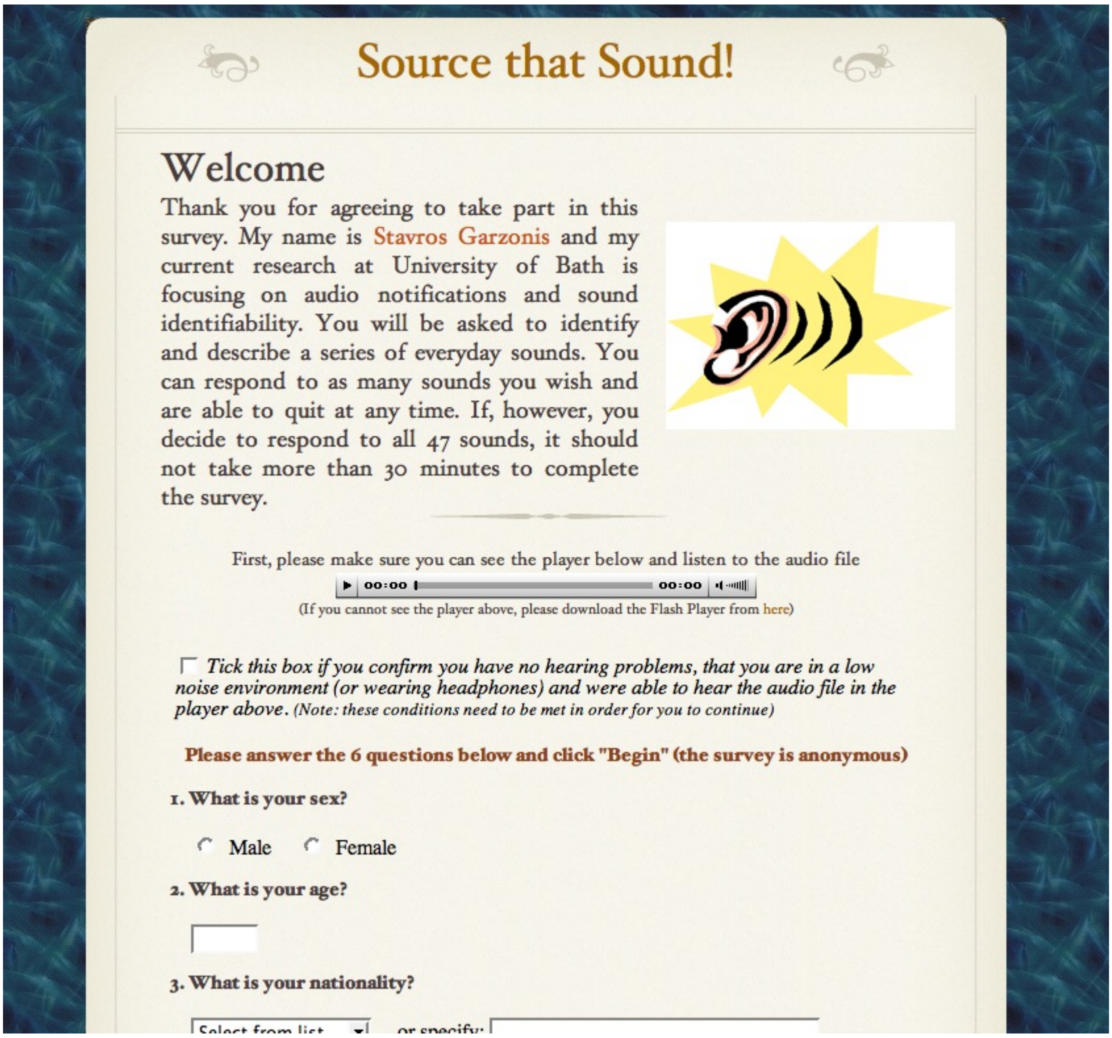
Comments:

Any other comments:

Thank you for your time.

