Design exploration in interactive sonification

Mikael Fernström
Interaction Design Centre
Department of Computer Science and Information Systems
Faculty of Science and Engineering
University of Limerick

A thesis submitted for the degree of
Philosophiae Doctor (Ph.D.) in Interaction Design
Supervisor: Prof. Liam Bannon

Submitted to the University of Limerick, August 2011.
PhD Committee

1. Supervisor: Professor Liam Bannon
2. External Examiner 1: Prof. John Bowers, Goldsmiths University of London
3. External Examiner 2: Dr. Dermot Furlong, Trinity College, Dublin
4. Chair: Dr. Michael English

Day of Defence: 23rd of August 2011
Revised: 23rd of February 2012.

Signature of Chair of PhD Committee: _______________________________
# Table of Contents

ABSTRACT .................................................................................................................. I

DECLARATION .......................................................................................................... III

PAPERS INCLUDED IN THE THESIS ........................................................................ V

OTHER RELATED PUBLICATIONS BY THE AUTHOR ............................................ IX

ACKNOWLEDGEMENTS ............................................................................................. XIII

LIST OF FIGURES ...................................................................................................... XV

1 INTRODUCTION ...................................................................................................... 1

1.1 BEYOND THE SONIFIED DESKTOP METAPHOR ............................................. 1

2 PERSPECTIVES ON SOUND .................................................................................. 5

2.1 SOUND IN THE HUMAN WORLD ...................................................................... 5

2.2 LISTENING ......................................................................................................... 6

2.3 DIFFERENT FORMS OF LISTENING ................................................................ 6

2.4 SUMMARY .......................................................................................................... 10

3 SOUND DESIGN ..................................................................................................... 11

3.1 FOLEY SOUND .................................................................................................. 12

3.2 IMPLICATIONS FOR SONIC INTERACTION DESIGN .................................... 13

3.3 SUMMARY .......................................................................................................... 14

4 AUDITORY DISPLAY ............................................................................................... 15

4.1 WARNINGS, ALARMS AND NOTIFICATIONS .................................................. 16

4.2 NON-INTERACTIVE SONIFICATION .................................................................. 17

4.3 INTERACTIVE SONIFICATION ......................................................................... 17

4.4 SONIFICATION IN GAMES AND ENTERTAINMENT ..................................... 18

4.5 USER INTERFACES MADE OF SOUND ............................................................. 19

4.6 SUMMARY .......................................................................................................... 21

5 SONIC INTERACTION DESIGN ............................................................................... 23

5.1 MAKING USER INTERFACE WIDGETS .............................................................. 24

5.2 AUDITORY AFFORDANCES ............................................................................. 25

5.3 DESIGNING SOUND OBJECT MODELS .............................................................. 28

5.4 DESIGNING INTERACTIVE SOUND OBJECT MODELS .................................. 29

5.5 FRAMEWORK FOR DESIGN OF AUDITORY WIDGETS ................................... 31

5.6 RESEARCH QUESTIONS ...................................................................................... 32

5.7 OPERATIONALISATION ...................................................................................... 32

5.8 SUMMARY .......................................................................................................... 33
6 CONTRIBUTIONS OF THE PRESENT WORK .................................................. 35

6.1 PAPER I & II: AFTER DIRECT MANIPULATION – DIRECT SONIFICATION; REFLECTIONS ON
SONIC BROWSING ........................................................................................... 35

6.1.1 Summary ................................................................................................. 38

6.2 PAPER III: SOUNDING OBJECTS ................................................................. 39

6.2.1 Sound object models ............................................................................. 39

6.2.2 Example: Liquid sounds ........................................................................ 40

6.2.3 Summary ................................................................................................. 43

6.3 PAPER IV: HCI DESIGN AND INTERACTIVE SONIFICATION FOR FINGERS AND EARS .... 45

6.3.1 Choosing sounds that are easy to recognise ......................................... 45

6.3.2 Example of an Auditory User Interface using direct manipulation ......... 46

6.3.3 Summary ................................................................................................. 48

6.4 PAPER V: THE SHANNON PORTAL INSTALLATION: INTERACTION DESIGN FOR PUBLIC
PLACES ................................................................................................................. 49

6.4.1 The challenge of auditory display in a public environment .................... 49

6.4.2 Evaluation ............................................................................................... 50

6.4.3 Summary ................................................................................................. 52

7 GENERAL DISCUSSION .................................................................................. 55

7.1 METHODOLOGY ISSUES ............................................................................. 59

8 CONCLUSIONS .................................................................................................. 63

9 FUTURE RESEARCH ....................................................................................... 65

10 REFERENCES ................................................................................................... 67

Published Papers I – V
Abstract
This thesis examines interactive sonification, in particular the design, implementation and evaluation of user interface components using sound. It consists of a series of design interventions and explorations, including preliminary empirical investigations, showing features of this space. A novel search method and tool for finding sound files was created. The Sonic Browser utilised the human ability to listen to multiple simultaneous sounds, and facilitated users to switch attention between the sounds while navigating a virtual soundscape. The notion of sound object models, based on physical modelling, or physics-inspired modelling, was developed and explored. The sound object model approach can be used for more responsive and expressive user interfaces based on sound. A method for measuring causal uncertainty in listening tests was improved and modified to also facilitate the design of user interface metaphors based on short narratives provided by participants in listening tests. A novel auditory widget, a pseudo-haptic soft-button for touch-based interaction, was exemplified and explored, which potentially can be used for non-visual user interfaces, in particular for mobile and wearable computing. Finally, an interactive sonification for improving interaction with computers in public places, for both single and multiple users using full body gestures, was exemplified and explored.

The research in this thesis covers perspectives on sound and listening, sound design, auditory display, the making of auditory user interface components and outlines a framework for exploring the design of auditory widgets.
Declaration

This thesis is presented as fulfilment of the requirements for a degree of Doctor of Philosophy at the University of Limerick, Faculty of Science and Engineering, in the Department of Computer Science and Information Systems. It is entirely my own work and has not been submitted to any other university, or higher education institution, or for any other academic award in this University. Where use has been made of the work of other people, it has been fully acknowledged and referenced.

Signature

Mikael Fernström
Papers included in the thesis
The effectiveness of providing multiple-stream audio to support multimedia browsing on a computer was investigated through the iterative development and evaluation of a series of Sonic Browser prototypes. The data set used was a database containing music. Interactive sonification was provided in conjunction with simplified human-computer interaction sequences. The paper explored the extent to which interactive sonification with multiple-stream audio could enhance multimedia browsing tasks, compared to interactive sonification with single-stream audio support. With ten users, it was found that with interactive multiple-stream audio the users could accurately complete the multimedia browsing tasks significantly faster than those who had single-stream audio support. This paper defines the concepts Direct Sonification and Sonic Browsing.
This work was mainly the work of M. Fernström in terms of concept development, initial software development, experiment design and evaluation. As iterative development was used, later implementations were developed with assistance from final year students (David Boland, John Burton, Hugh Roberts) and one graduate student (Caolan McNamara).
Paper I was originally published in the proceedings of the International Conference on Auditory Display (ICAD) 1998 in Glasgow. ICAD¹ publications are peer-reviewed. This paper was later subjected to further peer-review and re-published with a commentary (Paper II) in Transactions on Applied Perception of the Association of Computing Machinery (ACM).²

This paper reports and reflects upon continued research in the area of sonic browsing and how an interactive sonification approach can be used for browsing collections of recorded everyday sounds and for perception experiments.

¹ www.icad.org
² www.acm.org
This work was mainly the work of M. Fernström in terms of concept development and design. The paper covers the continued and expanded development of interactive sonification and sonic browsing in domains such as management of corpora of sound files and perceptual experiments. The later prototypes were implemented and evaluated in collaboration with a postgraduate student (Eoin Brazil).


Interactive systems, virtual environments, and information display applications may need dynamic sound models rather than simply audio reproductions. This implies three levels of research: auditory perception, physics-based sound modelling, and expressive parametric control. Parallel progress along these three lines leads to effective auditory displays that potentially can complement or substitute visual displays. This paper reports the definition and development of sound object models, their implementation and evaluation.

The work reported in this paper was a result of an EU-Future and Emerging Technology project, the Sounding Object. The main collaborators Davide Rocchesso, Roberto Bresin and Mikael Fernström account for our respective contributions, with additional contributions from postgraduate students. Fernström's work included positioning the research within the domain of Human-Computer Interaction, a substantial part of the interaction design, implementation, evaluation and analysis.

Paper III was peer-reviewed and published by IEEE Multimedia.


This paper reports an exploration of the use of auditory display for ubiquitous computing to extend the boundaries of human–computer interaction. Our design process is based on listening tests, gathering free-text identification responses from participants. The responses and their classifications indicate how accurately sounds are identified and help us identify possible metaphors and mappings of human action and/or system status to sound. The contributions in this paper are:

---

3 www.computer.org
• The improvement of Ballas’ $H_{cu}$ measure (1993), to support selection of what sounds to choose for auditory user interface design.
• The development and evaluation of an exemplar interface - auditory soft-buttons.

This work was mainly by M. Fernström: concept development, improvement of Ballas' (1993) method for measuring causal uncertainty, the experiment design and the development of a sound model for the new soft-button widget, and analysis. Postgraduate students assisted in sorting written participant responses from listening tests, as well as integrating a sound model developed by M. Fernström with application level code for the experiments. Prof. Liam Bannon contributed with constructive critique throughout the development.

Paper IV was peer-reviewed and published by IEEE Multimedia.


Interactive sonification was one of the elements of the Shannon Portal installation in Shannon Airport (11th to 31st of July 2006). The sonification was designed to catch the users’ attention when passing by and to provide auditory feedback about the system’s response to their actions when navigating through a virtual image gallery on a large back-projected visual display, based on the users’ body movements.

The work in this paper covers a number of issues when designing ubiquitous technology for a public space. It was a major deliverable in the Science Foundation Ireland funded project Shared Worlds (Principal Investigator Prof. Liam Bannon), with substantial contributions from most of the faculty, postgraduates and staff in the Interaction Design Centre at the University of Limerick. M. Fernström’s contribution to the work was the overall technical management and engineering, and in particular for this thesis, the research, development and evaluation of an auditory display component to enhance the user’s interaction with the system.

Paper V was peer-reviewed and published by IEEE Computer.
Other related publications by the author


Acknowledgements

Thanks to all the people who helped and supported me:

Mona Fernström
Anne Murphy
John Bowers
Eoin Brazil
Chaoimhe Doyle, Ardmore Sound
Randy Thom, Skywalker Sound
Mícheál ÓhAodha, UL Librarian
Laura Beatty, RTÉ Radio
Douglas Engelbart
Stewart Brand
Liam Bannon
List of Figures

Figure 1: Two examples of earcons........................................................................................................19
Figure 2: PD sound object model for bottle filling sound.................................................................42
Figure 3: Sonogram, filling a real 1-litre bottle with water. .............................................................42
Figure 4: Sonogram, sound object model for filling 1-litre bottle................................................43
# 1 Introduction

This is a multidisciplinary investigation in the field of sonic interaction design. The focus is on designing and evaluating auditory displays\(^4\) for “direct manipulation”-style interaction in the context of ubiquitous computing environments. The research examines the process of design at a widget-level, i.e. in terms of user interface components. The thesis consists of a series of design interventions and explorations, including several preliminary empirical investigations that demonstrate interesting features of this space. The use of sound in interaction design to date has been relatively minimal, and this thesis documents the author’s research spanning several years that involved the exploration of the relatively unchartered field of sonic interaction design.

Most of today’s operating systems and software applications such as email, instant messaging, error messages, etc., have the option to provide feedback through sound, but only to notify the user that an event has occurred and those events are signified by the playback of a short sound file. Very few of the user’s direct actions with a computer have auditory feedback to support the user’s activities. As the forms and embodiments of computers change and functionality is extended to support a plethora of human activities in different contexts, the ways that we interact with computers are constantly changing. Tools, that used to be purely mechanical or electromechanical, are becoming enhanced and augmented by computational functionality and interconnected to other computational artefacts in our environment. Unfortunately, the auditory feedback from many such artefacts is often limited (Norman 2007, pp. 58-60).

The work presented in this thesis is a design exploration in interactive sonification and argues for a more extensive and richer set of interactive auditory user interfaces, where the auditory display is fully responsive to the user’s activity, with tight coupling between actions and the auditory display, using non-speech audio.

## 1.1 Beyond the sonified desktop metaphor

After Mark Weiser’s (1991) seminal article on ubiquitous computing, some of his vision has been realised, or rather, the concepts he attributed to ubiquitous

---

\(^4\) The use of non-speech audio to convey information as defined by the International Community on Auditory Display, www.icad.org.
computing are now accepted in mainstream computing. Mobile devices show an increasing computational capability and outnumber personal computers by approximately a factor of three\(^5\); different kinds of devices communicate with each other, and with desktop, laptop and tablet computers, as well as with embedded computers in our immediate environment and via the Internet. Electronic whiteboards and digital projection systems have become common and less expensive, providing high quality displays and new forms of interaction opportunities. Weiser’s vision remains, and an increasing number of researchers have proposed and explored a *new computing* that seamlessly blends in with our environment and augments our abilities (Bergman 2000; Rheingold 2003; Shneiderman 2002; Mistry *et al.* 2009). Here, to paraphrase Ben Shneiderman, the difference between *old* and *new* is a matter of perspective, where the *old* focused on what computers could do, while the *new* focuses on what humans can do with technology.

The concept of ubiquitous computing covers a whole range of possible domains, devices and applications, from ambient room-sized computational spaces to personal and wearable technology. With ubiquity, the focus is not on a single machine that can be used for many different things in one place. Instead, there are numerous and various computational artefacts that directly support various activities in different contexts and environments. One issue with the *old computing* (i.e. PC, desktop, laptop, etc) is that it mainly focused on the use of visual displays, and many of the new emerging designs (in particular in the mobile category) suffer from small visual displays, or, the visual displays are inappropriate, as the visual modality is required for other things in our immediate environment. With wearable computers and mobile devices, the exploration of auditory interfaces has also progressed (see for example Brewster, 1999; Fernström, Brazil, & Bannon, 2005b; Helle *et al.*, 2001; Ronkainen, 2001; Rocchesso *et al.*, 2008). The following chapters of this thesis present a number of topics that contribute to a novel approach for designing sonic user interfaces.

\(^5\) Retrieved 17:45, May 30, 2011 from:

http://maps.grida.no/go/graphic/number_of_personal_computers
http://maps.grida.no/go/graphic/mobile_phones_per_1000_people
Chapter 2 reviews different perspectives on sound in the world and different forms of human listening. This is followed by Chapter 3, which considers how cinematic sound design practice in the context of this thesis can contribute to the design of auditory non-speech user interfaces. Chapter 4 outlines different aspects of non-speech auditory display, in particular user interfaces with sound. Chapter 5 describes sonic interaction design and interactive sound object models as user interface components – widgets. It also reviews the concept of affordance in the context of auditory user interfaces and describes how sound object models can be designed. Finally, the chapter proposes a framework for design of auditory widgets and states the research questions that the papers in this thesis cover, including how each question was operationalised. Chapter 6 presents the main findings in the papers, followed by a discussion in Chapter 7. Finally, Chapter 8 presents conclusions followed by suggestions for future work in Chapter 9.
2 Perspectives on sound

This chapter reviews four different perspectives on sound in the human world and how technologically mediated sound has traditionally been created. The perspectives range from Gaver’s (1994; 1993a; 1993b) work on auditory interfaces and ecological acoustics, to Schaeffer’s (1966) work in electroacoustic music, to Chion’s (1994) work on cinematic sound, to work in acoustic ecology and soundscapes by Schafer (1977) and Truax (1984; 1999).

2.1 Sound in the human world

Most people can hear, and the world is full of sound. Every natural event, creature and process makes sound. Humans only control some of the sounds in the natural environment. Background sounds contributes to awareness of time of day or night, the physical structure of the immediate environment (e.g. indoor, outdoors, reverberant, dampened), approaching potential dangers, etc. Listening carefully, numerous features of the surrounding soundscape can be noticed. For example, traffic noise is louder and more complex on weekdays; church bells, clock towers and call to prayers may demarcate certain times of the day; children’s voices in neighbouring gardens hint that they are off school and that the weather is good. Also, natural features like a tree or a hedgerow become audible when wind and rain sweeps across the landscape (O’Sullivan & Fernström 2002). The richness of natural soundscapes has been investigated by for example R. Murray Schafer in the World Soundscape Project\(^6\) since the 1960s, providing methodologies and terminology for improved understanding and appreciation of the world as an acoustic ecology (Schafer 1977; Truax 1984). In addition to all natural and traditional sounds, life in a technological society is increasingly filled with electronic sounds generated by microprocessors that are programmed to create an awareness of different kinds of information and events. For example, mobile phones make sound when someone is calling or texting, if the phone needs charging, if email arrives, etc. Household appliances such as cookers and microwave ovens make sound when operated and provide notification when a process is complete. Customer operated automatic checkouts in supermarkets bleep, ping and talk with synthetic or sampled voices. Cars make sound to alert drivers that the key is left in the ignition or that lights are

\(^6\) http://www.sfu.ca/~truax/wsp.html
on, when the door of the vehicle is opened. It is a major challenge for designers to facilitate people to pick up more and useful information in such a pandemonium.

2.2 Listening

Hearing can provide rich information about place and activity in the world. Without having to think about it, hearing contributes to an immediate awareness of surrounding space. For example outdoors, aspects of the soundscape can be rapidly picked up. A blindfolded person can be led around between different places and their hearing will immediately pick up subtle cues about the structure of the environment and surrounding activities (McGrath et al. 1999). Hearing also helps shifting attention between events happening around us, including behind, above or in places in the immediate environment where the actual source is visually occluded (Van Valkenburg et al. 2004). Hearing is also active while sleeping. In addition to alarm clocks or hearing our own name, unfamiliar sounds can wake us up to attend to potential dangers, or, if a person is used to sleep in the presence of a particular noise, its absence may work as an alert (Ruby et al. 2008).

Hearing can also be used interactively, for example learning to hear if a fruit is ripe by tapping on it and listening (Barrass 1997), or, as Hermann (2008) suggested, determining the tension of a string by listening to its pitch when plucked. Sound is used in many trivial everyday activities, for example when picking up a box of matches. By shaking the box it can be heard if it is empty or if there are a few or many matches in the box. Another example is when filling an opaque container with liquid the degree of fullness can be heard.

2.3 Different forms of listening

There are ample sources in literature about the psychology of hearing and psychoacoustics (see for example, Handel 1989; Bregman 1990; Warren 1999; Neuhoff 2004), but there are also other, more conceptual issues of listening. The following section examines how four different theoretical frameworks can be aligned for further exploration.

First, a simple example, listening to a violin playing a melody by, for instance, J.S. Bach. With the normal way of listening a beautiful melody is heard winding and weaving its way up and down the scales in a baroque fugue. If familiar with Bach’s work, the theme may be recognised and the piece of music identified and named. It would be noticeable if a novice, an amateur or a virtuoso is playing. If listening to a real, live musician, the violinist can be asked to perform the same piece, again, in
different moods, e.g. happy, sad, and angry. Most people with some previous listening experience of classical Western Art Music will immediately pick up the difference and be able to classify if it is, for example, a happy or angry performance (Bresin, 2000).

It is possible to listen in different ways to the performance. From a violinmaker’s perspective, the instrument itself can be listened to, rather than the melody or the performer (Fritz et al. 2012). A violinmaker will notice if it is a high-quality instrument, or a low-quality mass-produced violin. A violinmaker may also notice if it sounds like there is a crack developing in the top plate of the violin, or that there is a problem with the rosin on the bow.

Changing mode of listening again, as an acoustician or sound designer, it can be considered if it is a harmonic sound, how much noise it contains, amplitude and spectral envelopes, frequency modulation, amplitude modulation, etc. When listening in this mode it is possible to ignore that it is a violin and just focus on the acoustic properties of the sound. This form of listening is what Pierre Schaeffer (1966) called reduced listening.

Everyday sounds can be at least as complex and expressive as the sound in the example above. For instance, a sheet of paper can be torn apart in different ways. It can be torn apart in different moods, at different speeds, with different force, at different angles, etc. Instead of tearing paper apart, other materials such as tin foil or plastic bubble-wrap can be used. Instead of tearing, the material can be crumpled, tapped, scraped, etc. The list of possibilities is almost endless, but in most cases a listener not seeing the action would be able to pick up and identify both the material involved and the actions performed (Ballas, 1993; Ballas & Mullins, 1991; Fernström, Brazil, & Bannon, 2005a; Gygi, 2001).

What makes everyday sounds difficult to analyze and describe is that there is no precise vocabulary or score representation to accurately describe the sound, apart from normal language. This problem is not unique to the auditory domain, as people face similar difficulties with other modalities when trying to describe visual shapes and colours, scents, etc. For conventional music, a rich and very specific vocabulary and notation system7 has evolved over hundreds of years. This enables a composer to instruct a performer about fine details of a performance on a traditional musical

---

7 For some extreme examples, see [http://www.informatics.indiana.edu/donbyrd/InterestingMusicNotation.html](http://www.informatics.indiana.edu/donbyrd/InterestingMusicNotation.html) (accessed 2012-02-16)
instrument, and instruments are made according to a narrow specification that have evolved over a long period of time between instrument makers and musicians. For describing everyday sounds, there are no narrow formal descriptions but normal spoken or written language can be used, which often becomes fairly long sentences even for a relatively simple sound. Another alternative is to use onomatopoeia as, for example, reported by Cano et al. (2004) and Lemaitre et al. (2011). Onomatopoeic descriptions can be important to complement normal text descriptions involving actions, agents and contexts, or, to describe sounds that cannot be described by normal words, e.g. comic or sci-fi sound effects.