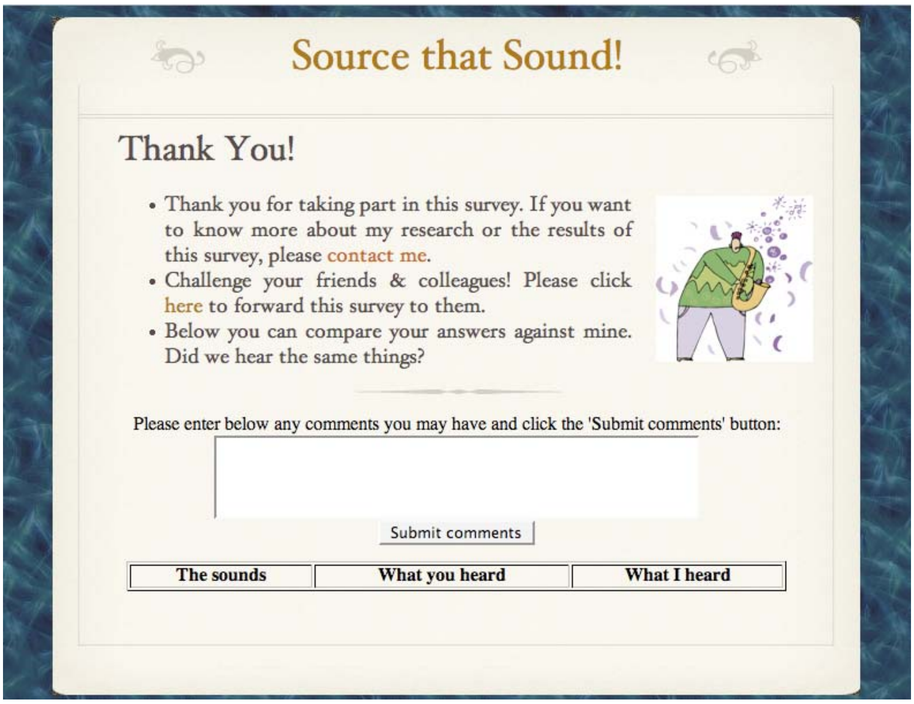
APPENDIX V – ONLINE SURVEYS MATERIAL

V.1 – Surveys Screenshots



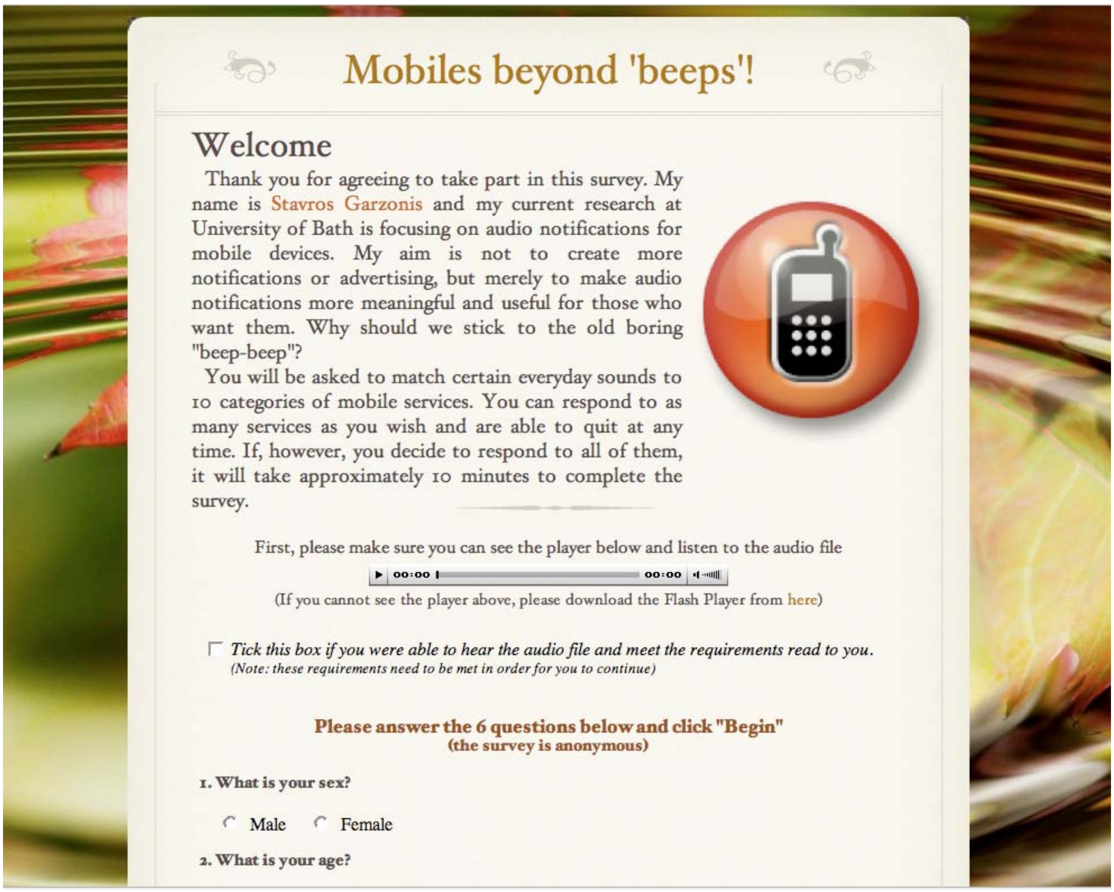
Survey 1: Intro page

Survey 1: Main Page

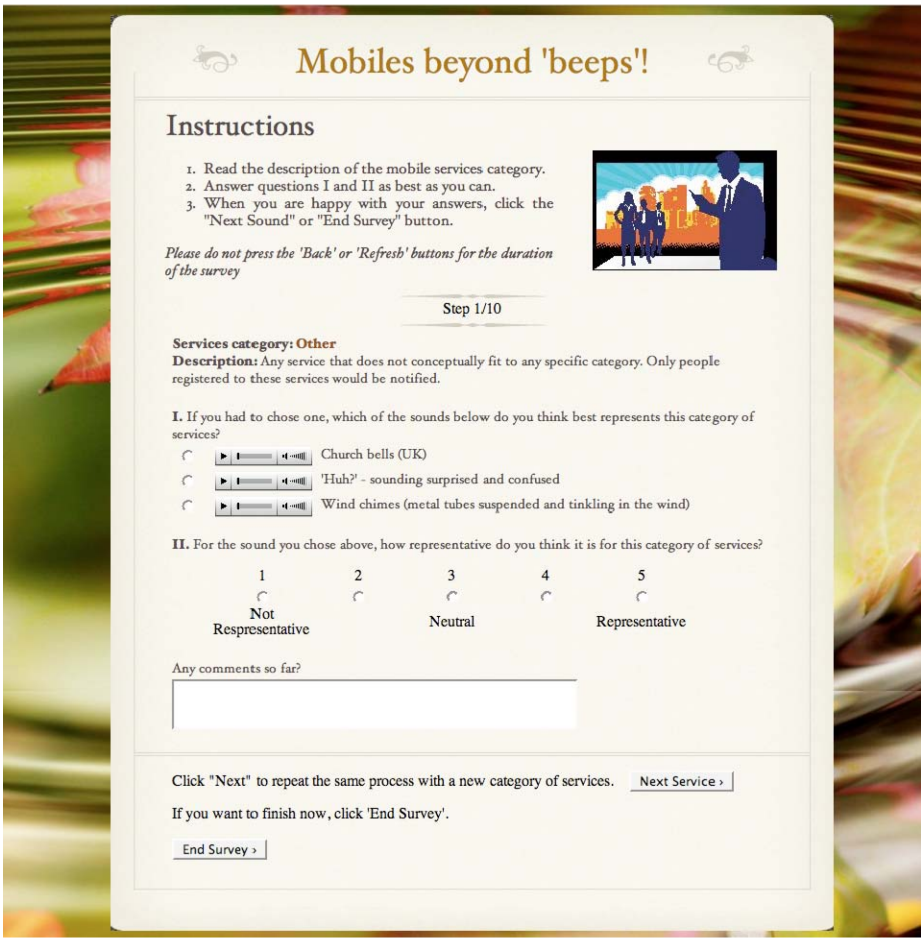


Survey 1: Final Page

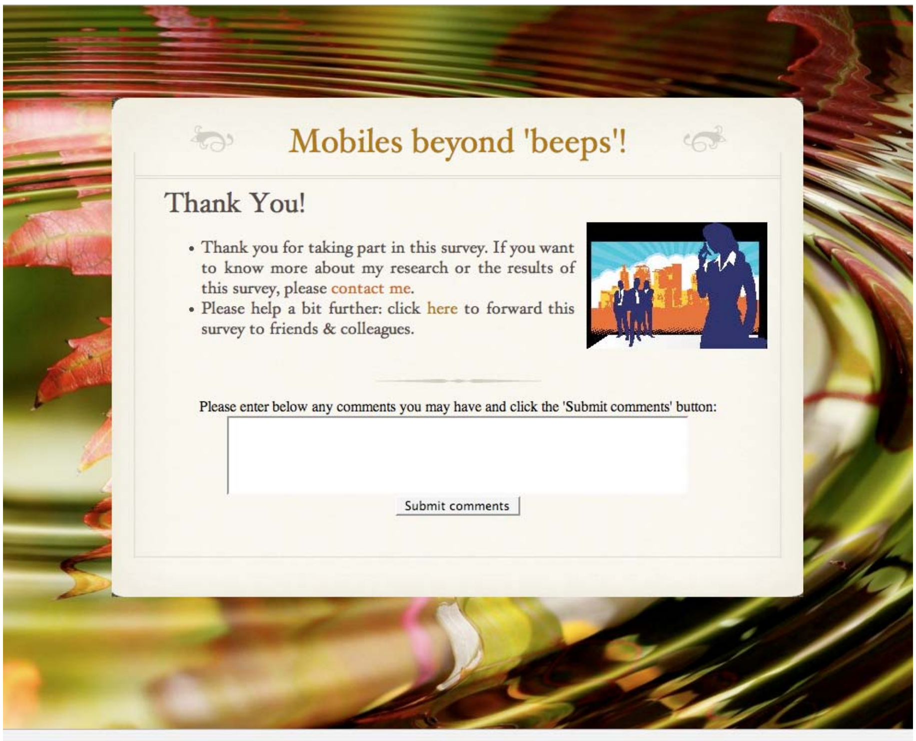
Note: List of responses and sound descriptions were appended at the bottom of this page.



Survey 2: Intro Page



Survey 2: Main Page



Survey 2: Final Page

V.2 – List of Categories of Services & Descriptions

	Category Name	Description
1	News Information	Information relating to news articles. Only people registered to receive specific type of news on specific topics would be notified (e.g. "UK internal affairs", "US economy")
2	Sports Information	Information relating to sports news. Only people registered to receive specific information on specific sports would be notified (e.g. "today's match results" or "Manchester United club affairs")
3	"Here and Now" Information	Information relating to users' current location or situation, such as driving/walking directions, travelling information, traffic reports or city-guide hotspots near them. Only people registered to specific topics and in specific situations will be notified (e.g., "directions to the nearest petrol station" or "going-out options")
4	Entertainment Downloads	Availability of entertainment material, such as games, ringtones, wallpapers, movie trailers, music videos for downloading. Only people registered to specific downloads will be notified (e.g. "sci-fi movie wallpapers", "U2 music video releases")
5	Entertainment Live	Availability of live entertainment material, such as radio broadcasts, TV shows or peer-to-peer video games. Only people registered to specific type of events will be notified (e.g., "The Simpsons broadcasts" or "Pacman competitions")
6	Incoming Calls	Incoming calls, such as voice, video or conference calls
7	Incoming Messages	Incoming text (SMS), multimedia (MMS) message or instant messaging communication (e.g. MSN messenger)
8	Self Reminders	Reminders users have set for themselves, such as calendar entries, birthdays or to-do list items (e.g., "John's birthday" or "buy milk")
9	Backup Reminders	Reminders related to backing up or synchronising devices (e.g., "backup pictures from mobile device to online account" or "conflicts while synchronising calendar between mobile device and desktop computer")
10	Other Services	Any service that does not conceptually fit in any specific category. Only people registered to these services would be notified

APPENDIX VI – EARCONS EXPERIMENTS MATERIAL

VI.1 – Consent Form for Experiments 1 & 2

INFORMED CONSENT TO PARTICIPATE IN

Learn the sound of mobile services

About this study

This experiment is being conducted by Stavros Garzonis (S.Garzonis@bath.ac.uk) for the Department of Computer Science at the University of Bath. Potential benefits associated with the study include a greater understanding for designing non-speech audio notifications for mobile services.

You will be presented with categories of mobile services and a series of non-speech sounds associated with them. You will then be asked to learn and recall these associations. You will be able to withdraw from this process at any time, should you request so. Your participation in this study should take no more than 30 minutes.

Confidentiality

The study will be recorded for analysis purposes only. Your responses will be stored anonymously to protect your privacy. If you wish, we shall send you a copy of any subsequent publications that use any of the data from the study.

Participation agreement

If you agree to participate in this experiment as described, and for any relevant responses to be used in publications anonymously, please indicate your agreement by writing your name, e-mail address, then sign and date below. Thank you for your participation in this research.

Name: ……………………………………………………

E-Mail: ……………………………………………………

Signed: …………………………… Date: ………………………………….