In human-computer interaction (HCI) and auditory interfaces, one major contributor is Bill Gaver. From his early work\(^8\) with the Sonic Finder application (Gaver, 1989), to his two companion papers on ecological acoustics (Gaver, 1993a; 1993b) he has had a substantial influence on the area of auditory interfaces. He proposed two forms of listening. His first mode of listening is *everyday listening*; the way people normally listen to sounds in everyday life where with minor effort sources and the activities can be recognized. His second mode of listening he called *musical listening*, which is when listening to sounds in terms of their pitch, duration, timbre, tempo, etc.

Another relevant researcher, in the domain of electroacoustic music, is Pierre Schaeffer. He developed a number of key concepts in his work towards a *Musique Concrète*, including his definition of four different modes of listening (Schaeffer, 1966). The reason why *Musique Concrète* is interesting in the context of this thesis is that it is about recorded (i.e. separated from their real sources) everyday sounds that traditionally would not have been considered to be musical. Schaeffer’s exploration of this field led him to theorise about different forms of listening. By *Écouter* (to listen to) Schaffer meant causal listening, i.e. similar to Gaver’s *everyday listening* where we actively listen for the source. In Schaeffer’s definition, the sound listened to in this mode is indexical in a semiotic sense. The sound is not the source but it points to the source. With *Ouïr* (to hear), he referred to listening to sounds in their most elementary way, the sound and nothing but the sound, the raw perception. In this mode we may consider pitch, timbre, timbral flux, bandwidth, duration, etc., but in a technical sense rather than a traditional Western Art Music sense. Shaeffer’s *Entendre* (to listen to selectively, to attend to) is basically how

---

\(^8\) Initially on recognition of everyday sounds, inspired by VanDerveer (1979)
certain parts of can be listened to, e.g. different auditory streams (Bregman 1990). Finally, Comprendre (to understand) is the high-level understanding of what is heard. From this, he developed the concept of reduced listening, which is mainly an Ouïr activity. In reduced listening, a sound's primitive properties is listened to, e.g. pitch, timbre, duration, bandwidth, trajectory, location, texture and spectral flux, without any concern about the actual source. One could argue that reduced listening is a highly technical form of listening, or, a radically different form of aural aesthetic appreciation. In either case, it takes practice and learning to develop the skill of reduced listening. Pierre Schaeffer also defined the concept of a sound object, l'objet sonore, as a phenomenological object for human auditory perception.

A third relevant researcher, but in the domain of film sound, is Michel Chion (1994). He based his ideas on Schaeffer's work, but describes three different modes of listening. Causal listening, which is the similar to Gaver’s everyday listening and Schaeffer’s Écouter; Reduced listening, such as described by Schaeffer, and Semantic listening that is similar to Schaeffer’s Comprendre. Several of Chion’s cinematic concepts about sounds may be relevant for designing sounds for other domains. For example, he explores how sound can expand the narrative space to exist beyond a visible screen and how sound can make visible action more believable and immersive.

A fourth relevant perspective on sound is offered by Barry Truax (1984; 1999) whose work is based on R. Murray Schafer’s work on Soundscapes and Acoustic Ecology (Schafer 1977). What is important in the soundscape work is that it is different from and yet complements the views of Schaeffer and Chion. In soundscape studies, sound is discussed based on distal and situated properties. Among the concepts Truax defines are Sound object, Sound event and Soundscape. In Truax’s definition, a sound object is basically a phenomenological sound formation, independent of any social or musical context, but one that can be classified according to its typology and morphology. At a higher level, Truax defines Sound events involving Sound objects, building sound sequences in spatial and temporal contexts, forming Soundscapes.

All of these concepts are important for understanding of what listening is and how sound works in the human world. The number of modes of listening described here is by no means exhaustive, but can help moving beyond Gaver's perspective and having a richer vocabulary that more precisely can describe sonic experiences. In Chapter 5 and in the papers included in this thesis, it is shown how these
perspectives can contribute to sonic interaction design. For example, Papers I, II and V are partly informed by soundscape concepts, while Paper III and IV are about the design of interactive sound objects, drawing upon concepts from both *Musique Concrète* and cinematic sound.

### 2.4 Summary

This section described a number of perspectives of what listening is, from Gaver’s everyday and musical listening, towards a richer framework that can be used for understanding different forms of listening that can help inform design of user interface sounds. Users can only be expected to engage in high-level everyday or causal listening, but for designing auditory widgets designers may need to learn to listen differently and to consciously change between modes of listening. With this expanded framework, sounds are situated in an environment, forming complex soundscapes, based on Schafer and Truax’s views, while Chion’s ideas may support thinking more about the mimetic aspects of sound, to support understanding of activity beyond traditional screen-space, for example for user interfaces with small visual displays where auditory display can expand or complement a user interface metaphor.
3 Sound Design

To design auditory interfaces, a number of issues need to be addressed. First of all, where and how auditory interfaces are appropriate, taking into account the capabilities of the users, while carrying out tasks in real environments and that a surrounding soundscape might mask the sound of an auditory display, or vice versa. Where sound may enhance interaction, ways of creating and testing auditory metaphors have to be explored. Also, as suggested by Gaver (1994), in many cases parametric control is required, for example so that small events can make small sounds; the effort of an action can be dynamically represented, etc. Truax (1984) reflected that before the development of sound recording equipment in the 19th and 20th century, nobody had ever heard exactly the same sound twice. He also remarked that fixed waveforms, as used in simple synthesis algorithms, sound very unnatural, lifeless and annoying. Just as visual graphical user interface designs benefit from the skills of graphic designers, there is a need for the art of sound design for auditory interfaces (Somers 2000). Sound design, as a profession, has emerged from work in theatre, radio, cinema and television production (Chion 1994; Sonnenschein 2001).

Over the past two decades, action related sound effects have also been added to live television broadcasts, for example at major sporting events (Maasø 2002; Maasø 2006). The use of sampled sounds or sound synthesis, controlled by a sound engineer watching the live television-feed and using a MIDI-keyboard and sampler with sound effects, improves the viewers’ feeling of engagement during the broadcast. This is especially applicable when the video images are based on long lenses (telephoto), where it is impossible or difficult to capture distant sound with acceptable quality. Sound design has also become recognized as contributing to product design in general (Özcan & Van Egmond 2006).

The following sections explore these developments, including a brief historical background of sound design in film production. The motivation for including this background is that cinematic sound preceded sound on computers, providing rich multimedia experiences for audiences (users) and there are concepts and a body of work that can help inform development of improved auditory interfaces.
3.1 Foley sound

When synchronized sound entered the scene in the late 1920’s many US studios had silent films already in production. In Universal’s originally silent movie *Show Boat* (1929), Jack Donovan Foley tried adding live-acted sound effects to a few silent clips. The idea worked and after re-shooting some scenes with synchronized sound, *Show Boat* was released with a full sound track. Almost every footstep, door opening or closing, etc., had been re-created, recorded and added to the film. The technique and ideas originating from Jack Foley are still in use. In most feature films today there are a number of postproduction people in the credits, e.g. Foley Artist, Foley Mixer and/or Foley Editor. Foley Artists are the persons who make and act out the sounds using various forms of props\(^9\) and a Foley Mixer is the person recording the Foley artist(s). The Foley Editor is the person who finally puts all the Foley layers together. There is often an important distinction between sound effects and Foley, partly due to the availability of vast sound effect libraries and sound synthesis. Most often, all human related action\(^{10}\) sound is recreated as Foley sound because *synchronicity* with live action and expressivity is critical (Chion 1994), while other effects such as wind, rain, engines, bells, whistles and gun shots are often sourced from sound effect libraries (Yewdall 1999). Sometimes, the Foley and sound effects crew is replaced by Sound Designers who may work both with Foley, sound effect libraries and also synthesizing new and purpose-designed sounds electronically (Mancini et al. 1985; Sonnenschein 2001). Foley sounds and sound effects tracks are then added to dialogue tracks and music tracks and a final mix is produced.

It has also been found by, for example, Heller & Wolf (2002) that sounds recorded from Foley artists can be better identified than recordings of the corresponding real events. Heller & Wolf also tried to cartoonify the sounds, using synthesised versions of the events, emphasising perceptually salient properties of the sounds. In 70% of their tests, listeners preferred the synthesised sounds as being more realistic.

---

\(^9\) Short for *theatrical property*, any object used in a theatrical performance.

\(^{10}\) For a Foley artist, human activity can cover anything from walking, to a clinking glass, opening a door, to the sound of screeching tyres in car pursuit scenes. In many movies, several sounds are added that we normally would not consciously notice in everyday life but in a movie, for example the rapid movement of a hand in a fight scene can sometimes be represented by the sound of a slender stick of bamboo swished through the air.
3.2 Implications for sonic interaction design

Sonic interaction design is about the use of artefacts where users experience artefacts as part of a soundscape. Designers have to understand how users may experience an artefact, situated in a complex dynamic soundscape. When designing auditory widgets for ubiquitous computing, designers are also creating the sound effects for users’ activities with their computational artefacts.

Technology and media are extending contemporary soundscapes. There are now a number of generations of people that know what a light sabre sounds like\(^\text{11}\). There are also widespread beliefs (not necessarily correct) of what things should sound like due to how they have been represented in cinema and TV. Sound designers have created familiar impressions of, for example, the sound of gunshots and explosions, scientific machinery or even in the movement of animals in documentary films about nature\(^\text{12}\) (Ament 2009).

What the work of a Foley artist can contribute is to make both actions and materials believable, for example, creating the illusion that actors are actually hitting each other in a fistfight, or, that a light polystyrene prop looking like a large rock sounds like a large rock when rolled. When designing auditory widgets, which are essentially virtual, an understanding of Foley art and sound design can help inform design.

There is also some evidence that with reproduced sound (technologically mediated), sound effects made by Foley artists or physically inspired sound synthesis can be more identifiable and experienced as better, or more believable, than recordings of real sound sources and events (Heller & Wolf 2002).

---

\(^{11}\) From the movie *Star Wars*, 1977

\(^{12}\) In some of BBC’s David Attenborough nature films, sometimes no sound was recorded when a wildlife scene was filmed. Sound designers and Foley artists later made up an impression what they believed the scene may have sounded like. See for example http://www.nealromanek.com/blog/2007/12/sounds-of-nature-part-ii.html (accessed 10th of December 2010).
3.3 Summary

This chapter briefly reviewed how sound design has become an important contribution to technologically mediated experiences. The origins of Foley art and sound design as distinct disciplines emerged from the idea of making cinematic experiences richer and more engaging. The skills and knowledge from those disciplines can contribute to sonic interaction design.

The next chapter outlines different aspects of auditory display, in particular user interfaces with sound.
4 Auditory Display

This section reviews different aspects of auditory display. Using non-speech sound to represent data can be applied for example to represent a graph or an alarm. In this case data can be listened to, without the user being able to manipulate the sound. The user can only make decisions to act upon the information that may have been picked up, e.g. if we have an auditory display of stock market data, to buy or sell a number of shares (Nesbitt & Barrass 2004). Another possibility is auditory display where the user can manipulate, explore and control how data is sonified, and the auditory display is fully interactive (Hunt & Hermann 2011). Finally, the possibility of making auditory display an integral part of the user interface itself is considered (Brazil & Fernström 2011).

One of the oldest technological examples of auditory display is the Geiger counter, invented by Hans Johann Wilhelm Geiger in 1908. The original device detected the presence of alpha particles, and later developments together with Walther Müller resulted in devices that could detect different kinds of radioactive decay. One important characteristic of Geiger-Müller counters is the use of auditory display, i.e. each particle detected is converted to an electrical pulse that can be heard through headphones or loudspeakers as a click. The more radioactive particles per second, the higher the click-rate. The technique of converting data directly into sound is called audification (Dombois & Eckel 2011). Another characteristic of Geiger-Müller counters is that they are normally handheld and can be moved around by a mobile user to detect hot-spots and gradients of radiation, while allowing the user’s visual modality to be used for navigating in the environment or making other observations. Several other kinds of instruments have been developed along the same principles, for example metal detectors and cable finders.

As a field of research, Auditory Display was formulated following the first international conference in 1992 (Kramer 1994). It involves research from disciplines ranging between HCI, Computer Music, Experimental Psychology, Auditory Perception and a number of sciences. The term sonification has become the general term for using non-speech audio to display data, ranging from data-controlled sound to direct playback of data.
"...to translate relationships in data or information into sound(s) that exploit the auditory perceptual abilities of human beings such that the data relationships are comprehensible."

(Walker & Nees 2011, p.29)

In general, auditory display can help to free up the visual modality, for example when using a wearable computer or mobile device (McGookin et al. 2009; Pirhonen, Brewster, & Holguin, 2002). It is also a way to draw attention to events and to support peripheral awareness (Gaver, Smith, & O'Shea, 1991; Gutwin et al. 2011); to give users confirmation that their actions have succeeded or failed (Brewster, Wright, Dix, & Edwards, 1995); to monitor processes (Fitch & Kramer 1994; Janata et al. 2004; Vickers 2011); for notification (Gaver & Smith, 1990; McGookin & Brewster 2011), and for displaying multidimensional or multivariate data (Smith, Picket, Williams, & Kramer, 1994; Bly 1994; Grond & Berger 2011). The following sections discuss and exemplify some of the general sonification principles.

4.1 Warnings, alarms and notifications

Warnings, alarms and notifications have been used throughout human history. With industrialisation and complex man-machine systems the need to design sounds for warnings, alarms and notifications has become increasingly important for the safe operation of systems ranging from aircraft to industrial plants (Stanton & Edworthy 1999; Guillaume 2011). The sounds have to be designed in respect of the context of use. A number of psychoacoustic and cognitive factors are considered, such as ambient noise levels and masking effects (Warren, 1999, pp.60-64), perceived urgency (Burt et al. 1999), identification of the sounds (Ballas 1999) and the sounds’ relation to visual signals. Several researchers have also noted the importance of macro-temporal structuring of warnings and alarms, i.e. that such structures significantly improve recognition of sounds, e.g. (Gygi, 2001; Patterson, 1989; Patterson, Datta, Stanton, & Edworthy, 1999; Warren, McAdams, & Bigand., 1993; Warren & Verbrugge, 1984). Common to all sonifications in this category is that they are intended to grab attention to get humans to act, ranging from the ordinary doorbell or phone ring, to fire alarms and collision avoidance alarms in aircraft (Guillaume 2011).
4.2 Non-interactive sonification

With non-interactive sonifications the user listens to a sonification but cannot directly manipulate the data or the sonification (Walker & Kramer 1996; Vickers 2011). This type of sonification is typically designed and used for monitoring tasks, ranging from stock market fluctuations (Mezrich et al. 1984; Janata et al. 2004) to vital signs in patients. For example, (Fitch & Kramer 1994) showed that the operation of life support systems could be enhanced through the use of auditory display. They developed a simulator with eight continuously changing variables representing different vital signs of a patient and found that their subjects (students who had received a short training as anaesthesiologists) performed faster and with fewer errors when using auditory display compared to visual display, especially with multivariate changes. Fitch and Kramer hypothesised that the ability of the human auditory system to handle multiple streams of audio simultaneously was the reason for this finding. The most interesting consequence of Fitch and Kramer’s experiment is that with auditory display the visual modality is free to be used for other tasks related to a patient’s well-being.

4.3 Interactive sonification

The Geiger-Müller counter described earlier in this chapter may be considered a system using interactive sonification as the user explores and navigates an environment, picking up information about its radioactive properties through sound. There are several possible and valid definitions of interactive sonification, for example where a user interactively controls the sonification either directly or indirectly, or, where the user can either navigate through a data-space to try to find items, locations or relations of interest, or, try to change the mapping between data relations and acoustic relations in real-time to find contrasts and comparisons. Interactive sonification is typically used for data exploration. Examples of interactive sonification range from interpreting geotechnical-prospecting data (Barrass 1997) to sonification of multi-channel EEGs for diagnosis of epilepsy (Baier & Hermann, 2004). All interactive sonification applications are characterized by allowing the user control of the sonification through tight coupling between the user’s actions and auditory feedback from the sonification system. Currently, there are two main approaches in interactive sonification: Parameter-mapped sonification and Model-based sonification. In a parameter-mapped sonification, the data ‘plays’ the sonification (as an instrument or set of instruments or orchestra) and the user can
modify the mappings to find items of interest and navigate through the data (almost like how a DJ/turntablist uses scratching). In model-based sonification, the data is used to create the ‘instrument’ and the user interacts with the virtual instrument, i.e. the user’s actions work as excitation of the model, to find or explore information (Hermann 2008).

Parameter-based sonification is often used when a data set to be sonified has some intrinsic order (Grond & Berger 2011). Model-based sonification may offer advantages when a data set lacks any natural order or dimensionality and mappings may be assigned arbitrarily to allow for exploration (Hermann 2011).

4.4 Sonification in games and entertainment

Sonification for entertainment is slightly outside the scientific and technical domain of auditory display, but as a source of inspiration and experimentation it deserves to be mentioned. Computer games have used sound, ever since the first Pong (1972), Space Invaders (1978), etc. Most of the early games used very simple forms of sound synthesis. Currently, most computer games in general use sampled sound, i.e. playback of sound files. It is well-known in the computer games industry that sound can increase the user’s feeling of engagement and immersion in a game (Viaud-Delmon et al. 2006; Hendrix & Barfield 1996). The worldwide web has become increasingly filled with multimedia and virtual environments such as Second Life and other MMORPGs. Most virtual gaming environments provide sound support. For example, the game World of Warcraft offers extensive sound designs, including support for spatialized sound, that help to increase the feeling of immersion and engagement in a virtual environment. While the sonifications are interactive, they still mainly rely on sampled sounds and use multiple and concurrent layers of sound files that play depending on the state of the game and the users’ actions in the virtual world, i.e. the selection and mix of sounds changes dynamically, but not the sounds per se.

Recently RjDj started to explore the possibilities using model-based sound synthesis and interactive audio effects for entertainment purposes. The computer games company Electronic Arts is also exploring such possibilities (Farnell 2008).

---

13 Massively-Multiplayer Online RolePlaying Game
14 http://www.worldofwarcraft.com/
15 http://www.rjdj.me/ (from Reality Jockey Ltd)
16 http://www.ea.com/
4.5 User interfaces made of sound

Sonifications that aim to improve the usability and quality of the user experience of computer applications may draw on aspects from all the aforementioned types of sonification. There are many different examples of auditory user interfaces (see for example Gaver (1997) for a general overview), but this particular kind of sonification is not about exploring any arbitrary underlying data set, but about the direct experience of using computer applications, e.g. the metaphorical model-worlds of user interfaces. Historically, components of auditory user interfaces are for example *earcons* (Blattner, 1996; Blattner, Sumikawa, & Greenberg, 1989; Brewster, Wright, & Edwards, 1994) and *auditory icons* (Buxton, Gaver, & Bly, 1989; Gaver, 1989; 1994). These components, or *widgets*, are building blocks for interface design, i.e. virtual or metaphorical objects that users can perceive. *Earcons* are short musical motifs (sequences) and the meaning of these motifs has to be learnt. They can be organized into families of musical messages to represent, for example, hierarchical information. To date, most earcon designs build on the principles of Western Art Music, i.e. 12 notes per octave and counterpoint, and can therefore also build on features known from musicology, such as polyphony, harmony, timbre, rhythm and melody. Earcons have for example been used to improve the usability of pull-down menus, soft-buttons, scroll-bars (i.e. most standard GUI-widgets) but also for providing hierarchical information to users of Computer-Telephony Integrated systems, e.g. voice-mail systems (Brewster, 1997; Brewster, Wright, & Edwards, 1994). Figure 1 below illustrates two simple earcons. For example, the first ascending major triad can be associated with ‘opening’, while the second descending major triad can be associated with ‘closing’. Timbre can also be used for adding properties to earcons, for example opening a text file could be rendered as a piano sound, while opening a spreadsheet could be a flute sound. Furthermore, polyphony can be used (more than one note at the same time) to create hierarchical earcons.

![Earcons](image)

Figure 1: Two examples of earcons

*Auditory icons* use a different approach as they mimic everyday sounds that people may be familiar with from their everyday experience of the real world, hence the meaning of the sounds seldom has to be learnt as they metaphorically draw upon
previous experiences. For example, deleting a document may be represented by the sound of crumbling paper; an application terminating in error may be represented by the sound of breaking glass. William Gaver (1989) pioneered the field of application of auditory icons with his *SonicFinder*. It extended Apple's file management application *Finder* (an integral part of Apple’s desktop metaphor) using **auditory icons** with some limited parametric control. The strength of the *SonicFinder* was that it reinforced the desktop metaphor, enhancing the illusion that the components of the system were tangible objects that could be directly manipulated.

With most operating systems for personal desktop computers today, users can activate sets of desktop sounds. These sets are often hybrids of **earcons** and **auditory icons** (information and events that do not have a real-world analogy represented by earcons, while auditory icons used for information that can be regarded as metaphorically analogous to real-world events). A fundamental difference is that earcons are considered to be arbitrary symbolic representations, while auditory icons are regarded as analogical representations (Kramer 1994).

In the European project *Sounding Object* (Rocchesso & Fontana, 2003) a novel approach for creating sound for interaction design was investigated. The project was inspired by Gaver’s ideas (1993a; 1993b) and aimed at developing physics-based sound models and control models for potential integration into human-computer interaction, in particular for ubiquitous computing. Sound objects were synthetic auditory icons based on real-time mathematical simulations of real-world physical events, for example a soft mallet hitting a glass, or, a marble rolling on a table. The simulations were executed in real-time and therefore properties of the simulated events could be parametrically changed, for example, the speed of the marble rolling or the size of the glass being hit. For another way to understand the difference in the Sound Object approach, a useful categorization can be found in an unfinished book manuscript by Buxton, Gaver, & Bly (1994). They distinguished between **fully formed objects** and **evolutionary objects**. With the former category, all variables are known at instantiation, hence when an object is activated a sound is produced, from beginning to end, e.g. like playing back a recording. With the latter category, variables controlling properties of the sound are updated while the sound is playing.