VI.2 – Experiments 1 & 2 Instructions

Note: The following instructions were presented for earcons training and testing in the form of a slide slow. The instructions for the simple tones (in Experiment 1 only) were identical, apart from the words in bold font.

Learn the sound of mobile services

- You will be presented with 10 mobile services
- You will hear 10 **musical sounds** (one for each service)
- You will be given 5 minutes to learn the sound-service associations
- Then, you will be tested on the associations (3 times)
- Do not worry if you do not perform very well – just try your best!

Mobile Services – Selection Keys

F1	News Information	Services from the world
F2	Sports Information	
F3	Here and Now Information	
F4	Entertainment Downloads	
F5	Entertainment Live	
F6	Incoming Calls	Services from other users
F7	Incoming SMS Messages	
F8	Self Reminders	Services from myself
F9	Backup Reminders	
F10	Other Services	Rest of services

Training

- You will now be given 5 minutes to learn the sounds that correspond to each service
- Click on the button (F1-F10) next to each service to hear its sound
- Click to begin

Learn the sound of mobile services

- You will now hear the **musical sounds** in a random order
- You are asked to identify the service each sound corresponds to by pressing one of the **F1-F10** keys on your keyboard
- Sounds will be repeated up to 3 times if you do not respond immediately
- You will be given feedback at the end of each round
- You will repeat this process 3 times

Please respond as quickly and accurate as possible

VI.3 – EXPERIMENTS 1 & 2 PRE-TASK QUESTIONNAIRE

A little about you

Please indicate your choice by circling your answer or writing on the line provided.

Name: …………………………………………………………………………..

Gender: M F

Age: …………

Occupation / course (if student): …………………………………………………………………………..

Do you have any hearing difficulties?

Y N

I am a professional musician

Y N

I play a musical instrument and/or sing (e.g. in a band or a choir)

Y N

I would rate my musical ability:

Very Low *Low* *Neutral* *High* *Very High*

VI.4 – Experiments 1 & 2 Post-Test Questionnaire

Note: The following questions were displayed on screen and answers were captured via MediaLab.

I was able to hear all sounds

- Very clearly
- Clearly
- Somewhat clearly
- Not clearly
- Not clearly at all

Learning the associations was

- Very Easy
- Easy
- Neutral
- Difficult
- Very Difficult

If I wanted to be notified for any of the mobile services presented to me, and in situations where sound notifications are generally welcome, I would prefer

- Any of the sounds I just heard
- Most of the sounds I just heard
- Some of the sounds I just heard
- Few of the sounds I just heard
- None of the sounds I just heard

I was able to distinguish which group of services each sound belonged to

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree

Please type below if you have any comments about the notifications you just heard

VI.5 – Experiment 1 Comparative Questionnaire

Note: The following questions were displayed on screen and answers were captured via MediaLab.

Comparing the 2 sets of sounds, I preferred listening to

- The simple tones better
- The musical sounds better
- Both sets of sounds equally
- Neither set
- Some of one set and some from the other set

If I wanted to be notified for any of the mobile services presented to me, and in situations where sound notifications are generally welcome, I would prefer

- To have either set of sounds
- To have the simple Tones
- To have the musical sounds
- To have a different set of sounds
- Not to have any audio notifications at all

If I wanted to be notified for any of the mobile services presented to me, and in situations where sound notifications are generally welcome, I would prefer

- To mainly have non-speech notifications
- To mainly have speech notifications
- To have some speech notifications and some non-speech notifications, according to the service I am notified about
- To have some speech notifications and some non-speech notifications, according to the situation (where I am, what I am doing, who I am with etc.)
- None

Comparing to the simple tones, learning the associations with the musical sounds was:

- Much harder
- Harder
- About the same
- Easier
- Much easier

Please type below if you have any overall comments

APPENDIX VII – COMPARISON STUDY MATERIAL

VII.1 – Consent Form for Labs 1 & 2

INFORMED CONSENT TO PARTICIPATE IN

Audio Notifications for Mobile Services (Lab 1)

About this study

This experiment is being conducted by Stavros Garzonis (S.Garzonis@bath.ac.uk) and Simon Jones (S.Jones@bath.ac.uk) for the Department of Computer Science at the University of Bath. Potential benefits associated with the study include a greater understanding of and design for audio notifications of mobile services.

You will be presented with categories of mobile services and a series of non-speech sounds. You will then be asked to guess the association between the sounds and the services. You will be able to withdraw from this process at any time, should you request so. Your participation in this study should take no more than 40 minutes.

Confidentiality

The study will be recorded for analysis purposes only. Your responses will be stored anonymously to protect your privacy. If you wish, we shall send you a copy of any subsequent publications that use any of the data from the study.

Participation agreement

If you agree to participate in this experiment as described, and for any relevant responses to be used in publications anonymously, please indicate your agreement by writing your name, e-mail address, then sign and date below. Thank you for your participation in this research.

Name: ……………………………………………………

E-Mail: ……………………………………………………

Signed: ……………………………   Date: …………………………………

VII.2 – Consent Form for Field Study

INFORMED CONSENT TO PARTICIPATE IN

Audio Notifications for Mobile Services (field study)

About this study

This experiment is being conducted by Stavros Garzonis (S.Garzonis@bath.ac.uk) and Simon Jones (S.Jones@bath.ac.uk) for the Department of Computer Science at the University of Bath. Potential benefits associated with the study include a greater understanding of and design for audio notifications of mobile services.

An application developed by us will be installed on your mobile phone. You will be expected to respond to approximately 20 non-speech audio notifications at random intervals throughout each day of the study (10am to 9pm). With each sound, you will be asked match it to a service category and to answer a couple of questions (presented on your phone) with regard to your current context and sound preference. You will have the choice to reject the incoming notifications and not hear the sounds every time. No sounds will be heard if your phone is set to 'silent' or 'meeting' profiles. You will be able to withdraw from this process at any time, should you request to do so. The duration of this study will be 1 week long and the 3 participants with the most on-time responses will be presented with prizes (Amazon vouchers of £50, £20 and £10).