---

To further clarify the difference between these two concepts, consider an analogy from the musical world: Striking a key on a piano, after there is no more control over the sound, compared to playing a note on a violin where there is full expressive control over pitch, loudness, timbre, vibrato, etc., throughout the duration of the note.

For interaction designers, the possibility of having widget libraries with sound objects implies that the auditory display domain can be as live and animated in direct manipulation interfaces as visual interfaces have become. It has also been shown that a sound object approach can work well in, for example, ubiquitous and wearable applications (Fernström, Brazil, & Bannon, 2005b; Rocchesso, Bresin, & Fernström, 2003).

### 4.6 Summary

This chapter briefly described the different sub-domains and aspects of auditory display. Applications using auditory display range from relatively passive forms where the listener can pick up information in the sound to make decisions, monitor processes, etc., while not having any direct control of the sound. At the other extreme there are user interfaces built from sound, where the user interface components can be directly responsive to a user’s actions, which in turn helps to reinforce the user’s experience with the auditory display in terms of believability.18

The next chapter reviews what constitutes sonic interaction design and auditory user interfaces, in particular its components or widgets, and reflect upon how an auditory affordance can be defined. This is followed by an outline of the design process used in the papers presented in this thesis. Finally, a framework for sonic interaction design of auditory widgets is proposed and the research questions stated.

---

18 The term believability is used here to describe that the user finds the audio from a sound object model plausible as being caused by the current activity and actions with a computational artefact.
5 Sonic Interaction Design

Design is a human activity - a process - for specifying artefacts to be used by humans in specific environments and contexts to accomplish their goals. Design can be expressed in many different ways, as the use of a design approach can range from problem solving or problem finding, to iterations of gradual and directed improvement of an artefact, as well as pure inspiration (‘What if…’) resulting in new artefacts where the potential use and goals are discovered by users over time (Ralph & Wand 2009).

Interaction design is a complex and multidisciplinary field of research, which draws upon other fields such as human-computer interaction, computer science, psychology and ergonomics, as well as traditional design disciplines, such as graphic design, industrial/product design and architecture. As interaction design can be seen to have emerged from human-computer interaction, and technology is continuing to be developed, there is a constant need to re-imagine what the actual field of research is, as discussed by for example Bannon (2011). With an increasing amount of user participation in the design process, interaction designers can sometimes be seen as mediators or facilitators to support user groups to articulate designs that can provide a relevant user experience (Howard & Melles 2011).

In the field of interaction design, Daniel Fallman has discussed the differences between design-oriented research, with the purpose of acquiring new knowledge, and research-oriented design, with the purpose of producing new artefacts (Fallman 2003; 2007; 2008). Interaction design is mainly focused on human behaviour and the use of artefacts, which makes it different to industrial/product design, which is focused on form. While the goals of design-oriented research and research-oriented design differ, the former may also produce artefacts (as a means rather than a goal) – prototypes – to explore new possibilities for humans to interact with artefacts. Such prototypes can become demonstrations that help communicate new interaction design concepts between interaction design researchers and interaction design practitioners (Zimmerman et al. 2007).

In interaction design, as in most design disciplines, there are a number of characteristic activities:

- Developing multiple perspectives on a problem: In addition to direct and personal experience, for example, interviews, observations, surveys and
ethnographies can support the development of alternative views of a problem. The different perspectives can be used for informing designers about the nature and context of a problem, or, as ‘lenses’ or ‘filters’ to support critique and analysis of different solutions.

- Synthesise concepts and generate ideas towards different solutions. Ideas are often expressed as sketches for externalising and sharing ideas.
- It is an iterative process that includes continuous reflection and evaluation, refining concepts through a series of higher fidelity prototypes.

Sonic interaction design is focused on the exploration and investigation of how auditory display can be used in interaction design, where sound is the main medium for conveying information and aesthetic/emotional experience in relation to interactive systems. The field of sonic interaction design emerged from a number of research projects and collaborations, for example, the Sounding Object project 2001-2002 (Rocchesso & Fontana, 2003), the EU COST Action 287 Gesture Controlled Audio Systems 2003-2007 (Godoy & Leman 2009), the CLOSED project19 2006-2009 and the EU COST Action IC0601 Sonic Interaction Design (Thomas Hermann et al. 2011). Sonic interaction design can be regarded as being an intersection between interaction design, sound and music computing and sound design.

5.1 Making user interface widgets

In human-computer interaction, a widget is traditionally defined as a component of a graphical user interface that displays an information arrangement controlled by a user (Smith & Alexander, 1988). Widgets are designed and displayed so that users can perceive them and understand how they can be manipulated as objects – more or less metaphorically analogous to objects in the real world. For example, a graphical user interface may have a number of rectangles in different shades or colours to the background. Each rectangle may have a text or icon symbolising its function, when activated. Each rectangle may have a different shade along its upper/left and lower/right outline, as if there was a light-source illuminating the display surface from somewhere above the rectangles, hence giving an illusion that they are almost three-dimensional and standing out from the background surface. If one of the rectangles is manipulated by pointing and clicking with a mouse and cursor, or with

19 http://closed.ircam.fr/ (accessed 2012-02-07)
a finger on a touch screen, the illusion of three-dimensionality will be reinforced by
the shade of the outline being reversed while the widget is activated, looking like it
has been pressed into the background surface. The visual appearance and spatial
structure of such rectangles can be picked up by visual perception. From previous
experiences with the world it may suggest the rectangles are buttons affording the
action potential of pressing or touching them in some way.

One possible way to understand how information can be picked up from our
environment and from actions performed in that environment is the concept of
affordances. The concept was originally introduced by J.J. Gibson (1979) as a
relation between an organism and its environment that potentially allows the
organism to carry out actions. Gibson’s work was firmly situated in theories of
action perception, in particular visual perception. Don Norman (1988) adapted the
term in the context of human-machine interactions and Norman’s book The
Psychology of Everyday Things popularised the concept. Norman’s use of the term
was problematic, as it introduced the notion of ‘perceived affordances’, which
differs from its original use by Gibson. The resulting controversy led to a later
clarification by Norman (1999).

5.2 Auditory affordances
The work initiated by VanDerveer (1979), inspired by Gibson’s work on visual
perception, formulated an ecological approach to acoustics. This was continued and
improved by Gaver (1988), who also applied the emerging ideas to human-computer
interaction, including the development of the SonicFinder application for the Apple
Macintosh, which added sound events to accompany user actions on the Apple
desktop interface. Gaver (1991) tried to bring some further clarity to the use of the
affordance concept in the context of human-computer interaction, discussing the
concept from a design perspective. He also noted that sound in a user interface can
afford users information about how successful an action is, and support
collaboration between users and supplement visual information, in principle
extending Norman’s definition to the domain of auditory user interface design.

Stanton and Edworthy (1998) defined auditory affordances in the context of auditory
warnings. While they referred to Gibson (1979), they applied the term in a similar
way to Norman (1988), discussing the concept as something that is perceived but
also learnt in a social and cultural context. Several contributors to John Neuhoff’s
(2004) book on Ecological Psychoacoustics refer to the affordance concept, but do
not bring any further clarity to Gaver’s (1988) definition. As outlined above, several researchers have worked towards an ecological understanding of auditory perception, each in their own domain and with varying success. For the work presented in this thesis, the following definition of auditory affordance applies:

“The fact that sounds are available to the ear implies that information about the sound-producing events is also present. This information is available in the form of higher-level relations among the physical parameters of the sounds that correspond to attributes of their sources. Of particular importance is that these relations remain invariant over other transformations of the sounds if the corresponding source attribute is also unchanging, and change if the source attribute changes. The perceptual system "picks up" this information, actively seeking it and attuning itself to its presence. In particular, the perceptual system is sensitive to "affordances", information specifying the functional relations between the source and the listener. These affordances are partly responsible for the significance of the sounds.”

(Gaver, 1988, p.20)

For further understanding of how affordances can work for auditory interfaces, consider how invariants are detected:

“The most appealing aspect of the [ecological] approach, however, concerns the endeavor to develop descriptions of the structure of the physical world that make evident the properties that are perceived as being invariant, even though other properties may be changing.”

(McAdams, 1993, p.147)

All objects resonate if excited by an external force, e.g. hitting, scraping or rolling. The resonances in objects are modal, i.e. the acoustic energy propagating through the object is reflected between the edges and surfaces of the object. Complex and compound objects have more modes (Avanzini et al, 2003). As McAdams (1993) noted, the set of modal resonances from an object can in terms of perception be thought of as structural invariants when we listen to the sound of the object being handled. Modal resonances of an object can be regarded as mainly micro-temporal,
i.e. in the range from 50µs to 50ms (assuming the human hearing range is 20Hz to 20 kHz). When, for example, dropping an empty glass bottle on a hard floor, it may bounce and rotate a few times, or break. If it does not break and rotates between impacts, different parts of the bottle will come into contact with the floor at each bounce event, causing some of the modes to be more excited, and other modes dampened (Warren & Verbrugge, 1984). This could be thought of as **transformational invariants**. If the bottle breaks, it scatters into many different smaller pieces with different modal resonances. As transformational invariants are dependent on the handling of an object, they can then be regarded as mainly macro-temporal, e.g. in the range of from 50 ms to several seconds.

It has been shown that it is possible to reduce the number of modes when modelling a sound object while retaining an acceptable degree of identifiability, e.g. (Gygi, 2001; Warren & Verbrugge, 1984), as long as the macro-temporal patterns remain intact. This implies that **cartoonification** of sound object models (Fernström, 2003; Gaver, 1993b) can make them computationally more efficient and easier to distinguish from ‘real’ sounds (in the context of using auditory icons that otherwise might cause confusion). Alternatively, we may think of this as **caricaturisation** as we can emphasise some parts of a model and ignore other parts, to improve recognition, making our sound object models different to real-world sounds – but still building on the users’ everyday experience of sounding objects (Heller & Wolf, 2002). If we were to make a musical analogy, when we play a violin by bowing or plucking, there will be structural resonances in the strings that are transferred via the bridge to the body of the violin that in turn will have structural resonances. The violin and what excites the strings forms a sound object. A listener may hear that it is a violin that is played, picking up the structural invariants of the sound. The listener will also distinguish if it is bowed or plucked, picking up the transformational invariants of the sound. The listener will pick up all this information in the context of the ambient acoustics produced, hearing the event as a soundscape, and will also be able to estimate the distance to the violin and the size and reflective properties of the surfaces of the room. These distinctions may also help in understanding why, for example, with a low-fidelity reproduction of sound (e.g. a small radio with a 25 mm loudspeaker), different sound sources in a complex mix can be identified.

---

20 It often takes slightly longer to pick up a structural invariant, as you need to hear at least a few periods of each mode for the auditory system to detect the signal properly.
5.3 Designing sound object models

Ever since Pythagoras anecdotally observed that hitting a large anvil created a sound with lower pitch than when hitting a small anvil, people have attempted to understand the vibrations of objects in the world (Riedweg & Rendall 2005). With significant contributions in what could be called “classical” acoustics ranging from Galileo, Newton and later for example Helmholtz and Rayleigh, today’s engineering methods in Finite Element Modelling have allowed us to create high-fidelity models of many kinds of physical and dynamic systems (Strang & Fix 1973). While the 20th century saw a number of sound synthesis principles emerge and evolve, mainly based on a signal-based approach, Julius Smith’s (1992) paper on digital waveguides inspired many researchers towards a physical modelling approach. At the same time Xavier Serra’s work (1989; 1997) on spectral methods with analysis/resynthesis of sinusoids plus noise, offered alternative possibilities to model complex sounds. In the Sounding Object project both approaches were used, while mainly focusing on physical modelling methods, but also using hybrid methods similar to Perry Cook’s PhISM21 (1997). See, for example, Rocchesso & Fontana (2003) and Rocchesso, Bresin, & Fernström (2003) for details.

A sound object model can be implemented by either simulating the modal resonances of a physical object being excited by an external force, or, simulating a signal with homomorphic modulation similar to that which is emitted from a physical object being excited by an external force. Such simulations are mainly micro-temporal and depend on the material structure of a simulated object. Different patterns of excitations, e.g. bouncing, can be modelled on how an object is handled and are mainly macro-temporal.

In the Sounding Object project, the sound object models had both a micro-temporal level for the sound objects and control models for the macro-temporal level of interaction (Bresin et al., 2003; Fernström, 2003). Similar approaches can be found in many contemporary sound synthesis applications where the audio simulation rate (sampling rate) is for example 44,100 Hz, while the control rate is one or two orders of magnitude lower. From this it follows that structural invariants as simulated by sound models represent material structure, while transformational invariants as simulated by control models represent the manipulation of the sound object models.

---

21 Physically Inspired Sonic Modelling
5.4 Designing interactive sound object models

In most creative design activities, the process includes brainstorming and concept generation. In this part of the design process, samples have to be available that can be discussed and critiqued, to help drive the creative process forward. In other design disciplines, mood boards are often used to gather ideas for stimulation and as exemplars. For sonic interaction design, one of the tools developed, the Sonic Browser, can be used for gathering and listening to groups of sampled sounds or individual sampled sounds (Fernström, 2005; Fernström & Bannon, 1997; Fernström & McNamara, 2005a). With a tool like the Sonic Browser, visual representations (symbols) of sound files in a starfield visualisation can be dragged around the screen and grouped into compound clusters of sounds that play simultaneously when an aura-cursor is over a group. This enables a designer or design team to rapidly select and try out how different combinations of multiple sounds work together. Alternatively, any modern MP3-player can be used to hold examples with collections of sounds, with user-defined playlist functions to create mood boards. The important issue is to have plenty of relevant sound examples at hand to support a creative process. It is also valuable to have sound collections that have been tested in terms of causal uncertainty, including information about what users participating in testing the sounds in isolation believed they heard (Fernström, Brazil, & Bannon, 2005b). The short sentences written by participants can support finding and building possible metaphors for interaction. If a suitable sound is not available in collections, there is also the possibility to try to make the sound live, and perhaps record it as a placeholder for a concept.

When a number of suitable candidate sounds have been found, sound files can be edited, processed and mixed and, for example, added to a video prototype of the system under development. Another possibility is to just act out the imagined use of a system under design, while making the sounds live, i.e. bodystorming (Oulasvirta, Kurvinen, & Kankainen, 2003). Potential users can participate directly or watch the video prototype and comment.

Sounds and interactions can then be chosen from the previous stage and rapid prototypes created, with tools such as, for example, PD22, Max/MSP23, or Supercollider24 and/or in combination with easy to use multimedia packages, for

---

22 http://puredata.info/
23 http://cycling74.com/
24 http://www.audiosynth.com/
example Adobe Flash\textsuperscript{25}, Adobe Flex\textsuperscript{26} or Processing\textsuperscript{27}. This approach has an overhead in terms of having both a sound synthesis engine and a multimedia player running together, and there will inevitably be some latency issues. In the Sounding Object project, PD was used as the common platform for sound object model designs, which was also effective to facilitate the next stage of the process. As PD is Open Source software, the final sound designs can be extracted and rebuilt by taking the relevant source code snippets from PD, and integrating it with other source code for the product or artefact under development.

An attempt to integrate this design process was proposed by the CLOSED project, creating a toolkit for sonic interaction design (Devallez, Fontana, & Rocchesso, 2008; Drioli et al., 2009; Stefano Delle Monache, Polotti, & Rocchesso, 2010; Polotti, Delle Monache, Papetti, & Rocchesso, 2008; Rocchesso & Serafin, 2009; Rocchesso et al., 2008).

\textsuperscript{25} http://www.adobe.com/products/flash/
\textsuperscript{26} http://www.adobe.com/products/flex/
\textsuperscript{27} http://processing.org/
5.5 Framework for design of auditory widgets

Based on the literature review in the previous chapters and the empirical work presented in the papers in this thesis, a framework for designing auditory widgets for direct manipulation style interaction is proposed.

Exploring and finding sounds

Designers have to be able to easily and effectively explore collections of sounds and find potentially useful sounds to facilitate the sound object model design process. Sounds that are considered to be useful should be possible to organise in subsets within collections that can inform the design and implementation of sound object models.

Auditory affordances and different forms of listening

To create auditory widgets from sound object models, it is necessary to have an understanding of auditory affordances in the context of user interface design and the fundamental issues involved in different forms of listening and modelling of acoustic phenomena.

Believability of sound object models

The believability of a sound object model depends on both micro and macro-temporal aspects of sound object model design. Just as a rectangle on a screen in a graphical user interface looks more “pushable” if rendered in pseudo-3D by introducing virtual shadows that change when manipulated, so do auditory widgets sound more believable if the dynamics of their macro-temporal control models correspond to what we are familiar with in the natural world. Sound object models have to respond to user actions in a plausible way.

Interactive auditory display in context

Auditory widgets have to be designed for use in real contexts. A user interface including audio does not exist in isolation. For example, a visual user interface is dependent on ambient light levels, and an auditory user interface is dependent on the ambient noise levels. The context and environment that an auditory user interface is designed for can be explored and analysed through soundscape studies and acoustic ecology.
5.6 Research questions

In relation to the proposed framework in the previous section, the five papers included in this thesis address the research questions below.

**Paper I and II**
How can an interactive auditory display be designed and used for augmenting the task of finding fuzzy or ill-defined targets in a collection of sounds?

**Paper III**
How can auditory widgets be constructed from simulations of sound objects? How can such sound object models be constructed and parametrically controlled in real-time to mimic both structural and transformational invariants, including dynamics, and be controlled by human gesture?

**Paper IV**
How can sound identification accuracy be measured to help inform design of auditory widgets? How can an interactive non-visual interface, based on auditory widgets from sound object models, be designed, implemented and evaluated?

**Paper V**
How can interactive auditory display be used for supporting interaction in ubiquitous computing in public environments?

5.7 Operationalisation

The research questions were operationalised as follows:

**Paper I and II**
A Sonic Browser application was developed with a participatory design approach. An experiment/probe, with 10 users was conducted. With multiple-stream audio, target sounds were found faster than with single-stream audio. Time to Completion was measured in 13 randomly ordered audio browsing tasks per user.

**Paper III**
Sound object models were developed and implemented in PD. Several experiments/probes were conducted, ranging from investigation of perception of everyday sounds, to the design and evaluation of novel auditory interaction styles based on sound object models.
Paper IV

An experiment was designed and conducted to investigate how users identify recordings of everyday sounds, with 14 participants listening to 104 different sounds in random order, describing each sound with a short written sentence. To analyse the results, James Ballas’ method for calculating causal uncertainty was used (Ballas et al. 1986; Ballas & Howard 1987; Ballas & Mullins 1991; Ballas 1993).

A second experiment was designed and conducted to investigate a novel pseudo-haptic interaction style using audio to represent soft-button widgets based on a sound object model of friction, in a non-visual user interface. 10 participants carried out a button layout identification task in two conditions, haptic and pseudo-haptic/auditory, with six different layouts presented in random order.

Paper V

As part of the design of an interactive system for a public space, an interactive sonification was developed with a participatory design approach and evaluated through video recording, interviews and observation of users.

5.8 Summary

The work presented in this thesis takes a design-oriented research approach. The goal of the work was to uncover new knowledge about how sound could be designed and used for augmenting human-computer interaction. The previous chapters presented different perspectives on sound in the human world and reflected upon different forms of listening. This was followed by a description of the evolution of sound design practices. Different aspects of auditory display were introduced, in particular, how user interfaces can be made with sound. The emergence of sonic interaction design was reviewed and a framework for design of auditory widgets proposed. Finally, the research questions were stated in relation to the papers included in this thesis. In the next chapter the contributions of the present work are discussed. Prototypes were developed and explored as “proof of concept”, demonstrating novel possibilities for humans to interact with artefacts.
6 Contributions of the present work

This chapter presents the contributions of the five papers included in this thesis. For examples on how this work has contributed to the field and used by other researchers, see Chapter 7.

6.1 Paper I & II: After Direct Manipulation – Direct Sonification; Reflections on Sonic Browsing

When exploring multimedia browsing, a set of real problems and initial requirements were gathered from potential users in a neighbouring research centre, the Irish World Music Centre (IWMC). Their faculty and students in ethnomusicology were using handwritten or printed traditional musical score or text based methods to annotate collected music material, and to search in those written materials for contrasts and comparisons. Working with visual symbolic representations of music in such a context is a highly specialised and demanding task, sight-reading and audiating the musical material (Gordon 2001). Some musicians and musicologists in IWMC suggested that they would like search for music by acoustic query-by-example (to play, whistle or hum a section of a tune), or, by quickly listening to one or two bars of each tune (instead of reading). It was also discussed with the user group that some musicians keep notebooks with one or two bars of the main theme of each tune in their repertoire, as an aide-memoire, for example enabling them to quickly respond to the audience requests.

The query by example alternative is problematic for a number of reasons. A user playing, whistling or singing would normally not have the same pitch accuracy as a computer system’s internal score-based or digitized audio-based representation. Furthermore, queries or samples would have to be normalized and quantized regarding key, note durations and tempo to be comparable. Finally, and perhaps the most difficult problem in relation to the chosen user group in IWMC, traditional tunes tend to exist in many, slightly different, versions with various levels of ornamentation added to each version of a tune, depending on who collected it in the first place and who is trying to find it and their respective views on the subject matter, the music and the culture in which it exists (Ó Maidín 1995; Ó Maidín 1998; Ó Maidin & Fernström 2000).

Following discussions with the users, an augmented multimedia browsing approach was explored, with some other problems and constraints. Another perspective on
this problem is based on listening experiences with multitrack tape, as described by Walter Murch in (Chion 1994, pp XIII-XV), i.e. that many layers of sound could be put together concurrently and listened to selectively at will, either as music, sound or acousmatic music.

The main technical challenge was that the operating systems of personal computers at the time would normally not support concurrent playback of multiple audio files. Other challenges were, for example, to find spatial mappings that were meaningful to users. In discussions with the IWMC user group, it was found that they would think of the organisation of a corpus of musical material in terms of type of tune, tonality, date, key, time signature, tempo, and composer. These ideas can also be seen reflected in physical paper-based versions of musical corpora, for example Aloys Fleishmann’s Sources of Traditional Irish Music (1997), where indices are structured in this manner.

Inspired by the work of Ben Shneiderman and his collaborators (for example, Shneiderman 1993; Ahlberg & Shneiderman 1994b; Jog & Shneiderman 1994; Ahlberg & Shneiderman 1994a), a first prototype was developed of a Sonic Browser. Later versions with further improvements were developed with support from undergraduate and postgraduate students in Computer Science. Early in this development and based on for example Benford and Greenhalgh’s ideas of aura and nimbus in virtual settings in CSCW28 (Benford & Greenhalgh 1997), a full 3-D audio spatialization was explored, where the user’s presence in the desktop screen-space was the cursor with an aura around it and the symbols around the cursor were tunes. Unfortunately personal computers were not powerful enough at that time to spatialize more than about three audio files using HRTFs29, but the exploration also hinted that there might be a problem with the fact that the up/down space on a screen did not feel natural in relation to the front/behind space of 3-D audio. What did work though was mapping of left-right panning; hence the development work was refocused on a simple stereo spatialization with the loudness of tune symbols being inversely proportional to the distance to the cursor. In our user testing, this proved to be an effective approach and in a comparison between using single stream audio (one tune after another) and stereo panning with multiple stream sound (several tunes at the same time; all tune symbols within the user’s aura playing).

28 Computer-supported cooperative work
29 Head-related transfer function
With the stereo-spatialized multiple-stream condition, user’s located target tunes, on average, 27% faster than users with single-stream audio (Fernström & Bannon, 1997; Fernström & McNamara, 1998; 2005b). A plausible explanation of why this spatial mapping worked well is probably the tight coupling in the user interface, in principle creating a similar experience as the so called *audivisual contract*, i.e. sounds get *magnetized* to the location of visual action (Chion 1994). Other researchers have adopted this approach while adding more detail to the visual appearance of the Sonic Browser (Brazil, Fernström, Tzanetakis, & Cook, 2002; Tzanetakis & Cook, 2001).

Since its inception, the Sonic Browser has also been modified, used, explored and evaluated in other domains needing interactive sonification. Moving on from traditional music to everyday sounds, the Sonic Browser was re-developed for two other domains in the *Sounding Object* project. Firstly, it was used as a content management system for the growing collection of sound samples and sound models. Secondly, it was used for exploring perception experiment designs where participants visually and spatially sorted sound models depending on their experience with each sound in relation to other sounds.

For content management, three different visual displays were implemented, that the user could switch between at any time - *Starfield*, *Hyperbolic Tree* and *TouchGraph*. The functionality of the auditory display was the same as before. Symbols within the user’s aura (around the cursor) would play concurrently, panned out in stereo with loudness inversely proportional to the distance from the cursor (Brazil & Fernström 2003).

In the perception experiments with sounds object models, users were presented with a starfield display where approximately half of the symbols were mapped to recordings of real sounds and the other half of the symbols were mapped to sound object models of the same kind of objects and events. Initially the symbols were evenly and randomly distributed in the screen space. For example for impact sounds, participants were asked to spatially sort the symbols on screen according to height of drop and size of objects. It was found that the Sonic Browser was effective for quick explorations of how the multimodal parameter space of sound object models could map to physical measurements, and also provide hints about the fidelity of our

---

30 This was measured as the cumulative time per user, the results subjected to a Mann-Whitney test, *p*<.05. The cumulative time for single-stream audio ranged from 130-191s, while for multiple stream audio 90-121s.
sound object models (Brazil, Fernström, & Ottaviani, 2003a; 2003b; Ottaviani & Fernström, 2003). The possibility of using the Sonic Browser concept in personal music players such as the Apple iPod was also informally explored. The explorations were implemented on desktop personal computers with iPod mock-ups on screen (Fernström 2005).

6.1.1 Summary
The Papers I and II produced an answer to the question as to how an interactive auditory display can be designed for augmenting the task of finding ill-defined targets in collections of sound through the development of an interactive multimedia browser - the Sonic Browser. Users that participated in the evaluation of the Sonic Browser completed their browsing tasks significantly faster when provided with multiple stream audio. It was also found that it was the tight coupling between the user’s action and the dynamic sonification that contributed to the user’s ability to navigate between different sounds or within a complex soundscape.
6.2 Paper III: Sounding Objects

The Sounding Object project (SOb) was a European research effort\(^{31}\) with collaborators in Italy, Sweden and Limerick. The main objective was to research and develop a novel approach for the creation and implementation of dynamic and expressive auditory displays and auditory widgets in ubiquitous computing applications. Before this project, almost all non-speech sounds used in user interfaces were either recorded sounds or simple waveforms. There were a few exceptions in the area of electronic musical instruments and electronic music systems, but those were only applicable for musical contexts. The project was inspired by Gaver’s work (1993a;1993b) and aimed at finding control parameters for physical model sound synthesis that were perceptually salient in terms of the identity and expressivity of the resulting sound object models.

Physics-based models are common in both science and engineering where a system’s dynamic behaviour is simulated through finite-element modelling, i.e. objects are analysed by subdividing them down into simple sub-systems that can be described by differential equations. Although it is possible to create high-fidelity simulations through this approach, solving a large matrix of differential equations numerically does not lend itself to real-time applications, particularly not on ubiquitous and wearable computing platforms with limited computational resources. The SOb project strived to simplify models, with the aim of only preserving perceptually salient aspects of the sound object models.

6.2.1 Sound object models

To develop a sound object model, initially Finite Element models of physical objects were created in \textit{Matlab}\(^{32}\) or \textit{Octave}\(^{33}\). While such models could not run in real-time, they could render audio files at non-real-time rate that could then be played back and evaluated in listening tests in terms of realism, fidelity, scaling, etc. Models were then reduced, while retaining perceptually salient aspects of the sound object, and finally implemented in \textit{PD}, or through extensions of \textit{PD} written in \textit{C}, resulting in

\(^{31}\) The project was part of the EU – Future and Emerging Technologies programme the \textit{Disappearing Computer}

\(^{32}\) http://www.mathworks.com/products/matlab/

\(^{33}\) http://www.gnu.org/software/octave/
sound object models that would run in real-time and that could be evaluated as part of auditory user interfaces. The reductions of the models also explored *cartoonification* of the sound, as previously described in Section 5.2. The main sound object models developed by the project were for bouncing, breaking, rolling, friction, crumpling and pouring (Rocchesso & Fontana, 2003).

A new version of the *Sonic Browser* application (described in Paper I and II) was developed and used by the project collaborators for some perception experiments and for audio content management, as part of the sound object model design process, as described in Section 5.3 (Brazil, Fernström, & Ottaviani, 2003a; 2003b; Fernström & Brazil, 2001; Ottaviani & Fernström, 2003).

A number of control models were explored, e.g. walking, running and hitting. In particular a virtual *Bodhran*\(^{34}\), a *Vodhran*, was developed. The motivation for this exploration was that the traditional musical instrument can be quite expressive as the player can both apply different tension to the drumhead (traditionally made of goat-skin) and dampen some of the modal resonances depending on the placement of the point of contact between one hand and the drumhead. The drum is traditionally struck with a stick made of wood and the player can excite different modal patterns depending on how and where the stick strikes the drumhead. As drumming requires low latency, this exploration was a critical part of evaluating the real-time performance of the sound object models (Bresin et al., 2003). Throughout the exploration of the Vodhran, three virtuoso Bodhran players in Ireland collaborated with the project team. The collaboration included early discussions about the differences in their playing styles, to measuring and analysing their playing gestures using a Polehemus Fastrak electromagnetic motion tracking system\(^{35}\). The final Vodhran design was tested both in a laboratory environment by one of the previously mentioned players, as well as in a live ensemble performance in Mestre in Italy, 18\(^{th}\) of June 2002. It was found that the maximum acceptable system latency for the Vodhran was approximately 30ms and the Vodhran was perceived as a promising expressive device for a computer augmented percussion performance.

### 6.2.2 Example: Liquid sounds

The sound of a bottle being filled with water is a sound that many have stated is easy to recognise and has several interesting features that can be heard (Ballas, 1993; 34 Traditional Irish frame drum.
35 6-degrees-of-freedom, 120 Hz sampling rate, 4 ms latency.
http://www.polhemus.com/?page=motion_fastrak
Several features can be picked up when listening to such sounds, e.g. the size of the vessel, if it is being filled or emptied, the viscosity of the liquid and the rate of pouring. The everyday perceived properties of this kind of sound suggests that it may lend itself well for user interface widgets such as progress bars, scrolling, or other auditory displays where a direction (up/down) or quantity is to be estimated (Walker 2000).

The most primitive micro-event in such a sound is a single drop of liquid falling into a vessel and/or into a liquid surface in the vessel. In the SOb project initial explorations were based on Gaver (1993b), i.e. the drop creating a small resonating cavity in the liquid that rapidly closes producing a very short wavelet. The pitch of such a sound is a very short burst of pitched sound – onomatopoeically speaking, a *blip*. For a more complex liquid sounds, a number of droplet sounds can be used, statistically distributed, to create the illusion of a stream of water being poured. A vessel being filled is often either cylindrical or a bottle (a Helmholtz resonator) and the modes of such a cavity can easily be modelled as high-resonance comb filters. As the vessel fills, the resonances rise in pitch.

With liquids like water, the rapid increase in pitch for a droplet cannot be heard. In fact, Gaver’s (1993b, pp. 296-297) explanation is not valid for the object falling into a liquid if it is only a drop of water falling into water, e.g. with drop size typically around 2 millimetres in diameter. The Helmholtz-resonance for such a small cavity starts at approximately 20 KHz, which is beyond the ability of human hearing. Instead, what is heard is the air resonance of the vessel holding the liquid, above the surface of the liquid, and in some cases the structural resonance of the vessel itself and the accumulated body of liquid, excited by the wavelet (working more as an impulse response). Hence, it is possible to simplify the model of the droplet and the vessel as simple oscillators with envelope control. For more complex turbulent flows, the model can be simplified by either using granular synthesis or subtractive synthesis, modelling the change in overall resonance in the vessel as high-resonance comb-filters.

For example, Fig 3 shows a simple PD model of a 1 litre cylindrical vessel being filled with water. On the right-hand side of the figure there are four simple resonator objects, each containing a white noise source, a bandpass filter and control logic for making the center-frequency of the filter ‘wobble’ to mimic a turbulent flow. The resonators are tuned and coupled together to generate the four first resonant modes of the vessel. On the left-hand side of the figure are the overall controls for the
model, i.e. degree of fullness and a timed ramp to generate the filling parameter.

Figure 2: PD sound object model for bottle filling sound.

In Figures 4 below and Figure 5 on the next page, the similarity in the sonograms between the real sound (4) of filling a bottle and the sound from the model (5) can be seen.

Figure 3: Sonogram, filling a real 1-litre bottle with water.
In the study reported in (Rocchesso, Bresin, & Fernström, 2003) previous research was expanded and 197 undergraduate students at the University of Limerick listened to 11 recorded sounds of plastic bottles being filled or emptied of water. As reported in (Rocchesso, Bresin, & Fernström, 2003, p.46), the participants judged if bottles were being filled or emptied, when bottles were half-full or half-empty, and when they were just about full or just about empty. When filling a vessel, the pitch is rising and when emptying, the pitch is falling. The pitch of a half-full bottle versus the pitch of an empty bottle has a ratio of a diminished fifth (6 semitones) for a bottle (Helmholtz model), while for a straight cylindrical vessel the ratio is one octave. From a musical listening perceptive, such intervals are easy to detect.

6.2.3 Summary
The work presented in Paper III showed that sound object models could be created based on a physical modelling approach. These models were simplified by cartoonification, which made it possible to implement auditory widgets based on sound object models on lean computing platforms. These auditory widgets belong to the category of evolutionary objects as defined by Buxton, Gaver, & Bly (1994), as they can be controlled parametrically in real-time (see Section 4.5).

Amongst the researchers in the Sounding Object project, a design process emerged which was based on listening tests, brainstorming, prototyping and evaluation as described in Section 5.4.
6.3 Paper IV: HCI Design and Interactive Sonification for Fingers and Ears

During the last year of the Sounding Object project, a couple of issues became more interesting: the recognition of auditory icons and how to design a non-visual, direct manipulation auditory interface. Both issues are of major importance. If users cannot recognise the sounds they hear when interacting with a computer, the outcome will be more uncertain and the user experience will be negative. Secondly, while many researchers have tried to improve existing GUIs with sound, where the GUIs are already optimized for visual interaction, the emphasis here was to create a non-visual interface that could be evaluated without any disturbance from visual information. This has been a challenge discussed over many years in the auditory display community, how to support work with computers when the eyes are busy with other things or when the user is visually impaired.

6.3.1 Choosing sounds that are easy to recognise

This paper covers two different, but related, issues. First, a method to measure how accurately people can identify short everyday sounds. Secondly, the paper reports on a design, implementation and evaluation of a non-visual user interface based on pseudo-haptic auditory widgets. The first issue was to investigate ways to better inform the choice of sounds when designing a user interface based on auditory icons in terms of how easy it would be for users to recognise the sounds. Several previous studies have looked at this problem, in particular James Ballas and his collaborators (Ballas et al. 1986; Ballas & Howard 1987; Ballas & Mullins 1991; Ballas 1993; Ballas 1994; Ballas 1999; Ballas 2002).

During exploration of the possibilities with listening tests, it had been noted that participants sometimes were more accurate in recognising the activity that caused a sound rather than the objects involved in the activity. To investigate this, a study was carried out with 14 postgraduate participants listening to 104 different recordings of everyday sounds, writing down what they believed they heard in a sentence or two per sound. This method is quite similar to the methods used by, for example, Gaver (1988) and VanDerveer (1979). Following this, three research assistants categorised and sorted the responses, according to Ballas’ (1993) method, but in addition to this the participants’ responses were categorised and sorted based both on the reported activity as well as the reported objects involved in the activity.
The study confirmed that the participants recognised the activity more accurately than the objects involved. As a result, separate measures of causal uncertainty ($H_{CU}$) are obtained, i.e. how easy it is to recognise the sound in terms of activity, objects involved and the two combined. If the $H_{CU}$ is low, most participants reported similar and correct identification of a sound, while if the $H_{CU}$ is high, participants reported different and sometimes incorrect identification of a sound. The implication of this is that when choosing sounds for use in an interface based on auditory icons, sounds with low $H_{CU}$ should be used.

The richness of the participants’ narrative descriptions of what they heard is also useful for guiding interaction designers when building auditory user interface metaphors. By first analysing the tasks to be carried out, e.g. open, move, upload, close, the database with the participants’ descriptions can be searched for potential matches. It can then be considered if the causal uncertainty measures for these sounds are sufficiently low, and if any of the sounds can fit a potential metaphor. By using tools such as the Sonic Browser, a sonic mood board can be created to explore if the combination of chosen sounds work (Fernström & Brazil 2001). Sections of the sounds that match a possible metaphor can be edited and extracted, and a rapid prototype can be implemented and evaluated. Finally, when the prototype design has reached acceptable levels of usability and acceptance, the user interface can be implemented as sound object models.

6.3.2 Example of an Auditory User Interface using direct manipulation

The second issue investigated was the possibility of designing a user interface that did not use visual display, but that relied on auditory display only. To contrast the auditory aspects, without using visuals, a tactile/haptic interface was used. Considering that the so-called soft button in GUIs is an important and more or less essential part of a visual user interface, an auditory soft button design was explored. In the case of a visual soft button, while the activation of a button is discrete (on/off/slip-off), finding a button is a continuous process, e.g. moving a visual cursor on a screen by moving a mouse until it is visible within the area of a visual icon representing a button. Visual interfaces with multiple buttons have a spatial structure, a layout, which allows the user to locate, choose and activate different functionalities of an application. A real physical interface with real pushbuttons has both visual and tactile/haptic affordances that allows the user exploration of its structure and activation of system functionality by pressing buttons. Such an
interface can, of course, also be explored and used without the visual modality, i.e. tactile/haptic only. The problem from a designer’s point of view is that real physical buttons are expensive and cannot, with today’s technology, be made to appear, disappear or to structurally reconfigure based on a few lines of code in software. To explore the possibilities with the auditory modality, it can be noted that when touching things there is normally a very faint contact and friction sound. By creating a simple sound object model that cartoonified a friction sound, the model can be dynamically controlled based on the detected location and movement of a user’s fingers on a generic touch-tablet device.

A rich and complex friction sound model as described by Avanzini et al. (2003) is computationally quite expensive and would not lend itself for implementations on lean platforms such as PDAs or mobile phones. A simpler, but fully sufficient sound model in the context of auditory soft-buttons can be created by subtractive synthesis, i.e. a white-noise generator, a band-pass filter with adjustable centre-frequency and bandwidth, and a variable gain factor. Entry, exit and slip-off click sounds can be modelled by the same sound object model, but with different parameters, i.e. amplitude envelope, frequency and bandwidth.

A simple experiment was designed, inspired by the work of Gibson & Stanley (1961) who had twelve blindfolded participants to identify the shape of a cookie cutter, either by just having the cookie cutter placed in their hand and try to ascertain its shape without moving, or, by actively exploring its shape using and moving their fingers to identify the shape. Gibson and Stanley concluded that it was almost impossible to know the shape under the first condition, but in the second condition the active exploration made the task quite easy.

Initially, some sample layouts were made using cardboard shapes glued onto a cardboard sheet. Four volunteers were recruited among our graduates. The participants were blindfolded and asked to identify the shapes on the cardboard sheets by exploring the sheets through touch. It was found that they could easily identify the different shapes and layouts. Encouraged by this result, it was decided to try a metaphor where friction sounds would represent the presence of buttons on a touch-pad, without any tactile structure (a flat surface). A first prototype was made in PD and Adobe Flash, where the Flash application connected to a sound object model in PD via network sockets. The advantage with this kind of prototyping environment is that button layouts and interaction are easy to program in Flash. As had been noted in the Sounding Object project, PD lends itself well to the design of
fully parametric auditory icons. The prototype was tested with three participants. The only disadvantage found with this prototype was some latency between users’ actions and changes in the sound output. This was due to both the lean platform (a Xybernaut\textsuperscript{36} wearable computer) and scheduling issues between Flash and PD. Still, as results in general were positive and confirming that users could explore and identify a spatial layout of metaphorical soft-buttons represented by sound, a fully integrated prototype was implemented to overcome the latency problem. Ten participants, both in a real haptic condition and in a pseudo-haptic condition, evaluated the final prototype where the button areas were represented by sound object models of friction. It was found that using auditory display was almost as accurate and usable as a real haptic display for conveying information about a spatial layout of soft-buttons.

6.3.3 Summary

Ballas’ (1993) method for causal uncertainty was extended, differentiating between identification of the actions and objects involved in the production of everyday sounds. The resulting measures are useful when choosing sounds for auditory icons, auditory widgets, etc., when designing an auditory user interface. The written responses from participants in the listening tests that form the basis for the causal uncertainty measure can help a designer with suggestions for building an auditory user interface metaphor.

Secondly, a non-visual user interface was designed, implemented and evaluated, based on a cross-modal effect created by the application of sound object models that responded to the participants’ movement of fingers on a touch-detecting surface. The results indicate that the participants could easily pick up, and form a mental model of, the spatial structure of soft-button layouts based on the sound they heard with a pseudo-haptic user experience, comparable to the same tasks using a surface with a real physical tactile structure.

\textsuperscript{36} http://www.xybernaut.com/

The Shannon Portal was a complex interactive installation and was part of the Shared Worlds project\(^{37}\), sponsored by Science Foundation Ireland. The overall design is described in Ciolfi \textit{et al.} (2007), while in this thesis the focus is on the design of the auditory display element of the installation (Fernström & Brazil 2009).

One of the issues that were explored in the project was the use of large ambient displays and public interaction with such displays. As the installation at Shannon required an interactive image gallery to be displayed, a purpose-built back-projection system was designed. An overhead camera tracked users’ movements in front of the display, allowing the users to move a virtual magnifying glass through the image gallery. One of the doctoral students working on the project, Paul Gallagher, developed several iterations of software for a visual display and after many trials with different forms of mapping of users’ actions to movement of the virtual magnifying glass, it was decided that it was not the users’ location that would control the movement, but the dynamics of users’ movement. The rationale for this was that while there was only one virtual magnifying glass, there could be multiple simultaneous users. The design intention with such a mapping was that it might encourage collaboration and engagement. During test and evaluation of the system in the lobby outside the Interaction Design Centre laboratory at the University of Limerick it was observed that users who occasionally passed by the display were not aware that the display was interactive. Furthermore, due to the nature of the mapping, some users found it difficult to understand that it was their movement and not their location that controlled the virtual magnifying glass. To address this problem it was decided to explore if an interactive auditory display element could be included in the system to provide users with auditory feedback that would be directly responsive to their movement in front of the display, and help reinforce the metaphor of controlling the virtual magnifying glass.