Confidentiality

The study will be recorded for analysis purposes only. Your responses will be stored anonymously to protect your privacy. If you wish, we shall send you a copy of any subsequent publications that use any of the data from the study.

Participation agreement

If you agree to participate in this experiment as described, and for any relevant responses to be used in publications anonymously, please indicate your agreement by writing your name, e-mail address, then sign and date below. Thank you for your participation in this research.

Name: ……………………………………………………

E-Mail: ……………………………………………………

Signed: ……………………………   Date: …………………………………

VII.3 – Lab 1 Pre-Test Questionnaire

Note: The following questions were displayed on screen and answers were captured via MediaLab.

Do you have any hearing difficulties?

- Yes
- No

Are you (or have been) a professional musician/singer?

- Yes
- No

Do you (or did you) sing and/or play a musical instrument?

- Yes - quite well
- Yes - not so well
- No

In general, how would you rate your musical natural ability?

- Very Low
- Low
- Medium
- High
- Very High

VII.4 – Lab 1 & Lab 2 Instructions

Note: The following instructions were presented in a form of a slide slow. They reflect only one condition of the experiments

Guess the meaning of sounds

- You will be presented with 12 mobile services (descriptions and examples will be given)
- You will hear a series of **everyday sounds** and a series of **musical sounds**
- Each sound corresponds to only 1 service
- You will be asked to guess which service corresponds to each of the sounds you hear

Part 1 - Everyday sounds

- You will now begin the evaluation of the **everyday sounds**
- Upon hearing a notification, choose the service you think it corresponds to by pressing one of the **F1-F12** keys on your keyboard
- You will be given no feedback on your guesses
- You may need to respond to some sounds more than once
- If you do not respond immediately, each sound will repeat up to 3 times

Please respond as accurately and quickly as possible

Part 2 - Musical sounds

- You will now begin the evaluation of the **musical sounds**
- Upon hearing a notification, choose the service you think it corresponds to by pressing one of the **F1-F12** keys on your keyboard
- You will be given no feedback on your guesses
- You may need to respond to some sounds more than once
- If you do not respond immediately, each sound will repeat up to 3 times

Please respond as accurately and quickly as possible

VII.5 – Lab 1 & Lab 2 Post-Test Questionnaire

Note: The following questions were displayed on screen and answers were captured via MediaLab. The musical sounds questionnaire was equivalent to the everyday sounds questionnaire presented here.

I was able to hear all everyday sounds

- Very clearly
- Clearly
- Not clearly
- Not clearly at all

On guessing the associations for the everyday sounds on average I was

- Very confident
- Confident
- Neutral
- Not confident
- Not confident at all

My certainty about the associations I proposed for the everyday sounds was

- The same for all everyday sounds
- Somewhat varying amongst everyday sounds
- Greatly varying amongst everyday sounds

If I wanted to be notified for any of the mobile services presented to me, and in situations where sound notifications are generally welcome, I would like to have

- All of the everyday sounds I just heard
- Most of the everyday sounds I just heard
- Some of the everyday sounds I just heard
- Few of the everyday sounds I just heard
- None of the everyday sounds I just heard

I would like to have the sound I am hearing right now as a notification on my mobile phone

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree

Please type below if you have any comments about the everyday sounds notifications

VII.6 – Lab 1 & Lab 2 Post-Test Comparative Questionnaire

Note: The following questions were displayed on screen and answers were captured via MediaLab.

Comparing the 2 sets of sounds, I preferred listening to

- Both sets equally
- The musical sounds better
- The everyday sounds better
- Neither set
- Some from one set and some from the other set

My certainty about the associations I proposed was

- Higher for the everyday sounds
- Higher of the musical sounds
- Nearly equal for both sets of sounds

If I wanted to be notified for any of the mobile services presented to me, and in situations where sound notifications are generally welcome, I would prefer

- To have either set of sounds
- To have the musical sounds
- To have the everyday sounds
- To have a combination of the two sets of sounds
- To have a completely different set of sounds
- Not to have any audio notifications at all

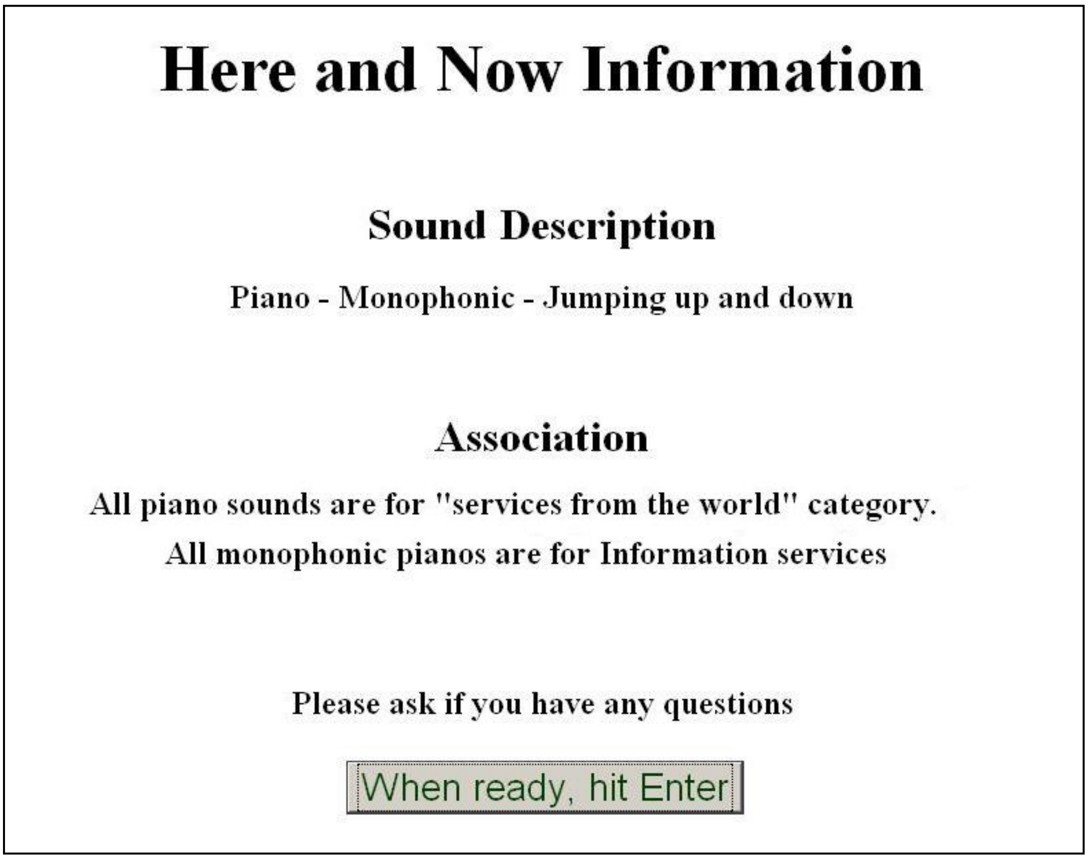
If I wanted to be notified for any of the mobile services presented to me, and in situations where sound notifications are generally welcome, I would prefer to have