6.4.1 The challenge of auditory display in a public environment

It has been noted in numerous papers (e.g. (Brewster, 2002; Sorkin, 1988)) that auditory display can be annoying. A location such as the transit hall of an airport has

\(^{37}\) 2004-2007, Principal Investigator Prof. Liam Bannon.
an inherently high background noise level, at peak-time often over 85 dB(A). In the transit hall in Shannon airport, the main contributors to the soundscape are people walking, talking, handling luggage, eating and drinking, mobile phone ring tones, tills in the Duty Free shop, all the activity in the Sheridan Bar and bursts of announcements about flights over the PA-system. Having analysed the airport soundscape and the nature of the interaction with the system, it was obvious that traditional auditory icons and earcons would not be suited for the design. Auditory icons would easily be confused by all the normal activities in the space. Earcons would easily be confused by mobile phone ring tones and the bleeps and pings from PA announcements and tills. Due to this, a number of different auditory displays based on a more abstract approach were explored. The basic requirements of the auditory display were that it should:

- Only sound when users actively moved in front of the image gallery.
- Indicate up, down, left and right as non-discrete, continuous feedback. The parameters received from the overhead motion tracking camera system were the average amount of movement, left-right position and front-back (closeness to display) position.
- Be audible and not be easily masked by the airport’s existing soundscape.
- Be audible but not too loud or mask other sounds, in particular, speech.

The aesthetics of the visual metaphor had a certain sci-fi feel and the setting was in an airport, hence it was decided that an aerospace-inspired sound design might be appropriate. PD was used as the sound designer’s sketching tool to rapidly generate a number of possible sonifications. As there was no direct constraint by the computational platform, an average-performance PC was used for generating a low-latency sonification. All four sound designs described below were created and evaluated over a four-week period, allowing instant access to adjust parameters, or quick modifications based on comments from the participants.

### 6.4.2 Evaluation

All the designs throughout the design process were evaluated. Initially heuristic evaluation was used, considering issues such as masking and mapping. This was followed by informally evaluation of the designs with 3 to 5 people in the Interaction Design Centre at UL. The installation was then, temporarily, put up in

---

38 *ShuttlePC*, Intel Pentium, 2.4 GHz, 512MB RAM, www.shuttle.com
the lobby area outside the Interaction Design Centre, in a semi-public space. In this location, 9 people participated in the evaluations. They were first observed in free casual use of the system and then interviewed. Finally, the installation was moved to Shannon Airport, and over 500 people used the system. Approximately 50 hours of video of real use was recorded and later analyzed. 20 users were also interviewed about their subjective experience.

6.4.2.1 First design: Ping

The first display used an 80 ms decaying pulse from two sinewave oscillators 4 Hz apart, pitched from 200 Hz to 800 Hz. The rate of pulses ranged from 0.3 to 5 Hz. The output was amplitude panned left-right. In essence one could describe this simple sound design as “a machine that goes ping”. The following mapping was used:

- Amount of movement – Pulse rate
- Left-Right – Panning
- Front-Back - Pitch

Participants found this display to be clear but too intrusive (“sounds like a truck reversing”). Part of the intrusiveness in the sound design was due to the sharp onset, which was strongly attracting attention. The pitch used was in the middle of the background soundscape spectrum, hence would require to be loud to avoid masking. All participants found the mapping easy to understand.

6.4.2.2 Second design: PhotonDrive

The second display used a pulse with approximately 0.3 s attack and 0.3 s release from two sawtooth oscillators tuned 2% apart, pitched from 50 Hz to 650 Hz. The output was amplitude panned. The mapping used was the same as in the previous trial.

The softer onset and richer spectrum resulted in participants feeling that the sound design was less intrusive. Due to the richer spectrum, the sound could be played at lower loudness, but was still sometimes masked.

6.4.2.3 Third design: Wind

In the third display a less intrusive wind-like sound model was used. It was based on subtractive synthesis using white noise sources that were bandpass filtered with an amount of jitter in the center-frequencies to make it sound more realistic. The average center-frequency was mapped to the amount of movement.
This sound was noted as not being intrusive as there was no distinct onset, but due to its naturalness, basically being a parametrically controlled auditory icon, and only two mappings (activity – frequency, left-right panning), the sound would not attract sufficient attention.

6.4.2.4 Fourth design: Shepard tones

In the fourth and final display a more radical approach was used. This sound design was based on the Shepard tone illusion (Shepard 1964). This illusion is perceived as eternally rising or falling pitch and is due to the continuous frequency change and mix of ten partials. The partials were carefully adjusted to an unnatural timbre (bell-like) to try to avoid masking effects in relation to the surrounding soundscape in the airport.

- Amount of movement – loudness
- Left-Right – Panning
- Front-Back – Pitch direction

Participants found this display to be positively intriguing, as it did not sound like anything they had heard before. Due to a soft onset combined with an unnatural timbre, it was noted as being both pleasing as well as attention catching. A participant or group of participants walking past and accidentally being tracked and sonified immediately understood that it was their movement that caused the sound. This observation is similar to that found with Todd Winkler’s (2000) installation Light around the edges39 where people passing through a lobby area of a building discovered that their movements were directly mapped to humorous sonifications.

The installation was moved to Shannon airport where it was installed for 3 weeks. Both naturalistic observations and interviews were carried out during this period. Both passengers and staff at the airport found the sound design contributed to their experience and was non-intrusive, with the previous findings being confirmed.

6.4.3 Summary

In this paper the development and evaluation of an interactive computer system for casual use in a public environment was reported. The system included an interactive sonification that supported users to engage with the system. The findings include the importance of first studying the soundscape of the location and taking psychoacoustic issues into account when creating prototypes. The importance of

39 http://www.brown.edu/Departments/Music/sites/winkler//lightaroundedges/
rapid prototyping and continuous evaluation was also noted.
For the actual sound object design presented in this paper, tight coupling between the users’ actions and dynamics in the resulting sound are critical for the success of the interaction style with full-body movements in relation to a movement of virtual magnifying glass across an image gallery on a back-projected screen. The soundscape was also analysed with the sonification, on location, to confirm that other people in the place where the system was deployed, while not using the system, did not experience the sound as annoying or disturbing.
7 General discussion

This chapter summarises the contribution of the five papers presented in this thesis and discusses how the results answer the research questions, as stated in Chapter 5. The five papers cover a range of issues in auditory display and, in particular, the design and evaluation of auditory widgets and sonic interaction design.

In Paper I and II a software application to facilitate faster and easier access to a corpus of sound files on computers was created, initially for music and then for any kind of sound file. The human ability to selectively attend to auditory streams is utilised (Entendre in Shaeffer’s terminology (Schaeffer 1966); or, the cocktail party effect (Cherry 1953; Cherry & Taylor 1954; Arons 1992). While the initial aim was to have a full 3-D spatialization using generalized HRTFs to utilize the human ability of binaural listening, the spatialization had to be reduced to stereo using panning due to the computational power available at the time, but this was fortunate as it allowed the discovery that the tight coupling between the user’s actions and the visual and auditory feedback was sufficient for the participants to quickly browse through a set of sound files and find a target of interest. The reason for this may be similar to what in film sound is called the audiovisual contract and magnetization, i.e. that sounds get ‘locked’ by the viewer/listener’s perception to where the visual action is (Chion 1994). It was found that it was significantly faster to locate sounds when listening to multiple sounds concurrently, compared to listening to one sound at a time.

Having developed the initial sonic browser application, other new domains were explored for this kind of tool. It became a tool for sound file/sound object model content management as well as a tool for perception experiments (spatially sorting sounds depending on the perceived property of sounds, either real recorded sounds or sound object models) in the Sounding Object project. An interesting feature of using the Sonic Browser in perception experiments is that participants can be asked to respond in two dimensions, simultaneously, by sorting and dragging icons of sound files into groups and arrangements both vertically and horizontally in a starfield visualization (Brazil, Fernström, & Ottaviani, 2003b).

Several other researchers have since explored sonic browsing and direct sonification. For example Heise et al. (2009) developed a system called SoundTorch for browsing sound effect libraries. SoundTorch closely resembles the Sonic
Browser in terms of stereo-spatialized multiple-stream audio, a starfield visualisation and user navigation. Uzuki et al. (2011) developed *iSoundScape*, an iPhone app for browsing audio clips, which closely resembles the work in Paper II. They cite Paper III but refer to sonic browsing. Stewart & Sandler (2011a; 2011b) developed *ambir*, a music browser and playlist manager for mobile use, partly inspired by Paper I & II. Dupont et al. (2009) also refers to the Sonic Browser work in their *AudioCycle* system for browsing musical loop-libraries. While not explicitly citing the work in this thesis, Björk’s *Biophilia* application for *iPad* includes a pseudo-3D starfield visualisation and sonic browsing concepts for navigating between albums\(^{40}\) (Cragg 2011).

Today, the main distribution of music and media is moving to the Internet, giving people access to millions of music and sound files. The issue of finding easier and alternative ways to locate content of interest has become increasingly important. The traditional user interfaces for browsing music via the Internet are mainly based on text-based searches on titles, artists and genres that then require users to click on an icon to hear a single snippet of audio. By enabling users to quickly navigate and switch between different sounds, utilising the human ability to hear and segregate multiple-stream audio, the human activity of browsing for sounds on computers is augmented.

Paper III, IV and V are mainly about new forms of interactive sonification. Paper III reports the main findings of the Sounding Object project. The overall brief for the project was to develop and investigate sound object models that could easily be integrated in computational artefacts and be made responsive to the user’s handling and gestures. One important issue was to make the sound object models expressive, i.e. to respond proportionally to human action similar to how real objects respond (a small movement or object would make a *small sound*; larger movements or objects would make *larger sounds*, etc.). Novel forms of continuous action and feedback were explored, similar to what Buxton, Gaver, & Bly (1994) suggested for *evolutionary objects*. On the level of sound design, the main approach from Italian colleagues was physical modelling (Avanzini et al. 2003), while the work reported in this thesis explored both their physical models as well as creating physically inspired models. The reasons for this was primarily to be able to have the sound

---

\(^{40}\) The term *albums* here refers to interactive sub-applications that allow the user to create a personalised mix of Björk’s work.
object models run on lean platforms such as wearable computers, as the full physical model approach would require more computing power. At the end of the Sounding Object project, the first prototype of pseudo-haptic soft-buttons in a non-visual user interface had been implemented on a wearable computer. This version of the soft-buttons had some latency issues due to the lean computing platform. After the Sounding Object project, the work continued on soft-buttons and an experiment was designed to evaluate this new type of user interface widget. The results reported in Paper IV show that it is possible to use an interactive sonification with a friction-like sound object model to provide the user with continuous feedback when moving his or her fingers across a touch detecting device, giving the user a pseudo-haptic experience through sound of a spatially structured interface area.

Based on the Sounding Object project 2001-2002 (Rocchesso & Fontana, 2003), a number of research projects and collaborations followed, for example, the EU COST Action 287 Gesture Controlled Audio Systems 2003-2007 (Godoy & Leman 2009), the CLOSED project 41 2006-2009 and the EU COST Action IC0601 Sonic Interaction Design (Thomas Hermann et al. 2011). Sonic interaction design as a specialisation emerged from this and can be seen as where interaction design meets sound and music computing and sound design.

Paper IV also presents a study with listening tests to determine how accurate recordings of everyday sounds are identified. Ballas’ (1993) method of causal uncertainty was used and expanded to differentiate between object and action identification. It was found that actions were better identified than objects. One possible explanation for this can be found in David Worrall’s (2010) paper on Parameter Mapping Sonic Articulation and the Perceiving Body, where he points to the work of Kohler et al. (2002) and the discovery that audio-visual mirror-neurons respond both to our own actions and when we perceive the sound of similar actions by others.

In two listening experiments involving sound identification Lemaitre & Heller (2012) have also confirmed that actions are more identifiable than materials. Both accuracy as well as reaction time was better for actions than materials.

The work in Paper V was carried out by a multi-disciplinary design team in the Interaction Design Centre at UL. To design sonifications for a public environment

41 http://closed.ircam.fr/ (accessed 2012-02-07)
such as an airport required a different approach than that used in Paper III and IV.

The design ambition was to develop a ubiquitous computer system for a public setting. Instead of using traditional metaphors, for the goal was a system that would be immediately obvious to the users instead of being weighed down by traditional user interface clichés or forcing users to read instructions before using the system. This is always a problem in computing for public environments, as most users would never have encountered a system like this before (at least, not the first time) and within a few seconds the system has to communicate to potential users that it can be interacted with, and then guide users as to how to proceed.

From the perspective of this thesis, the main challenge was to create sonification that would be responsive, attention grabbing and non-intrusive. The sonification had to be audible in the vicinity of the installation, without masking other important sounds such as conversations or PA announcements. The sonification also had to communicate the amount of user activity detected by the computer-vision system, and as a result the up-down and left-right movement of a virtual magnifying glass that users could control, based on their body movements, in a projection of a photo gallery. From a sound design perspective, this included the same challenges as Ben Burtt faced when designing the light sabre sounds for the movie Star Wars (1977), i.e. how to design a sound of something that does not exist or that would not make a sound in the real world.

In the design process, it was noted that while it might appear to be easy, it is sometimes difficult to get a design team to simply onomatopoetically vocalize ideas for an auditory display in a bodystorming session, probably because participants feel it is what children do when playing, adding vocal sound effects to their actions, and as an adult ‘we don’t do that’. Still, it is the most immediate way to communicate a sound – what can be vocalized can be heard.

The final design was not an auditory icon, but a symbolic sound object model with continuous control, using the features of the Shepard tone illusion for up-down features (Shepard 1964), panning for left-right, and loudness for the amount of activity, hence communicating three different dimensions with one sound, as well as catching the attention of users near the installation without being intrusive. From a design perspective, there are many possibilities. The difficulty is to quickly home in on plausible alternatives, in this case sounds that can be played at lowest possible loudness while not being masked by the surrounding soundscape, yet still be informative. In the evaluation, in the interviews, all users stated that they never
encountered a system like this before but they were all able to use it immediately. On the issue of annoyance, there is only one way to find out if a particular design is acceptable or annoying, by following the users’ experience of the auditory display over a long period of time. The research team was allowed to be on-site with the finished system for three weeks. None of the users and none of the non-users (for example airport staff in the vicinity of the system) reported any annoyance with the auditory display element of the system. However, this is simply an informal observation and the issue would benefit from a more thorough investigation. A similar approach and findings have been discussed by for example Serafin et al. (2011).

Finally, the work in this thesis has shown that the proposed approach aligning the frameworks of Gaver (1993a; 1993b), Schaeffer (1966), Schafer (1966), Truax (1984; 1999) and Chion (1994) can be informative, productive and applicable to sonic interaction design. With this combination, a richer vocabulary is available creating a deeper understanding of what an auditory interface can offer the user. This is necessary for improving the user experience allowing us to move on from discussions about earcons and auditory icons. Interaction designers may need to consider how interactive audio can contribute to the mimesis of user interaction, with interaction shifting from the screen to the world. With well-designed interactive systems, it can be an intricate interplay between humans and computers, as affordances are dynamic properties of the environment and artefacts that afford action, giving users the experience of a more natural interaction (Norman 2007; Juul 2009).

7.1 Methodology issues
The auditory material used in Paper I was music from the genre that our users were familiar with, i.e. traditional Irish music. This may have contributed to the users’ success and speed in finding target tunes, but it can be assumed that as the same kind of material was used for both conditions (single sound at a time versus multiple concurrent sounds), such an effect can be discounted. In Paper II when using the sonic browser for other domains and tasks such as managing collections of sound files with everyday sounds or sound object models, it was not investigated if there was any difference in the time to completion of tasks, as in the previous work, but the application was used and explored by a small user population (the researchers
participating in the Sounding Object project ≈ 20 people).

In Paper III there was a plethora of methods involved. The work included application and adaptation of traditional methods from HCI, such as thinking-aloud (Lewis 1982), measurement of time-to-completion and number of errors, as well as some work on psychophysics investigating how users could detect bottles filling and emptying with water, based on the recorded sound of such events (inspired by Bruce Walker’s (2000) work on Magnitude Estimation). The experiment was conducted with undergraduate computer science students in their final year. The experiment was in the form of a small downloadable program that each student downloaded and executed. Some students carried out the experiment in a computer science lab, while others used their own laptops in their home environment with fewer distractions.

In Paper IV there are two different contributions, a method to inform designers about suitable sounds for auditory icons and the design and evaluation of pseudo-haptic soft-buttons. Based on James Ballas’ (1993) work on a measure for causal uncertainty in sound identification, a study was conducted that looked at this measure both in terms participants identifying actions as well as objects involved in activities making everyday sounds. It was found that the participants were better at identifying action than the objects involved. The second contribution in Paper IV is the evaluation of pseudo-haptic soft-buttons. From a methodological perspective, to compare real haptic perception with pseudo-haptic perception (using sound), with identical structural configurations presented in random order in each condition gave a real insight into how cross-modal effects can be applied. Still, one reason that real haptic buttons were faster and slightly more accurate for participants to detect is probably because the way in which the contact between a fingertip and a real tactile object works. A single fingertip can pick up differences in tactile structure within the whole area of the contact, i.e. containing more information about for example an edge, while the pseudo-haptic condition only provides information about a single point, the centre-point of the contact area.

In Paper V, using the experimental platform of PD and the approach of sound object models allowed rapid prototyping of a number of different sounds and informally evaluating them in a laboratory and in a lobby outside our laboratory, before arriving at a final design that was used in the installation. Without the sound, most users did not get the idea that the virtual magnifying glass was interactive and that their body movements would control it. This statement is based on our casual observations of users and informal interviews during the development. With the
sound, people walking past the installation almost immediately reacted to the sound responding to their movements. Their surprise experience appears to be similar to when a walking past a mirror, seeing own movements reflected.
8 Conclusions

This thesis has presented a number of topics concerning the design of highly interactive auditory user interfaces and widgets. A framework for designing auditory widgets has been proposed in Chapter 5. The five papers included in this thesis and the research questions they answer support this framework.

As an alternative or complement to traditional search methods and visual symbolic representations of corpora of sound files, it has been shown how software applications such as the Sonic Browser can augment human ability to selectively attend to multiple auditory streams and enable users to serendipitously navigate an interactive soundscape in search for items of interest. A software application such as the Sonic Browser can also be used for general music retrieval tasks, as well as for content management of corpora of sounds and sound object models, or as a tool for auditory perception experiments (Fernström & McNamara 2005a; Fernström 2005).

A sound object model approach have been described and shown that it can provide novel ways to create interactive sonifications for ubiquitous computing and be well suited for design and implementation in this domain. When the sound of interaction becomes as responsive (or more responsive) than visual user interfaces, new and richer possibilities in ubiquitous computing may arise, in particular for non-visual user interfaces for mobile and wearable use (Rocchesso, Bresin, & Fernström, 2003).

Design methods for auditory user interfaces based on auditory icons have been reported, where methods such as measurement of causal uncertainty in terms of both actions and objects to determine the identifiability of sounds were used. As the method is based on listening tests with participants giving short descriptions of what they believe they hear, their responses can also be used as suggestions for metaphor construction when designing an auditory user interface. For example, to represent ‘open’, the database of participants’ responses can be searched and, for example, finding ‘opening a door’, ‘opening a can of fizzy drink’, ‘opening and old rusty door’, all with low causal uncertainty values. These responses may inspire the design of a general system metaphor (Fernström, Brazil, & Bannon, 2005b).

With the sound object model approach, novel forms of auditory widgets can be designed, for example for wearable and mobile computing, that can completely free the visual modality. With the example of the design and evaluation of pseudo-haptic
soft-buttons this thesis have shown that it is possible to give users a pseudo-haptic experience of a spatially structured touch device through sound (Fernström, Brazil, & Bannon, 2005b).

Finally, it was shown that novel forms of auditory display can be designed and used for augmenting interaction in a public place, taking into account how the auditory display is situated in a soundscape, and that the auditory display needs to be audible and informative while not disturbing people (Ciolfí et al. 2007).
9 Future research

The idea of Sonic Browsing can be further explored in many different ways. Over the last decade an increasing amount of audio content has moved to distribution via the Internet but the mechanisms provided for finding interesting content have largely remained based on textual indices, or, recommendation algorithms (e.g. ‘if you like this, people who liked this, also likes that’. See for example (Zhou et al. 2011; Tan et al. 2011; Ò. Celma 2010)). Musicians, composers and music lovers often search for multiple versions of a piece of music and doing so by a literal search query can be quite time-consuming and demotivating, for example, listening to one snippet at a time via Apple’s iTunes Store. It may be possible to combine textual search with Sonic Browsing and fewer user actions would then be needed to find content of interest. Such a combination may be a move towards Ben Shneiderman’s information seeking mantra (Shneiderman 1996): “Overview first, zoom and filter, then details on demand” - in a true multimedia sense. It would also be interesting to investigate the possibilities with Sonic browsing and multi-touch surfaces, both for perception experiments as well as a tool for audio content management for sound designers. The aura-function (as described in Section 5.4) may be controlled by multiple fingers or bimanual gestures.

There is ample scope for continued exploration of sound object models, for example investigating and developing auditory widget libraries that can be used by interaction designers without them having any extreme expertise in auditory perception, digital signal processing and sound synthesis. There might also be scope for research on both operating system level and electronic hardware, i.e. if audio processing can be improved in a similar way as graphic adapters and co-processors have evolved.