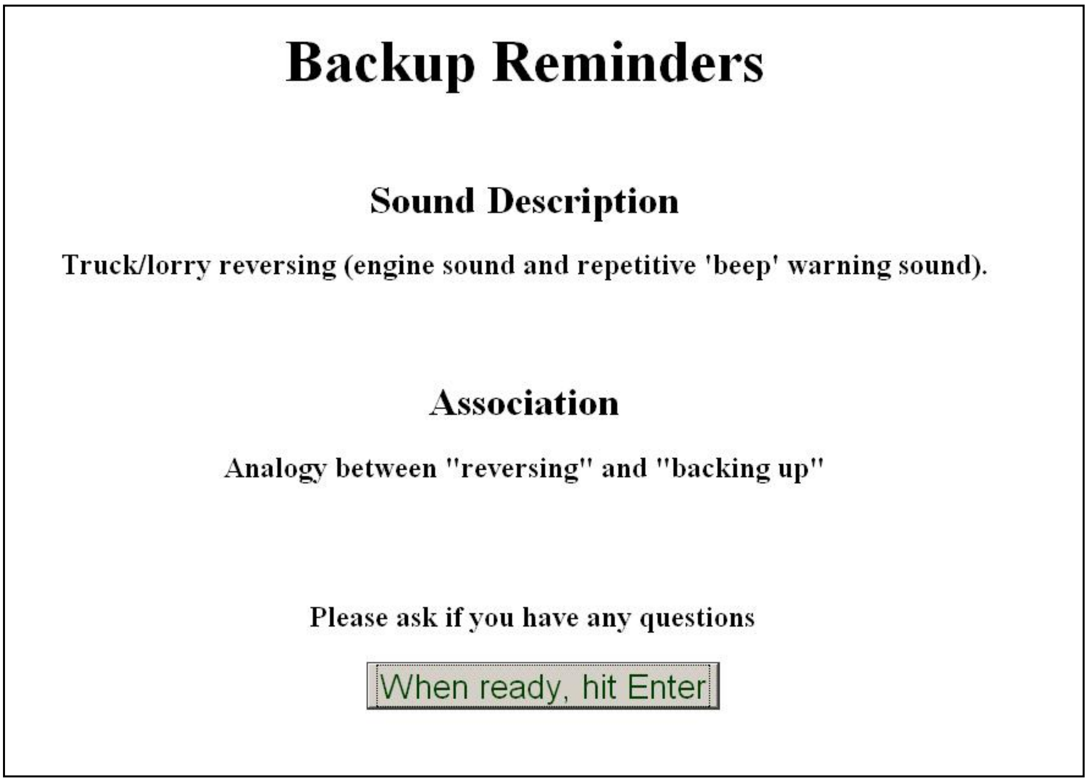
- Mainly non-speech notifications
- Mainly speech notifications
- Some speech notifications and some non-speech notifications, according to the service I am notified about
- Some speech notifications and some non-speech notifications, according to the situation (where I am, what I am doing, who I am with etc.)

Thank you!