For auditory display research it is important to continue exploring how multidimensional data can be mapped to, for example, sound object models, and how to make auditory displays highly responsive to users actions. From an artistic perspective it is time to set sound free, instead of being the neglected sibling behind stunning 3D graphics. The sound object model approach may be further explored in the context of virtual Foley as the increasing number of inexpensive and accurate gestural controllers (e.g. Nintendo Wii, Microsoft Kinect) can be used for controlling sound object models to be potentially more expressive than real physical props.
10 References


Maasø, A., 2002. Se-Hva-Som-Skjer!": En Studie Av Lyd Som Kommunikativt Virkemiddel I TV. , Ph.D.


Somers, E., 2000. Abstract Sound Objects to Expand the Vocabulary of Sound Design for Visual and Theatrical Media. In Georgia Institute of Technology, Atlanta, Georgia, USA.


75


Paper I
After Direct Manipulation—Direct Sonification

MIKAEL FERNSTRÖM and CAOLAN McNAMARA
Interaction Design Centre, University of Limerick, Ireland

The effectiveness of providing multiple-stream audio to support browsing on a computer was investigated through the iterative development and evaluation of a series of sonic browser prototypes. The data set used was a database containing music. Interactive sonification was provided in conjunction with simplified human–computer interaction sequences. It was investigated to what extent interactive sonification with multiple-stream audio could enhance browsing tasks, compared to interactive sonification with single-stream audio support. It was found that with interactive multiple-stream audio, the ten users could accurately complete the browsing tasks significantly faster than those who had single-stream audio support.

Categories and Subject Descriptors: [General Literature—General]—Conference Proceedings

General Terms: Experimentation, Human Factors, Performance

Additional Key Words and Phrases: Sonification, auditory display

1. INTRODUCTION

This paper concerns the development of an interface that allows musicologists to browse musical data sets in novel ways. The data set (in the users’ language often called a collection) is used by musicologists in their research. It contains over 7000 tunes, where each tune is represented by its score and a number of properties, such as tonality and structure. The traditional format for a collection is a printed book with various indexes. A common problem that musicologists have to deal with is to determine if tunes they collect in their field work exist in a particular collection and, if so, how they are related to other tunes in the collection, e.g., in chronology, typology.

1.1 Browsing

Browsing has become a popular term in recent years with the emergence of hypertext systems and the World Wide Web, but the concept of browsing goes well beyond these fields of application. There are many ways integrating text, sound, images, and video to provide richer and more interesting systems that would allow us to use more of our natural abilities. Marchionini and Shneiderman [1988] defined browsing as:

- “an exploratory, information seeking strategy that depends upon serendipity”
- “especially appropriate for ill-defined problems and for exploring new task domains”

This research was supported by Forbairt (ST/94/715)/Enterprise Ireland (SC/97/603).

Authors’ addresses: Mikael Fernström, Department of Computer Science and Information Systems The University of Limerick, Ireland; email: mikael.fernstrom@ul.ie; Caolan McNamara, Department of Computer Science and Information Systems The University of Limerick, Ireland.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 1515 Broadway, New York, NY 10036 USA, fax: +1 (212) 869-0481, or permissions@acm.org.

© 2005 ACM 1544-3558/05/1000-0495 $5.00

This is the case when musicologists are searching for tunes in a collection. Tunes collected through fieldwork can often be different than older original versions. They can still be the same tunes but with the addition of an individual performer’s style. This makes it difficult to use normal computer-based search algorithms [Ó Maidín 1995]. Humans have an outstanding ability to recognize similarities in this domain, which suggests that in a good solution we should make use of our auditory abilities. However, current interfaces do not provide much in the way of support for browsing sounds.

1.2 Direct Manipulation

Most of today’s interactive multimedia applications use direct manipulation. Items on display that can be interacted with can, for example, be highlighted when the cursor is over them, or the shape of the cursor can change. When an object is selected by a single mouse-button click, the object that it has been selected shows and, when double-clicked, the functionality associated with the object is activated.

We can summarize direct manipulation in the following [Hutchins et al. 1986; Shneiderman 1983]:

- Continuous representation of the objects of interest.
- Physical actions instead of complex syntax
- Rapid incremental reversible operations whose impact on the object of interest is immediately visible.
- Users get immediate feedback from their actions.

1.3 Browsing with Sound Support

In everyday listening, one is often exposed to hundreds of different sounds simultaneously and is still able to pick out important parts of the auditory scene. With musical sounds, or tunes, many different factors affect our ability to differentiate and select between the sources. Using instrumental sounds, the timbre, envelope, tonal range, and spatial cues support the formation of auditory streams [Bregman 1990, pp. 455–528]. The tunes themselves also assist the formation of streams, as music has its own inherent syntactic and semantic properties [Serafine 1988]. It is also interesting to note the “cocktail party” effect, i.e., that it is possible to switch one’s attention at will between sounds or tunes [Arons 1992; Cherry 1953; Cherry and Taylor 1954; Schmandt and Roy 1996; Wickens 1992, p. 103].

Albers [1996]; Albers and Bergman [1995] added sounds to a web browser, but kept the use of sound at a fairly low level of interactivity. Various “clicks” were used when users clicked soft buttons and selected menus. To indicate system events, such as data transfer, launch of “plug-ins” and, for errors, he used “pops and clicks,” sliding sounds, and breaking of glass sounds. For feedback about content, various auditory icons were used to indicate what kind of file a hyperlink was pointing to and the file size of the content indicated by piano notes (activated when the cursor was on a hyperlink). He also created hybrid systems using combinations of auditory icons, auralization, and sound spatialization to enhance operator performance in mission control work settings [Albers 1993, 1995].

LoPresti & Harris’ loudSPIRE system [1996] added auditory display to a visualization system. This system is an interesting hybrid as it used at three different layers for sonification. System events were represented by electronic-sounding tones associated with computers; data set objects were represented by percussive or atonal auditory icons parameterized for object properties; domain attributes were represented by themes of orchestral music, harmonious tonal sounds, and parameterized for attribute value of a region.

Begault [1994] demonstrated the use of 3D sound spatialization for use in cockpits and mission control, in order to enhance speech perception. Kobayashi and Schmandt [1997] showed that multiple-stream speech perception can be enhanced through 3D sound spatialization, including the existence of a spatial/temporal relation for recall of position within a sample of streamed speech, i.e., that the auditory content can be mapped to spatial memory.
After Direct Manipulation—Direct Sonification

With multiple auditory streams it is interesting to note the problem with differences in the individual ability to differentiate between multiple sound sources. A metaphor for a user controllable function that makes it visible to the user is the application of an aura [Benford and Greenhalgh 1997], which in this context, is a function that indicates the user’s range of interest in a domain. The aura is the receiver of information in the domain.

2. PROTOTYPE DEVELOPMENT

Three design iterations were performed. In the first iteration, exploratory interviews with potential users were conducted. Mock-ups and use scenarios\(^1\) were created together with potential users. In the second iteration, a prototype was created in a high-level authoring package and informally tested through subjective evaluation. In the third iteration, a prototype was developed in MS Visual C++. This prototype was then tested in a Thinking-Aloud study [Gould 1988].

2.1 Users, Tasks, and Environment

Throughout the development process, groups of users participated in the design and evaluation. They were all musicologists familiar with traditional methods and resources for their work, i.e., fieldwork and access to collections of music in paper-based formats. All users had some experience with computers, for example, word processing and email.

A task list was developed for the testing of the final prototype based on a scenario developed in the first iteration. The task list was considered realistic by users from the first and second iteration. The idea behind the task list was to get users to work primarily in the auditory modality, with the visual modality as additional support. A total of 13 tasks were created, of which 3 were visual. The reason for having three visual tasks was to make the overall session more realistic and to break the “monotony” that otherwise might develop. The order of tasks was randomly allocated to each user.

A workstation was set up in an open plan office, which was similar to the normal work setting of the users. One PC\(^2\) was running the sonic browser, another PC was used by the experimenter playing sample tunes that the user should try to locate by browsing. The users’ speech and actions were recorded on video.

2.2 Sonic Software

Normal multimedia PCs cannot play multiple sound files concurrently. This would, of course, prohibit the desired development. To work around this problem, new intermediate drivers for the sound devices were developed. The problem with existing drivers is that when a sound is to be played, the operating system allocates the physical sound device exclusively. To solve this problem, the intermediate drivers have to read sound files and transform them into a common output format. Sound spatialization was implemented to assist the users in differentiating and locating tunes. With sampled sounds, 3D spatialization can be used, but currently there is no existing support for 3D spatialization of MIDI synthesizer sounds on PC sound cards. Only stereophonic “pan” with difference in loudness between the left and right channel is available on standard sound cards [CreativeLabs 1996; Microsoft 1996]. The problem with different speeds and formats of source files applies to both sound files (such as, WAV) and sound-controlling files (such as, MIDI). As the users had expressed a preference for melody lines with MIDI-controlled synthesizer sounds, all further implementation work focused on stereophonic spatialization with only the difference in loudness between the left and right channel as a cue for auditory spatial location.

\(^1\)Describing the context of use, the users and their work and environment.

\(^2\)Intel Pentium, 120 MHz, 32 MB RAM, 17” display with 1024 x 768 pixels in 16-bit colour, Creative Labs SoundBlaster 16 sound card with OPL3 FM synthesis, loudspeakers Altec Lansing ‘Multimedia’ stereo 2 x 5 W, Microsoft Windows 95 v4.

The users found that they sometimes wanted the aura on, sometimes off, as this allowed them to shift their focus between the neighborhood of tunes to finer differentiation between just a few tunes. The number of tunes within the aura can vary due to the location of the cursor in relation to the density of the data set. Therefore an on–off function was added and the radius of the aura was made user controllable.

3. PROTOTYPE EVALUATION
The prototype was evaluated to test the hypothesis that the application of multiple-stream sound enhances browsing tasks, compared to browsing with single-stream sound. It was evaluated by ten musicologists divided into two groups. The first group had the aura function disabled, i.e., they only had one tune played at a time when the cursor was positioned on a tune object. The second group had the aura function enabled and could switch the aura on or off or resize it at any time during the tests. With the aura on, this group could obtain up to 16 simultaneous auditory streams.

4. RESULTS
In each specific task, the users were allowed to move the cursor around freely in the display and soundscape trying to find the target tune (the sample tune presented to them at the start of each task). Occasionally, the single-stream tasks were solved faster than multiple-stream tasks, but in no case were these differences statistically significant. Cumulative times were significantly faster in the multiple stream condition \((p < .05)\). Overall, for the ten auditory tasks, the total time it took the users to find the target tunes show that all users with multiple-stream sound support were faster than the users with single-sound support. It was verified that there was no correlation between the users familiarity with computers and the task completion times. The users with multiple-stream sound support were approximately 27% faster at locating the target tunes.

From the Thinking Aloud study, there is a good indication that users remember where they heard a tune before, since users that browse with the aura on, hear more tunes. This indication is also supported by, for example, Kobayashi and Schmandt [1997].

5. DISCUSSION
There are many limits to the traditional forms of data sets. The paper-based version is merely a well-structured repository of information, but requires a substantial amount of skilled work (by the end user) to be usable. A straight, text-based database version only slightly improves the accessibility. It takes substantial effort to compare a “fuzzy” sample (target tune) to hundreds of possible score representations.

The interfaces in many standard applications from some of the larger software developers have become overloaded and complicated in the interaction sequences. Through a simplified interaction sequence, users can work efficiently and with a high degree of satisfaction. The results also show that through tight coupling of the interaction, we can create a more engaging interface. By shifting some of the load from the visual to the auditory modality, we can perceive more information and make better use of our natural ability to recognize complex and “fuzzy” patterns through seeing and hearing.

6. CONCLUSION
We could add a new word audibility to Don Norman’s [1988] two key principles for good interaction design: visibility and affordance, because we are dealing with multimedia systems and sound, in particular. Audibility, in this sense, is the concept of how well a system can use auditory representation in the human–computer interaction. If the audibility is good, the users will perform their work better,
faster, with fewer errors, and a higher degree of satisfaction. If the use of sound in the user interface can provide more affordances, or affordances that are complementary to the visual interface, we have a system with good audibility.

This is also important for users with different abilities. By using sonic representations (or auditory display) in the human–computer interaction, the resulting applications will potentially be usable to visually impaired people.

There are many issues that need to be further investigated if we want to develop guidelines and tool kits for good audibility. Further investigations in perception and cognition at high levels of environmental complexity are required. Many guidelines are based on extremely isolated experiments. Hence, it is difficult to apply such guidelines in real-work settings. To get more realistic models for what we, as human beings, can process, combinations of seeing, hearing, and interaction should be studied.

REFERENCES


Received March 2005; accepted June 2005

Paper II
Reflections on Sonic Browsing: Comments on Fernström and McNamara, ICAD 1998

MIKAEL FERNSTRÖM
Interaction Design Centre, University of Limerick

In the original Fernström and McNamara paper at ICAD 1998, we described our work on a novel user interface for sonic browsing of music collections. In this commentary, we reflect upon the continued research in this area and how our interactive sonification approach has come to be used also for browsing collections of everyday sounds and for perception experiments.

Categories and Subject Descriptors: [General Literature—General]—Conference proceedings
General Terms: Experimentation, Human Factors, Performance
Additional Key Words and Phrases: Sonification, auditory display

1. HISTORICAL CONTEXT

The idea of Sonic Browsing originated during the time I was working as a graduate student at the University of Limerick on a project entitled Novel Multimedia Browsing Mechanisms (1995–1996). Our focus was on enhancing the auditory aspects of interfaces, as our personal experiences and review of the literature, indicated that few researchers had addressed this particular field of research. In exploring how sound might assist people in their computer-based work activities, we became interested in the problems encountered by people browsing music files.¹ These observations made us question, why, for example, normal multimedia computers could, at that time, only render a single sound file at a time, while humans are fully able to handle perhaps hundreds of sound sources simultaneously. This led to the development and subsequent iteration of a number of prototype interface mechanisms that attempted to exploit the human ability to pay attention to multiple sound sources simultaneously. In our studies, we were able to show that through such a multimodal (aural and visual) interface—providing multiple-stream audio support—users could find an auditory target substantially faster than with traditional human–computer interfaces [Fernström and Bannon, 1997a, 1997b; Fernström and McNamara 1998].

¹Our colocation with researchers from the Irish World Music Centre was fortuitous, providing us with interesting people and tasks to observe. Subsequent interviews with them also assisted in the development of our ideas.
We continued exploring this approach in a series of subsequent research projects, funded both nationally, and at the European level. These projects included Sound of Action (1997–2000), the EU project the Sounding Object (2001–2003), and a follow-up project on Multimodal Interface Browsing Mechanisms (2003–2005). Throughout this period, the ICAD community and conferences have been one of our primary forums for information exchange and discourse.

2. RESEARCH PROCESS

The underlying philosophy of our research group is to explore the use of computers to augment human ability, i.e., by all possible hardware and software means to adapt the machine to support human activities, rather than view the computer as a substitute for human skill. This approach takes inspiration from the early work of Douglas Engelbart [Bannon 1989]. Our investigations are also set in the context of the human activities, in which people are engaged, rather than focusing on laboratory tasks defined by the experimenter. This does not mean that we do not, at times, examine certain behaviors in the lab, but the original motivation for our studies come from actual tasks that people perform in their everyday life. It was in observing and discussing with our users that we realized how the then-current interfaces to computer systems were so inadequate in handling auditory material. We decided to focus on the interaction process and to search for new ways to understand and enhance our interaction with auditory displays.

3. BODY OF WORK

Early results from the evaluation of our first Sonic Browsing prototypes showed that people could find a target melody in a corpus of melodies substantially faster than with other methods (see Figure 1). Initially we had some difficulty fitting the concept sonic browsing into the existing ICAD vocabulary. Hence, we coined the term Direct Sonification to refer to this particular kind of auditory display, which can be characterized as:

- The information sought is also the information displayed,
• Tight coupling and direct manipulation, i.e., a low latency and highly responsive interaction,
• An *aura* function setting the user’s spatial range of hearing in a virtual soundscape, allowing for multiple melodies to be heard simultaneously.

This facilitated the human ability to deal with multiple-stream audio and spatial organization [Fernström and McNamara 1998].

The direction of our work then shifted to research on auditory user interface widgets based on auditory icons. This was our main focus in the Sounding Object project [Rocchesso et al. 2003], where we also used new versions of the sonic browser as a research tool both for cataloguing sound object models [Fernström and Brazil 2001] and for perception experiments [Brazil and Fernström 2003; Brazil et al., 2003; Ottaviani and Fernström, 2003]. This version of the Sonic Browser (see Figure 2) handled the most common kinds of sound files, e.g., wav and aif.

Finally, in our project on Multimodal Interface Browsing Mechanisms, we have used sound objects to create interactive surfaces without any visual display, where the auditory display creates a pseudohaptic user experience [Fernström et al. 2004]. We have also started to prototype sonic browsing on wearable and handheld computers (see Figure 3).

4. RELATIONS TO THE FIELD OF AUDITORY DISPLAY

Our research has been motivated by a fascination with the myriad of ways in which sound perception plays an important role in our everyday lives and how we might develop our computer-based systems to take greater advantage of the auditory dimension to support human activities. Kramer’s book [1994] has served as a defining document, providing a coherent vocabulary to an emerging field. Gaver’s work...
Reflections on Sonic Browsing: Comments on Fernström and McNamara

Fig. 3. Apple iPod. Mockup Java application for sonic browsing on a handheld device.

[1989, 1993a, 1993b, 1994] has also been highly informative and inspiring, especially his focus on the use of everyday sounds in the user interface. Other researchers, e.g., Brewster et al. [1994a, 1994b, 1995]; Brewster [1997, 1999, 2002a, 2002b] have emphasized the importance of evaluation and usability of auditory interfaces, which has also influenced our approach. Finally, there is more to auditory display than the pure sciences, as pointed out by Eric Somers [2000]. To get users to accept auditory display, we need to consider the aesthetics as well.

5. CONCLUDING THOUGHTS

Some of the visionary predictions we made in our early work, e.g., that people would soon have immediate access to thousands of music and sound files, have become reality. For example, on a 40 GB Apple iPod you can have approximately 11,000 songs. It can be quite cumbersome using traditional text-based browsing mechanisms to find your sounds and songs, while a Sonic Browser approach is sometimes more immediate and does not require the user to remember the names of songs, artists, and albums. We hope that the issues we have raised in the ICAD community, such as observing everyday activities, understanding interaction, taking usability seriously, linking the artistic and scientific communities, has helped others to think creatively about auditory display.

ACKNOWLEDGMENTS

Many thanks to all the graduate students who have worked with me on the generations of Sonic Browsers: David Boland, Caolan McNamara, John Burton, Hugh Roberts, and Eoin Brazil; all staff and students who have volunteered in our experiments and evaluation; and my friend and colleague Liam Bannon who has provided endless critique and encouragement over the years.

REFERENCES


Received March 2005; accepted June 2005
Paper III
Interactive systems, virtual environments, and information display applications need dynamic sound models rather than faithful audio reproductions. This implies three levels of research: auditory perception, physics-based sound modeling, and expressive parametric control. Parallel progress along these three lines leads to effective auditory displays that can complement or substitute visual displays.

When designing a visual widget we attempt to understand the users’ needs. Based on initial requirements, we can sketch the idea on paper and start to test it. We can ask potential users for advice and they might tell us that, for example, “it should be like a rectangular button, with a red border, in the upper left hand corner, with a frame around it.” As testing and new sketches evolve, we can then code the widget and do further tests.

Designing auditory-enhanced interfaces immediately poses a number of problems that aren’t present in the visual case. First, the vocabulary for sound is vague—for example, “when I do this, it goes whoosh,” or “it should sound like opening a door.” What does whoosh or an opening door really sound like? Is one person’s whoosh the same as another person’s whoosh? What can I do, as a designer, if I want to continuously control the sound of a door opening? A sample from a sound effect collection is useless in this case. This article aims to shed some light on how psychologists, computer scientists, acousticians, and engineers can work together and address these and other questions arising in sound design for interactive multimedia systems.

**Sounding Object**

It’s difficult to generate sounds from scratch with a given perceived identity using signal-based models, such as frequency modulation or additive synthesis. However, physically based models offer a viable way to get naturally behaving sounds from computational structures that can easily interact with the environment. The main problem with physical models is their reduced degree of generality—that is, models are usually developed to provide a faithful simulation of a given physical system, typically a musical instrument. Yet, we can build models that retain direct physical meaning and are at the same time configurable in terms of physical properties such as shape, texture, kind of excitation, and so on. We used this approach in the Sounding Object project that the European Commission funded to study new auditory interfaces for the Disappearing Computer initiative (http://www.disappearing-computer.net).

Based on our experience with the Sounding Object project, and other recent achievements in sound science and technology, we’ll explain how

- the sound model design should rely on perception to find the ecologically relevant auditory phenomena and how psychophysics can help in organizing access to their parameters;
- physics-based models can be “cartoonified” to increase both computational efficiency and perceptual sharpness;
- the physics-related variables have to be varied in time and organized in patterns to convey expressive information (this is the problem of control); and
- we can construct interactive systems around sound and control models.

We’ll illustrate each of these points with one example. You can download the details, software implementations, and sound excerpts of each example from http://www.soundobject.org.

**Perception for sound modeling**

Humans sense the physical world and form mental images of its objects, events, and processes. We perceive physical quantities according to nonlinear, interrelated scales. For instance, humans perceive both frequency and intensity of sine waves according to nonlinear scales (called, respectively, *mel* and *son*), and intensity perception is frequency dependent. The discipline of psychoacoustics has tried, for more than a century, to understand how sound perception
Previous Research

The introduction of sound models in interfaces for effective human–computer interaction dates back to the late 1980s, when William Gaver developed the SonicFinder for Apple’s MacIntosh, thus giving an effective demonstration of auditory icons as a tool for information and system status display. The SonicFinder extended Apple’s file management application finder using auditory icons with some parametric control. The strength of the SonicFinder was that it reinforced the desktop metaphor, creating an illusion that the system’s components were tangible objects that you could directly manipulate. Apple’s later versions of MacOS (8.5 onward) implemented appearance settings—which control the general look and feel of the desktop metaphor, including sound—based on the SonicFinder.