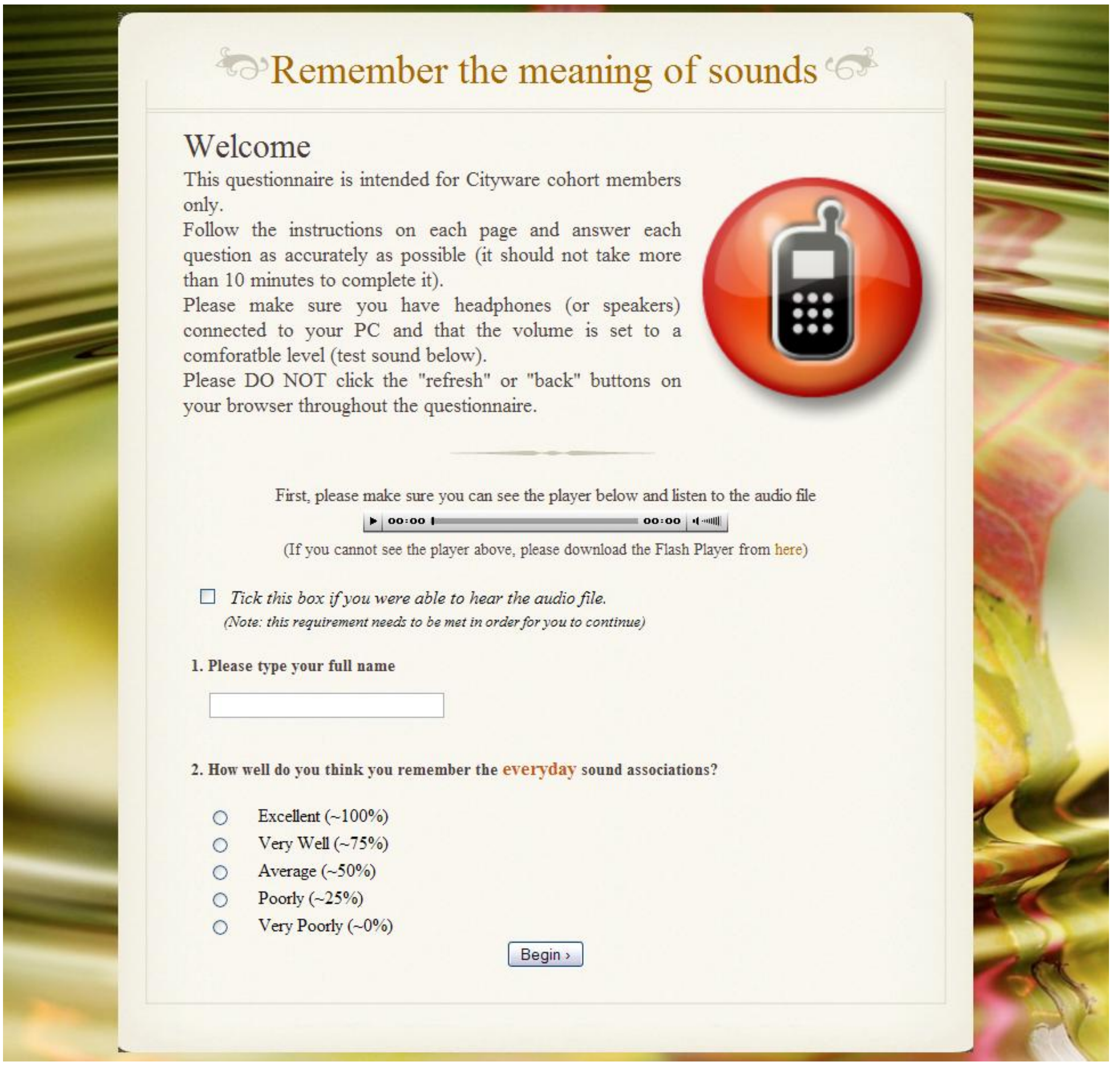
- You have completed the experimental procedure
- Please make your way to the buffet area
- Instructions for the mobile phone study will be given shortly

VII.7 – Examples of Screenshots for the Sound-Service Links in Lab 2

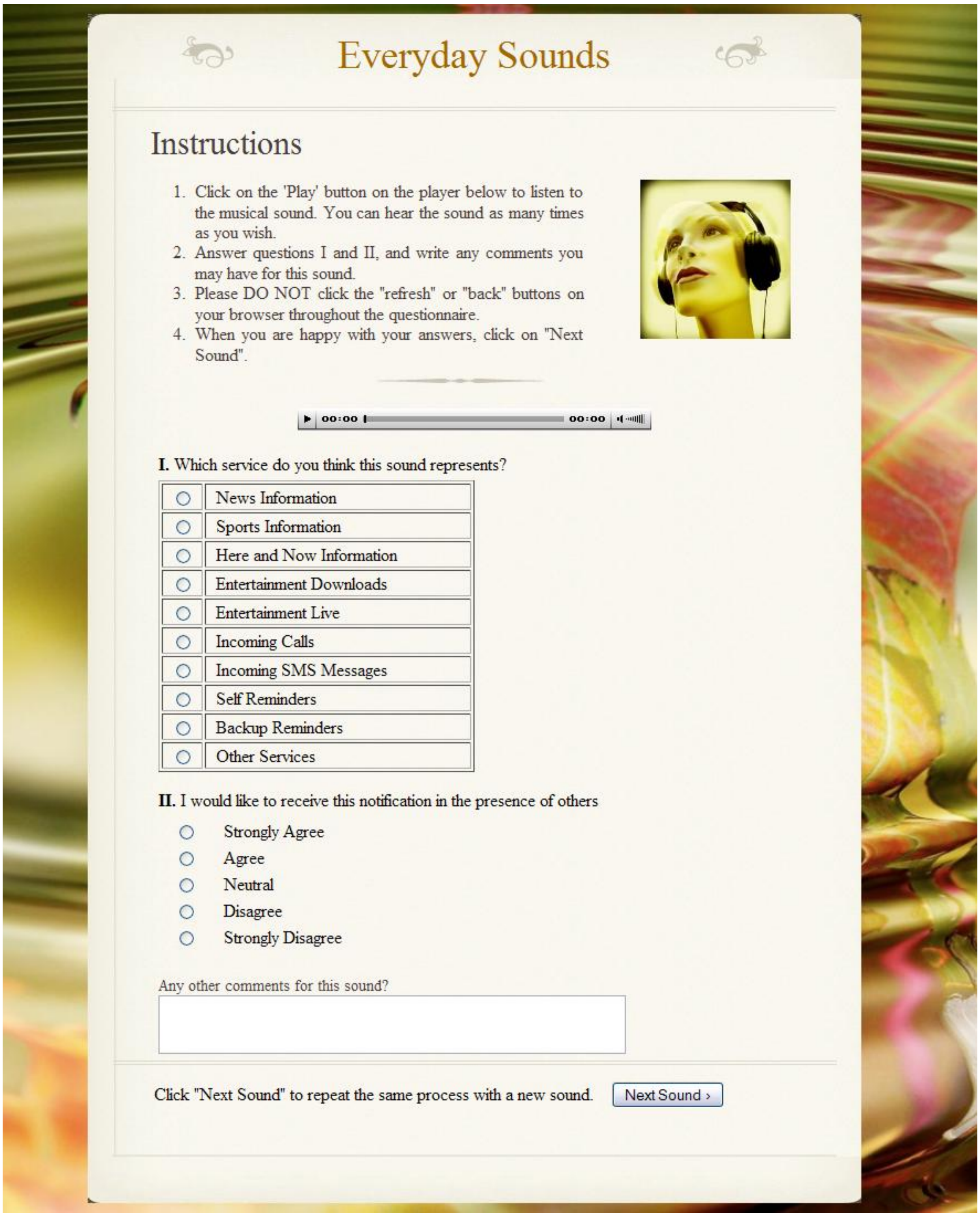




VII.8 – Web 1 & Web 2 Example Screenshots

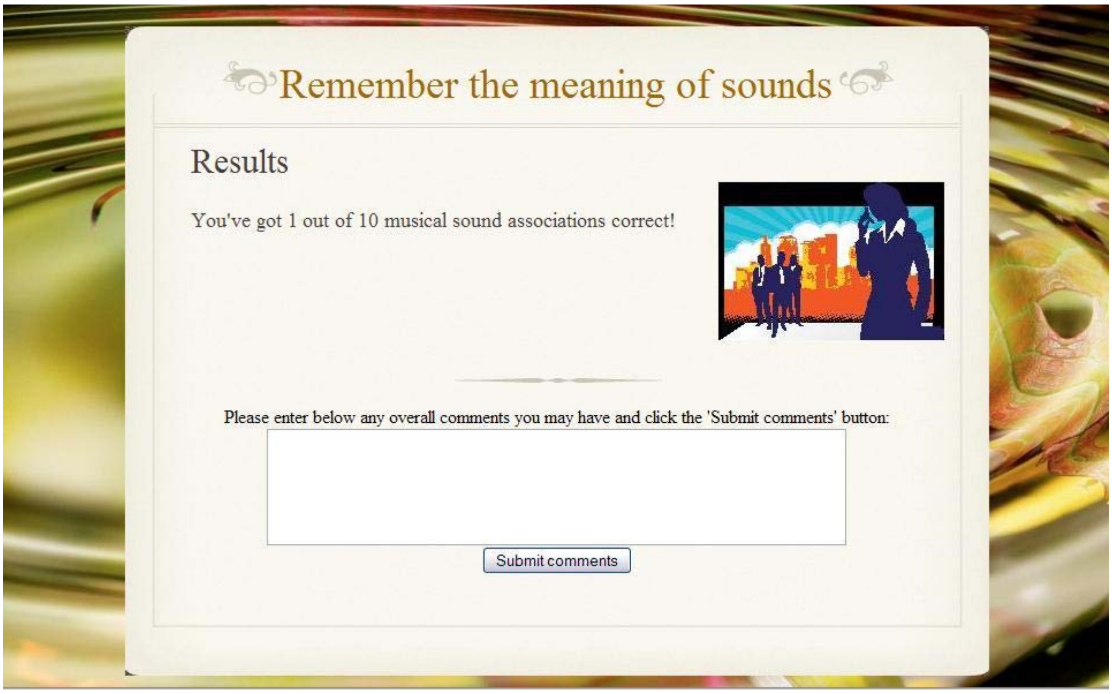


Web 1 and Web 2 Intro Page

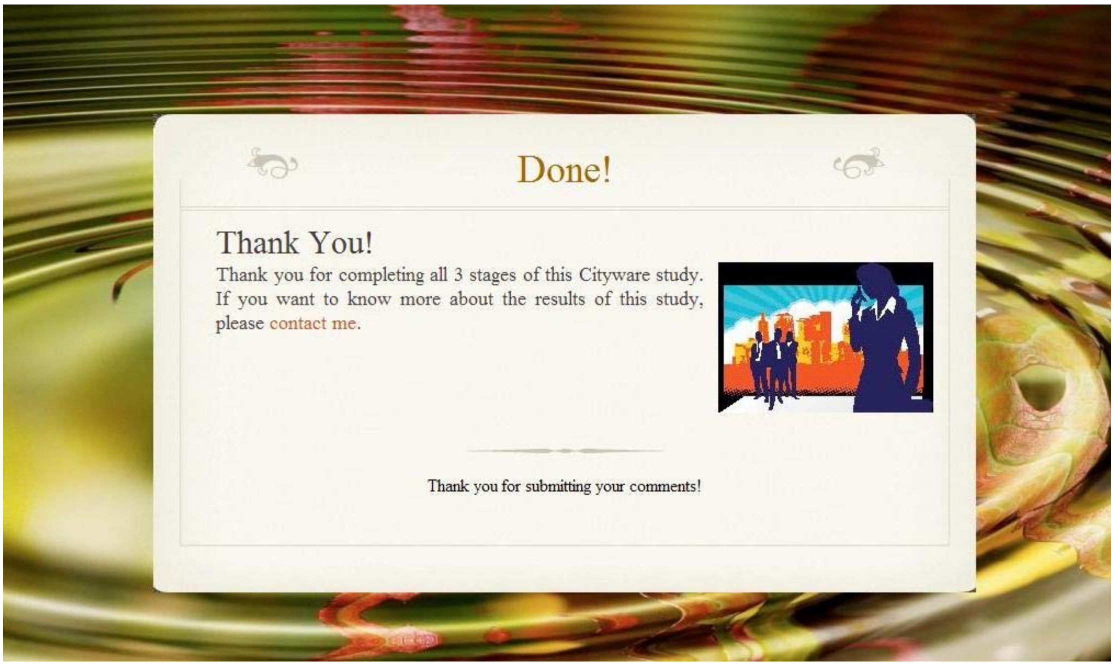


Web 1 & Web 2 Main Page

Web 1 & Web 2 Results Page 1



Web 1 & Web 2 Results Page 2



Web 1 & Web 2 Thank You Page