During the last 10 years, researchers have actively debated issues such as auditory icons, sonification, and sound design while the International Community for Auditory Display (http://www.icad.org) has emerged as an international forum, with its own annual conference. In addition, several other conferences in multimedia and computer–human interaction have hosted an increasing number of contributions about auditory interfaces. Researchers have developed several important auditory-display-based applications that range from computer-assisted surgery to continuous monitoring of complex systems to analysis of massive scientific data. However, the relevant mass of experiments in sound and multimedia have revealed a substantial lack of methods for designing meaningful, engaging, and controllable sounds.

Researchers have attempted to fill this gap from two opposite directions. Some researchers have tried to understand and exploit specific sound phenomena. For example, Gaver studied and modeled the sound of pouring liquids. Others have constructed general and widely applicable sound information spaces. For instance, Barrass proposed an auditory information space analogous to the color information spaces based on perceptual scales and features.

The need for sounds that can convey information about the environment yet be expressive and aesthetically interesting, led to our proposal of sound spaces constructed on dynamic sound models. These sound spaces build on synthesis and processing models that respond continuously to user or system control signals. We proposed that the architecture for such a sound space would be based on perceptual findings and would use the sound modeling technique that’s most appropriate for a given task. Sound events have both identity and quality aspects, so physical models are appropriate for representing a sound’s identity and signal-based models are more appropriate for adjusting a sound’s quality. In fact, the nature of a sound’s physical structure and mechanisms (such as a struck bar, a rubbed membrane, and so on) determines a sound source’s identity while the specific instances of physical quantities (for example, metal bars are brighter than wood bars) determine its qualitative attributes.

References

auditory substitute of a visual progress bar. The advantages for this application are evident, as the user may continuously monitor background activities without being distracted from the foreground work. We used 11 recordings of 0.5- and 1-liter plastic bottles being filled or emptied with water. The sound files’ lengths ranged from 5.4 to 21.1 seconds. The participants used head phones to listen to the sounds. They were instructed to respond using the “0” and “1” keys on their keyboards. We divided the experiment into three sections to detect

- the action of filling or emptying,
- whether the bottle was half full or empty, and
- whether the bottle was almost full.

We randomized the order between sections and between sound files. When we asked the participants to respond if the sound was filling or emptying, 91.8 percent responded correctly for emptying sounds and 76.4 percent for filling sounds. In the section where they responded to when the bottle was half full or empty, responses during filling had a mean of 0.40 (range normalized to 1) with a standard deviation of 0.13. During emptying responses had a mean of 0.59 with a standard deviation of 0.18. In the section where users responded to whether the bottle was almost full (just about to overflow) or almost empty, the mean value for filling sounds was 0.68 with a standard deviation of 0.18, and for emptying sounds the mean was 0.78 with a standard deviation of 0.18. Based on these results, we envisage successfully using this kind of sound, for example, as an auditory progress bar, because users can distinguish between full or empty and close to completion. From informal studies, we’ve noted that users also can differentiate between the bottle sizes, pouring rate, and viscosity.

Perception is also important at a higher level, when sounds are concatenated and arranged into patterns that can have expressive content. Humans have excellent capabilities of deducing expressive features from simple moving patterns. For instance, the appropriate kinematics applied to dots can provide a sense of effort, or give the impression that one dot is pushing another. Similarly, temporal sequences of sounds can convey expressive information if properly controlled.

**Cartoon sound models**

In information visualization and human–computer visual interfaces, photorealistic rendering is often less effective than nicely designed cartoons. Stylized pictures and animations are key components of complex visual displays, especially when communication relies on metaphors. Similarly, auditory displays may benefit from sonic cartoons— that is, simplified descriptions of sounding phenomena with exaggerated features. We often prefer visual and auditory cartoons over realistic images and sounds because they ease our understanding of key phenomena by simplifying the representation and exaggerating the most salient traits. This can lead to a quicker understanding of the intended message and the possibility of detailed control over the expressive content of the picture or sound.

Sound design practices based on the understanding of auditory perception might try to use physics-based models for sound generation and signal-based models for sound transformation. In this way it’s easier to impose a given identity to objects and interactions (such as the sound of impact between two ceramic plates). Some quality aspects, such as the apparent size and distance of objects, may be adjusted through time-frequency manipulations when they’re not accessible directly from a physical model.

**Physics-based models**

The sound synthesis and computer graphics communities have increasingly used physical models whenever the goal is a natural dynamic behavior. A side benefit is the possibility of connecting the model control variables directly to sensors and actuators since the physical quantities are directly accessible in the model. We can use several degrees of accuracy in developing a physics-based sound model. Using a finite-element model of interacting objects would make sense if tight synchronization with realistic graphic rendering is required, especially in the cases of fractures of solids or fluid-dynamic phenomena. Most often, good sounds can be obtained by simulating the rigid-body dynamics of objects described by piecewise parametric surfaces. However, even in this case each impact results in a matrix of ordinary differential equations that must be solved numerically. If the display is mainly auditory, or if the visual object dynamics can be rendered only poorly (for example, in portable low-power devices), we can rely on perception to simplify the models signifi-
cantly, without losing either their physical interpretability or dynamic sound behavior.

As an example, consider the model of a bouncing object such as a ball. On one extreme, we can try to develop a finite-element model of the air cavity, its enclosure, and the dynamics of impact subject to gravity. On the other hand, if we have a reliable model of nonlinear impact between a striking object and a simple mechanical resonator, we can model the system as a point-like source displaying a series of resonant modes. The source can be subject to gravity, thus reproducing a natural bouncing pattern, and we can tune the modes to those of a sphere, thus increasing the impression of having a bouncing ball.

Of course, this simplification based on lumping distributed systems into point-like objects introduces inevitable losses in quality. For instance, even though we can turn the ball into a cube by moving the modal resonances to the appropriate places, the effect wouldn’t resemble that of a bouncing cube just because the temporal pattern followed by a point-like bouncing object doesn’t match that of a bouncing cube. However, we can introduce some controlled randomness in the bouncing pattern in such a way that the effect becomes perceptually consistent. Again, if the visual display can afford the same simplifications, the overall simulated phenomenon will be perceived as dynamic.¹²

The immediate benefit of simplified physics-based models is that they can run in real time in low-cost computers, and gestural or graphical interfaces can interactively control the models. Figure 1 shows the graphical Pure Data patch of a bouncing object, where sliders can control the size, elasticity, mass, and shape. (For more information about Pure Data, see http://www.pure-data.org.) In particular, the shape slider allows continuous morphing between sphere and cube via superellipsoids and it controls the position of resonances and the statistical deviation from the regular temporal pattern of a bouncing ball.

**Drinking lemonade with a straw**

In physical sciences it’s customary to simplify physical phenomena for better understanding of first principles and to eliminate second-order effects. In computer-based communication, when the goal is to improve the clarity and effectiveness of information, the same approach turns out to be useful, especially when it’s accompanied by exaggeration of selected features. This process is called *cartoonification*.⁵

Consider the cartoon in Figure 2, displaying a mouse-like character drinking lemonade with a straw. The illustration comes from a children’s book, where children can move a carton flap
that changes the level of liquid in the glass. With the technology that’s now available, it wouldn’t be too difficult to augment the book with a small digital signal processor connected to sensors and to a small actuator, so that the action on the flap would also trigger a sound. Although many toys feature this, the sounds are almost inevitably prerecorded samples that, when played from low-cost actuators, sound irritating. Even in experimental electronically augmented books with continuous sensors, due to the intrinsic limitations of sampled sounds, control is usually limited to volume or playback speed.

The interaction would be much more effective if the sound of a glass being emptied is a side effect of a real-time simulation that responds continuously to the actions. Instead of trying to solve complex fluid-dynamic problems, we use a high-level analysis and synthesis of the physical phenomena as follows:

1. take one micro event that represents a small collapsing bubble;
2. arrange a statistical distribution of micro events;
3. filter the sound of step 2 through a mild resonant filter, representing the first resonance of the air cavity separating the liquid surface from the glass edge;
4. filter the sound of step 3 with a sharp comb filter (harmonic series of resonances obtained by a feedback delay line), representing the filtering effect of the straw; and
5. add abrupt noise bursts to mark the initial and final transients when the straw rapidly passes from being filled with air to being filled with liquid, and vice versa.

The example of the straw is extreme in the context of cartoon sound models, as in reality this physical phenomenon is usually almost silent. However, the sound model serves the purpose of signaling an activity and monitoring its progress. Moreover, because we built the model based on the actual physical process, the model integrates perfectly with the visual image and follows the user gestures seamlessly, thus resulting in far less irritating sounds.

On the importance of control

Sound synthesis techniques have achieved remarkable results in reproducing musical and everyday sounds. Unfortunately, most of these techniques focus only on the perfect synthesis of isolated sounds, thus neglecting the fact that most of the expressive content of sound messages comes from the appropriate articulation of sound event sequences. Depending on the sound synthesis technique, we must design a system to generate control functions for the synthesis engine in such a way that we can arrange single events in naturally sounding sequences.

Sound control can be more straightforward if we generate sounds with physics-based techniques that give access to control parameters directly connected to sound source characteristics.

Control models

Recently, researchers have studied the relationship between music performance and body motion. Musicians use their body in a variety of ways to produce sound. Pianists use shoulders, arms, hands, and feet; trumpet players use their lungs and lips; and singers use their vocal chords, breathing system, phonatory system, and expressive body postures to render their interpretation. When playing an interleaved accent in drumming, percussionists prepare for the accentuated
stroke by raising the drumstick up higher, thus arriving at the striking point with larger velocity.

The expressiveness of sound rendering is largely conveyed by subtle but appropriate variations of the control parameters that result from a complex mixture of intentional acts and constraints of the human body’s dynamics. Thirty years of research on music performance at the Royal Institute of Technology (KTH) in Stockholm has resulted in about 30 so-called performance rules. These rules allow reproduction and simulation of different aspects of the expressive rendering of a music score. Researchers have demonstrated that they can combine rules and set them up in such a way that generates emotionally different renderings of the same piece of music. The results from experiments with expressive rendering showed that in music performance, emotional coloring corresponds to an enhanced musical structure. We can say the same thing about hyper- and hypo-articulation in speech—the quality and quantity of vowels and consonants vary with the speaker’s emotional state or the intended emotional communication. Yet, the phrase structure and meaning of the speech remain unchanged. In particular, we can render emotions in music and speech by controlling only a few acoustic cues. Therefore, we can produce cartoon sounds by simplifying physics-based models and controlling their parameters.

**Walking and running**

Music is essentially a temporal organization of sound events along short and long time scales. Accordingly, microlevel and macrolevel rules span different time lengths. Examples of the first class of rules include the Score Legato Articulation rule, which realizes the acoustical overlap between adjacent notes marked legato in the score, and the Score Staccato Articulation rule, which renders notes marked staccato in the score. Final Retard is a macrolevel rule that realizes the final ritardando typical in Baroque music. Friberg and Sundberg demonstrated how their model of final ritardando was derived from measurements of stopping runners. Recently, researchers discovered analogies in timing between walking and legato and running and staccato. These findings show the interrelationship between human locomotion and music performance in terms of tempo control and timing.

Friberg et al. recently studied the association of music with motion. They transferred measurements of the ground reaction force by the foot during different gaits to sound by using the vertical force curve as sound-level envelopes for tones played at different tempos. Results from listening tests were consistent and indicated that each tone (corresponding to a particular gait) could clearly be categorized in terms of motion. These analogies between locomotion and music performance open new possibilities for designing control models for artificial walking sound patterns and for sound control models based on locomotion.

We used the control model for humanized walking to control the timing of the sound of one step of a person walking on gravel. We used the Score Legato Articulation and Phrase Arch rules to control the timing of sound samples. We’ve observed that in walking there’s an overlap time between any two adjacent footsteps (see Figure 3) and that the tendency in overlap time is the same as observed between adjacent notes in piano playing: the overlap time increases with the time interval between the two events. This justifies using the Score Legato Articulation rule for walking. The Phrase Arch rule used in music performance renders accelerandi and rallentandi, which are a temporary increase or decrease of the beat rate. This rule is modeled according to velocity changes in hand movements between two fixed points on a plane. We thought that the
Phrase Arch rule would help us control walking tempo changes on a larger time scale.

Human running resembles staccato articulation in piano performance, as the flight time in running (visible as a vertical white strip in the spectrograms in Figure 3) can play the same role as the key-detach time in piano playing. We implemented the control model for stopping runners by applying the Final Retard rule to the tempo changes of sequences of running steps on gravel.

We implemented a model for humanized walking and one for stopping runners as Pure Data patches. Both patches allow controlling the tempo and timing of sequences of simple sound events. We conducted a listening test comparing step sound sequences without control to sequences rendered by the control models presented here. The results showed that subjects preferred the rendered sequences and labeled them as more natural, and they correctly classified different types of motion produced by the models. Recently, we applied these control models to physics-based sound models (see the “Web Extras” sidebar for a description of these sound examples available at http://computer.org/multimedia/mu2003/u2toc.htm).

The proposed rule-based approach for sound control is only a step toward the design of more general control models that respond to physical gestures. In our research consortium, we’re applying the results of studies on the expressive gestures of percussionists to impact-sound models and the observations on the expressive character of friction phenomena to friction sound models. Impacts and frictions deserve special attention because they’re ubiquitous in everyday soundscapes and likely to play a central role in sound-based human–computer interfaces.

### New interfaces with sound

Because sound can enhance interaction, we need to explore ways of creating and testing suitable auditory metaphors. One of the problems with sounds used in auditory interfaces today is that they always sound the same, so we need to have parametric control. Then, small events can make small sounds; big events make big sounds; the effort of an action can be heard; and so on. As Truax pointed out, before the development of sound equipment in the 20th century, nobody had ever heard exactly the same sound twice. He also noted that fixed waveforms—as used in simple synthesis algorithms—sound unnatural, lifeless, and annoying. Hence, with parametric control and real-time synthesis of sounds, we can get rich continuous auditory representation in direct manipulation interfaces. For instance, a simple audiovisual animation in the spirit of Michotte’s famous experiments (see the geometric forms animation in the “Web Extras” sidebar) shows how sound can elicit a sense of effort and enhance the perceived causality of two moving objects.
Design

Barrass\textsuperscript{21} proposed a rigorous approach to sound design for multimedia applications through his Timbre-Brightness-Pitch Information Sound Space (TBP ISS). Similar to the Hue-Saturation-Lightness model used for color selection, sounds are arranged in a cylinder, where the radial and longitudinal dimensions are bound to brightness and pitch, respectively. The timbral (or identity) dimension of the model is restricted to a small collection of musical sound samples uniformly distributed along a circle derived from experiments in timbre categorization. A pitfall of this model is that it’s not clear how to enrich the palette of timbres, especially for everyday sounds. Recent investigations found that perceived sounds tend to cluster based on shared physical properties.\textsuperscript{22} This proves beneficial for structuring sound information spaces because there are few sound-producing fundamental mechanisms as compared to all possible sounding objects. For instance, a large variety of sounds (bouncing, breaking, scraping, and so on) can be based on the same basic impact model.\textsuperscript{22} Bezzi, DePoli, and Rocchesso,\textsuperscript{8} proposed the following three-layer sound design architecture:

- **Identity**: A set of physics-based blocks connected in prescribed ways to give rise to a large variety of sound generators.
- **Quality**: A set of signal-processing devices (digital audio effects\textsuperscript{19}) specifically designed to modify the sound quality.
- **Spatial organization**: Algorithms for reverberation, spatialization, and changes in sound-source position and size.\textsuperscript{9}

Such a layered architecture could help design a future sound authoring tool, with the possible addition of tools for spatio-temporal texturing and expressive parametric control.

Organization and access

The complexity of the sonic palette calls for new methods for browsing, clustering, and visualizing dynamic sound models. The Sonic Browser (see Figure 4) uses the human hearing system’s streaming capabilities to speed up the browsing in large databases. Brazil et al.\textsuperscript{23} initially developed it as a tool that allowed interactive visualization and direct sonification. In the starfield display in Figure 4, each visual object represents a sound. Users can arbitrarily map the location, color, size, and geometry of each visual object to properties of the represented objects. They can hear all sounds covered by an aura simultaneously, spatialized in a stereo space around the cursor (the aura’s center). With the Sonic Browser, users can find target sounds up to 28 percent faster, using multiple stream audio, than with single stream audio.

In the context of our current research, we use the Sonic Browser for two different purposes. First, it lets us validate our sound models, as we can ask users to sort sounds by moving visual representations according to auditory perceptual dimensions on screen. If the models align with real sounds, we consider the models valid. Second, the Sonic Browser helps users compose auditory interfaces by letting them interactively access large sets of sounds—that is, to pick, group, and choose what sounds might work together in an auditory interface.

Manipulation and control

Real-time dynamic sound models with parametric control, while powerful tools in the hands of the interface designer, create important problems if direct manipulation is the principal control strategy. Namely, traditional interface devices (mouse, keyboards, joysticks, and so on) can only exploit a fraction of the richness that
emerges from directly manipulating sounding objects. Therefore, we need to explore alternative devices to increase the naturalness and effectiveness of sound manipulation.

The Vodhran

One of the sound manipulation interfaces that we designed in the Sounding Object project is based on a traditional Irish percussion instrument called the bodhran. It’s a frame drum played with a double-sided wood drumstick in hand A while hand B, on the other side of the drum, damps the drumhead to emphasize different modes. This simple instrument allows a wide range of expression because of the richness in human gesture. Either the finger or palm of hand B can perform the damping action, or increase the tension of the drumhead, to change the instrument’s pitch. Users can beat the drumstick on different locations of the drumhead, thus generating different resonance modes. The drumstick’s macrotemporal behavior is normally expressed in varying combinations of duplets or triplets, with different accentuation. We tuned our impact model to reproduce the timbral characteristics of the bodhran and give access to all its controlling parameters such as resonance frequency, damping, mass of the drumstick, and impact velocity.

To implement a virtual bodhran, the Vodhran, we used three devices that differ in control and interaction possibilities—a drumpad, a radio-based controller, and a magnetic tracker system (see Figure 5). Each device connects to a computer running Pure Data with the impact model used in Figure 1.

Drumpad controller. For the first controlling device, we tested a Clavia ddrum4 drumpad (http://www.clavia.se/ddrum/), which reproduces sampled sounds. We used it as a controller to feed striking velocity and damp values into the physical model. The ddrum4 is a nice interface to play the model because of its tactile feedback and the lack of cables for the drumsticks.

Radio-based controller. We used Max Mathew’s Radio Baton24 to enhance the Vodhran’s control capabilities. The Radio Baton is a control device comprised of a receiving unit and an antenna that detects the 3D position of one or two sticks in the space over it. Each of the sticks sends a radio signal. For the Vodhran, we converted the two sticks into two radio transmitters at each end of a bodhran drumstick, and played the antenna with the drumstick as a normal bodhran. The drumstick’s position relative to the antenna controlled the physical model’s impact position, thus allowing a real-time control of the instrument’s timbral characteristics. Professional player Sandra Joyce played this version of the bodhran in a live concert. She found that it was easy to use and that it had new expressive possibilities compared to the traditional acoustical instrument.

Tracker-based controller. With the Polhemus Fastrack, we can attach the sensors to users’ hands or objects handled by them, so that they can directly manipulate tangible objects of any kind with the bodhran gestural metaphor. During several design sessions, leading up to the public performance in June 2002, we evaluated numerous sensor configurations. With a virtual instrument, we could make the instrument bodycentric or geocentric. A real bodhran is bodycentric because of the way it’s held. The configurations were quite different. We felt a bodycentric reference was easier to play,
although the geocentric configuration yielded a wider range of expression. In the Vodhran’s final public-performance configuration, we attached one Polhemus sensor to a bodhran drumstick (held in the right hand), and the player held the second sensor in her left hand. We chose the geocentric configuration and placed a visual reference point on the floor in front of the player. We used normal bodhran playing gestures with the drumstick in the player’s right hand to excite the sound model. The distance between the hands controlled the virtual drumhead’s tension, and the angle of the left hand controlled the damping. Virtual impacts from the player’s hand gestures were detected in a vertical plane extending in front of the player. We calibrated this vertical plane so that if the player made a virtual impact gesture with a fully extended arm or with the hand close to the player’s body, the sound model’s membrane was excited near its rim. If the player made a virtual impact gesture with the arm half-extended, the sound model was excited at the center of the membrane.

**Conclusion**

An aesthetic mismatch exists between the rich, complex, and informative soundscapes in which mammals have evolved and the poor and annoying sounds of contemporary life in today’s information society. As computers with multi-media capabilities are becoming ubiquitous and embedded in everyday objects, it’s time to consider how we should design auditory displays to improve our quality of life.

Physics-based sound models might well provide the basic blocks for sound generation, as they exhibit natural and realistically varying dynamics. Much work remains to devise efficient and effective models for generating and controlling relevant sonic phenomena. Computer scientists and acousticians have to collaborate with experimental psychologists to understand what phenomena are relevant and to evaluate the models’ effectiveness. Our work on Sounding Objects initiated such a long-term research agenda and provided an initial core of usable models and techniques.

While continuing our basic research efforts, we are working to exploit the Sounding Objects in several applications where expressive sound communication may play a key role, especially in those contexts where visual displays are problematic.

**Acknowledgments**

The European Commission’s Future and Emergent Technologies collaborative R&D program under contract IST-2000-25287 supported this work. We’re grateful to all researchers involved in the Sounding Object project. For the scope of this article, we’d like to acknowledge the work of Matthias Rath, Federico Fontana, and Federico Avanzini for sound modeling; Eoin Brazil for the Sonic Browser; Mark Marshall and Sofia Dahl for the Vodhran. We thank Clavia AB for lending us a ddrum4 system and Max Mathews for providing the Radio Baton.

**References**


For further information on this or any other computing topic, please visit our Digital Library at http://computer.org/publications/dlib.

Davide Rocchesso is an associate professor in the Computer Science Department at the University of Verona, Italy. His research interests include sound processing, physical modeling, and human–computer interaction. Rocchesso received a PhD in computer engineering from the University of Padova, Italy.

Roberto Bresin is a research assistant at the Royal Institute of Technology (KTH), Stockholm. His main research interests are in automatic music performance, affective computing, sound control, visual feedback, music perception, and music education. Bresin received an MSc in electrical engineering from the University of Padova, Italy, and a PhD in music acoustics from KTH.

Mikael Fernström is director of the master’s program in interactive media at the University of Limerick, Ireland. His main research interests include interaction design, sound, music, and electronics. Fernström studied electronic engineering and telecommunications at the Kattegatt technical college in Halmstad, Sweden, and has an MSc in human–computer interaction from the University of Limerick.

Readers may contact Davide Rocchesso at the Computer Science Dept., Univ. of Verona, Strada Le Grazie 15, 37134 Verona, Italy; Davide.Rocchesso@univr.it.
Paper IV
We examine the use of auditory display for ubiquitous computing to extend the boundaries of human–computer interaction (HCI). Our design process is based on listening tests, gathering free-text identification responses from participants. The responses and their classifications indicate how accurately sounds are identified and help us identify possible metaphors and mappings of sound to human action and/or system status.

With the emergence of ubiquitous and wearable computers, we need to explore alternatives to visual displays, as users sometimes need to focus their visual attention on the surrounding environment rather than on the computer. For example, using a GUI on a handheld computer while walking is extremely difficult, and indeed dangerous when standing on scaffolding on a building site high above ground.

On the other hand, we can use other kinds of equipment in such situations, such as walkie-talkies, mobile phones, and various forms of electronic instruments (for example, Geiger counters, metal detectors, and personal entertainment systems). Such devices usually have a few fixed buttons or controls (which can be attached to the user’s clothing) that the user can operate with fingers. The user then learns to use the device over time with practice.

In this article, we report on the auditory displays that we devised and tested with users. We focus on two issues that we feel are important when designing interaction with auditory display:

- A method to inform the choice of suitable sounds for auditory icons, based on the analysis of listening tests.
- The design of soft buttons using auditory display to provide users with a pseudohaptic experience.

For background information on other researchers who are exploring sound, as well as basic information on some of the issues we’ve considered, please see the sidebar, “Considerations for Designing Auditory Displays.”

Exploring what people hear

While it’s assumed that everyday sounds have inherent meaning, learned from our everyday activities, hearing such sounds in isolation without context can be quite confusing. The sound of a single isolated footstep can, for example, be heard as a book being dropped on a table. Interestingly, this problem is somewhat similar to how linguistic homonyms work (words of the same spelling or sound can have different meanings depending on context).1

To further develop our understanding of people’s perception of auditory events, we conducted listening tests, an approach also used by other researchers.2,3 We made high-quality (44.1-kilohertz, 16-bit) recordings of 104 everyday sounds (durations between 0.4 and 18.2 seconds) and had 14 postgraduate students listen to the recorded sounds in random order using headphones, responding in free-text format to what each sound was. In most cases the descriptions they gave were quite rich. For example, the following responses (for three different recordings) describe the events quite accurately:

- “A person walking on a carpet with their hands in their pockets hence the clanging of keys or coins, taking five steps and turning to retrace their footsteps.”
- “A metal spoon in stirring motion in an empty ceramic cup, tapping the cup as if to displace the liquid from the spoon and then placing the spoon onto a table.”
- “Breaking of a cup (not really a glass sound more ceramic I think).”

Several ways exist to analyze the responses from such listening tests. The most obvious way would be to count how many responses could be deemed correct for each sound by linking the sound to the participants’ reported understanding of the objects and actions involved in producing the sound. A somewhat more interesting measure is Ballas’ method of causal uncertainty.4 Ballas et al.5 found that identifi-
Considerations for Designing Auditory Displays

In novel human–computer interaction (HCI) paradigms, such as ubiquitous, pervasive, wearable, and disappearing computing, interactive sonification might offer useful alternatives to the otherwise dominant visual displays, freeing up our eyes to see the surrounding world or do what small visual displays don’t do so well. In brief, using sound for display in interaction design is useful for attracting attention to events or locations, for non-visual communication in general (including speech), alarms, notification, and feedback. Sound is less useful for continuous display of objects, for absolute readings (most people perceive auditory dimensions such as pitch, loudness, and timbre as being relative), and for fine-detail spatial display. Sound is also problematic in noisy or noise-sensitive environments.

Designing interactive sonifications for HCI requires that we address numerous issues. We have to consider where and how sonification is appropriate. As designers, we also need to take into account the users’ capabilities while carrying out tasks in real environments, and consider that surrounding noise levels might mask the system’s sounds. If the sound will enhance the interaction, we need to explore ways of creating and testing auditory metaphors. We should also investigate to what extent the use of sound contributes to the users’ performance and subjective quality of use.

To be able to design with sound, we need a high-level understanding of what and how we hear (for example, see Gaver’s research4-5). While an extensive catalogue of studies exists on the perception of musical sounds and speech, researchers know relatively little about other kinds of nonspeech sounds—in particular everyday sounds (such as footsteps, creaking doors, water filling a container, bouncing, and breaking).

In the new HCI paradigms, we can explore new concepts of interaction and human activity. In previous work on auditory interfaces, ranging from Gaver’s Sonic Finder6 to Brewster’s hierarchical earcons,7 human action has to a large extent been thought of in a discrete way—like kicking a football, where a user action starts a process that then completes without any further user control. This view might be appropriate when typing, clicking, or flicking switches. An alternative view is action as a continuous flow, such as a pen stroke, where we continuously move a pencil on a surface, relying on our learned gesture through proprioception, as well as haptic, visual, and auditory feedback. This latter view is becoming important, now that several input devices (such as pens, digitizers, and cameras) are capable of detecting complex human actions.

Still, at the core of our design space, a fundamental problem is how to classify and select suitable sounds for a particular interaction design. Depending on our intended users, tasks, and context, initially a broad continuum exists in this design space, ranging from concrete to abstract displays (that is, from auditory icons to earcons).89 If we’re designing for casual everyday use, we probably need to consider concrete forms. If we’re designing for highly specialized domains (such as cockpit or process-control applications) where our users will be selected and trained for high-performance requirements, we might need to focus on psychoacoustic issues such as detection accuracy or time and perceived urgency. In the latter case the design space can be more abstract.10

References

1. E. Bergman, Information Appliances and Beyond, Morgan Kaufmann, 2000, p. 385.

\[
H_{CU} = \sum_{i} p_i \log_2 p_i
\]

(1)

\(H_{CU}\) is a measure of causal uncertainty for sound \(i\), \(p_i\) is the proportion of all responses for sound \(i\).
Being able to parametrically control sound models in real time can also, potentially, help make sonifications less annoying.

sorted into event category \( j \), and \( n \) is the number of categories for responses to sound \( i \). Applying this equation implies that if all participants in a listening test give the same response, the causal uncertainty is 0 (all participants agree). For example, with 14 participants if the responses are distributed 50/50 between two alternatives, the causal uncertainty is 1.0. If the distribution of responses is skewed—such as 13 of the 14 responses are the same but one response is different—the causal uncertainty is 0.37. If all 14 responses are different, the causal uncertainty is 3.8. From this we can see that calculating causal uncertainty according to Ballas’ method gives a good measure of how easy it is for users to identify everyday sounds.

With our collected data (responses from 14 participants listening to 104 different sounds) the responses were sorted and categorized, as well as evaluated for correctness, by two of the authors and a research assistant. The reliability between the evaluators was significant (weakest \( r = 0.78, p < 0.0016 \)). From the responses, we extracted and categorized action and object segments of the texts, such as how the objects/materials interacted and what objects/materials were used. We found that in general 32 percent of the sounds were identified correctly, while for action segments it was 38 percent and for object segments it was 25 percent.

The collected data set with all responses, categorizations, and measurements of causal uncertainty can also be used for suggesting the possible use of sounds in interaction design, somewhat similar to Barrass’ method of collecting stories about when sound is useful in everyday life.\(^6\) From a designer’s point of view it’s interesting to note that the responses from the listening tests contain information about how people describe everyday sounds as well as measurements of causal uncertainty.

**Sounding objects**

With the results of the listening tests, we can begin to suggest possible auditory displays and metaphors for interaction design. Based on Barrass’ TaDa approach,\(^6\) we can do a task and data analysis that lets us select sounds that can communicate the dimensions and directions that give users adequate feedback about their actions in relation to the system as well as the system’s status and events. We then need to create ways so that the system can produce the selected sounds and finally evaluate the resulting design with users.\(^7\) If we were to just play sound files as feedback to user actions, it would always sound the same and never (or seldom) be expressive (for example, to be mapped to the user’s effort or the size of the data objects involved).

This was one of the issues addressed by the European Union’s Sounding Object project (see http://www.soundobject.org), where new methods for physically inspired modeling of sounds for sound synthesis were explored. Our work was initially largely informed by ecological acoustics and Gaver’s work on auditory icons.\(^8\)

We also worked toward **cartoonification** of sound models—that is, simplifying the models while retaining perceptual invariants. We implemented the models in Pure Data (commonly known as PD; see http://www.puredata.org) and tested them in a number of ways, ranging from perceptual experiments to artistic performance.

Compared to ordinary sound files, sound objects can provide “live” sound models that we can parametrically control in real time with reasonable computational power.

Being able to parametrically control sound models in real time can also, potentially, help make sonifications less annoying. With prerecorded sound files, sounds used in an auditory interface always sound exactly the same. In contrast, with sound objects and parametric control we can vary properties of the sounds—for example, mapping the size of objects or the effort of actions—so that small objects or actions make small sounds and large objects or actions make large sounds.

Revisiting the overall results from the Sounding Object project,\(^9\) it’s interesting to note that all the sound models developed throughout the project point toward an epistemology that differs from Gaver’s trichotomy of primitives of solids, liquids, and gases. An alternative emerging view indicates that the primitive classes might be better understood if we think of the world of sound-producing events as composed of impacts, frictions, and
deformations. The simulated material properties of the objects involved in such interactions are controllable through parameters passed on to our sound models. The analysis of listening tests, as previously described, also suggests that actions are better identified than objects. This might suggest that interaction design using auditory display should focus on mapping human activity to actions rather than objects.

**Example: Auditory soft buttons**

The idea of software-defined buttons—soft buttons—emerged from research on direct manipulation and GUIs (from early work in Xerox PARC—Palo Alto Research Center10) and is now an important part of all GUI widget libraries. Most GUIs use soft buttons extensively, ranging from desktop personal computers and laptops (see, for example, Figure 1), to personal digital assistants (see Figure 2). With soft buttons the designer—and sometimes also the user—can easily modify, add, or remove a software application’s interactive controls.

The ways that users can activate soft buttons vary. On desktop computers the most common way is to move a pointing device, such as a mouse, that in turn indirectly moves a visible cursor on screen into the rectangle surrounding the soft button. The user then activates it by clicking with the pointing device. Visual soft buttons are often animated to improve feedback to the user—for example, when the user clicks, the visual button displayed temporarily changes its appearance so that the graphical symbol looks like it’s moving inwards, into the display surface. On other kinds of computers, such as handheld computers, the user can point directly to a visual soft button, either with a handheld stylus or simply with a finger.

User interface widgets such as these make the design of GUIs highly malleable and flexible, as the designer can display and represent highly complex underlying functionality with simple graphical symbols that, ideally, look like concepts or entities in the user’s task domain. Because it’s usually the software rather than the hardware that defines such widgets, the same physical display surface can be used for different widgets at different times, supporting the varying needs for the user to carry out different tasks. These features make soft buttons attractive components of user interfaces, both for designers and users. Along these lines, our interest is in using forms of display other than vision to create similar affordances, in this particular case through auditory feedback mimicking what it would sound like to touch differently structured surfaces.

**Pseudohaptic soft buttons using auditory display**

In three experiments we investigated the use of auditory display to create a pseudohaptic experience of soft buttons.

**Pilot 1: Real haptics**

First, we conducted a pilot experiment, based on Gibson’s11 cookie-cutter study. With four different paper shapes glued on paper, a participant found it easy to feel the shapes and then draw an image of them, picking up the shape of the objects through haptic perception and visualizing the shapes.

**Pilot 2: Pseudohaptics using auditory display**

Based on the first pilot experiment, we designed a second study of a soft-button proto-
We tested the idea of having soft buttons displayed by audio instead of graphics. We asked three users to make drawings of three different soft-button layouts. We used a wearable computer and touch-sensitive display subsystem from Xybernaut (http://www.xybernaut.de; see Figure 3), which is normally worn on the arm. In this experiment, however, we affixed the subsystem to the user’s belt in a position so that the users could comfortably rest their hand on the unit, with fingers free to access the touch area (see Figure 4). The size of the active touch area was 120 × 90 mm. We only used the touch detection of the device, not the visual display.

We created three different layouts with soft buttons (see Figure 5a). When a user moved a finger on a button area, a simple friction-like sound was produced. To emphasize the boundaries of each button, click sounds were produced when entering or exiting a button area (see Table 1 for the mapping between actions and sounds). The sounds were heard in mono, using a simple headphone in one ear.

We recruited three participants. Each participant spent approximately 10 minutes getting familiar with the design and making three drawings. We found that the participants were able to feel their way around the touch device and make quite accurate drawings of the soft-button layout, as we show in Figure 5b. (This example is from one user, although all three made similar drawings.) This indicated that this kind of auditory display of soft buttons lets users have a pseudo-

Table 1: Mapping between actions and sound.

<table>
<thead>
<tr>
<th>Action</th>
<th>Sound</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>No touch</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Touch area outside button</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Enter button area</td>
<td>Tick</td>
<td>N/A</td>
</tr>
<tr>
<td>Move finger on button</td>
<td>Friction sound</td>
<td>N/A</td>
</tr>
<tr>
<td>Exit button area</td>
<td>Tack</td>
<td>N/A</td>
</tr>
<tr>
<td>Lift finger off button</td>
<td>Tock</td>
<td>Select/activate function</td>
</tr>
</tbody>
</table>
haptic experience, giving them a mental model of the device’s layout.

Pilot 3: Haptics and pseudohaptics

The focus in our final experiment was the user detection of and interaction with soft buttons using an auditory display. We refined our experiment by first selecting a different touch device—a Tactex touch tablet called MTC Express (see Figure 6)—that differs from the previously used Xybernaut device in that the MTC Express device doesn’t have a visual display.

The interactive area of the Tactex device matches the size of a human hand quite well. The active touch area is 145 × 95 mm. The point of contact where the hand rests is a spatial and haptic reference point for finger movements on the device. We also redesigned our software to minimize any latency and implemented six different soft-button layouts (see Figure 7).

Procedure. We recruited 10 participants among our postgraduates. We tested all six layouts both as paper shapes on cardboard and with the Tactex touch tablet connected to a Windows PC. Each stimulus was tested twice, resulting in 24 drawings per participant (12 haptic, 12 auditory/pseudohaptic). The order between stimuli was randomized. Users had headphones to listen to the sounds (in mono) while interacting with the system. They were allowed to use either their left or right hand to explore the layouts and to draw their understanding of the layouts.

To prevent our participants from seeing where they were moving their fingers on the Tactex tablet, we covered it with a cardboard box with a cut-out for the user’s hand to reach the active touch area. On top of the box, a video camera was fitted for recording the participants’ hand and finger movements (see Figure 8). The same box was used for both haptic and pseudohaptic stimuli. The participants were given 3 minutes of explo-
ration time per stimulus. During their exploration they used a black pen to sketch, and at the end of each 3-minute period they used a red pen to mark their detected layout on a blank sheet of paper.

**Results.** Figure 9 details 12 typical examples out of 240 drawings, all of which are available at http://www.idc.ul.ie/mikael/softbuttons/. As Figure 9 shows, the participants were good at detecting the layouts, both in the haptic and pseudohaptic (using auditory display) conditions. With the more complex layouts, the number of errors increased—or on occasion they ran out of time (particularly in the pseudohaptic condition). In the debriefing sessions, participants reported the haptic and pseudohaptic conditions to be almost as easy (or difficult, for the more complex layouts).

**Discussion.** Our results indicate that using an auditory display to create a pseudohaptic experience, based on sound object models synthesized in real time, is almost as efficient and accurate as a haptic display. The participants needed more time to detect more complex layouts using auditory display, but because they didn’t know anything about the layouts in advance, we assume that if we allowed them to familiarize themselves for a longer period of time with a particular set of layouts, the difference would become smaller. The same applies to many other human activities, such as shifting gears in a car or typing. As our actions become automatic, we need fewer clues regarding the success of our actions.

Our findings indicate that this kind of auditory display of soft buttons lets users have a pseudohaptic experience that supports the development of a mental model of the device’s layout. A similar pseudohaptic approach was investigated by Müller-Tomefele, who in one of his demonstrations communicated differences in surface texture through friction-like sounds in a pen-based digitizer application. In the commercial world, Apple Computer’s Ink application for handwriting input with a digitizer tablet also attempts to enhance the user experience through friction-like sounds as feedback to the user’s pen strokes with a stylus.

**Future research**

In this article, our approach has been that designs should be based on the results from listening tests with possible metaphors being extracted from users’ descriptions of everyday sounds. The listening tests can also provide guidance in our understanding of how users interpret combinations of auditory icons.

More studies are needed on what people hear when they listen to everyday sounds to increase our understanding of the perceptual and cognitive processes involved. In particular, studies of the effects of combinations of different auditory icons in sequence or in parallel are lacking.

We’ve found that the PD environment and Sounding Objects project are both highly productive approaches for prototyping sound designs for interactive sonification. However, for fully integrated applications we need to seriously consider if we can more closely integrate a set of sonification primitives with operating systems. This can in turn result in the development of toolkits for developers, similar to what’s available for GUIs today. A need also exists to educate and support interaction designers so that they can open up their creative thinking toward interac-
tive sonification, and realize that it’s possible to provide continuous feedback in real time for gesture-based devices.

All components in HCI also have aesthetic properties. It’s probably possible to design sonifications that are psychoacoustically correct and quite efficient but unpleasant to listen to (just as it’s possible to design visual interfaces that users find unpleasant to view). As Somers\textsuperscript{13} has suggested, we need to draw upon the knowledge and ideas of Foley artists (sound design for film, radio, and television) as well as lessons learned from various theories of acousmatic music. MM

Acknowledgments

We would like to express our thanks to all the postgraduate students at the University of Limerick who participated in our experiments. We thank Xybernaut for giving us access to their wearable computers and Tactex for giving us access to their device driver source code. We’re indebted to all our colleagues in the Sounding Object project for an ongoing and engaging dialogue. Finally, we thank our colleagues Annette Aboulafia, Thomas Waldmann, and Michael Cooke for their comments on earlier versions of this article; Stephen Shirley for software development for the Tactex device; and the anonymous reviewers for their creative and encouraging comments. This work was partly financed by the European Union Future and Emerging Technologies program; the Disappearing Computer/Sounding Object project IST-2000-25287; the Irish Research Council for Science, Engineering and Technology/Multimodal Systems project SC-02-302; and the Science Foundation Ireland: Shared Worlds Project, 2003–2007.

The research in this article was originally published in the Proceedings of the International Workshop on Interactive Sonification in January 2004. The article has been revised and updated to fit with IEEE MultiMedia’s standards.

References


Mikael Fernström is the director of the master’s program in Interactive Media at the University of Limerick, Ireland. His main research interests are interaction design, sound, music, and electronics. Fernström studied electronic engineering and telecommunications at the Kattegatt technical college in Halmstad, Sweden, and has an MSc in human–computer interaction from the University of Limerick.
Eoin Brazil is a research assistant in the Interaction Design Centre at the University of Limerick, Ireland. His main research interests are interaction design, sound, software visualization, and music retrieval. Brazil has a BSc and MSc in computer science from the University of Limerick.

Liam Bannon is the director of the Interaction Design Centre and professor of computer science in the Department of Computer Science and Information Systems at the University of Limerick. He’s interested in improving the usability and desirability of computational artifacts. He also examines alternative conceptual and methodological frameworks for understanding human activities (and their mediation by various technologies). Bannon has an MSc in computer science from Trinity College in Dublin, Ireland, and a PhD in experimental psychology from the University of Western Ontario, Canada.

Readers may contact Mikael Fernström at the Interaction Design Centre, Computer Science Dept., Univ. of Limerick, Limerick, Ireland; mikael.fernstrom@ul.ie.

For further information on this or any other computing topic, please visit our Digital Library at http://www.computer.org/publications/dlib.

---

IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

Stay on top of the exploding fields of computational biology and bioinformatics with the latest peer-reviewed research.

This new journal will emphasize the algorithmic, mathematical, statistical and computational methods that are central in bioinformatics and computational biology including…

- Computer programs in bioinformatics
- Biological databases
- Proteomics
- Functional genomics
- Computational problems in genetics

Publishing quarterly
Member rate: $35 print issues
$28 online access
$46 print and online
Institutional rate: $375

Learn more about this new publication and become a subscriber today.

www.computer.org/tcbb

Figure courtesy of Matthias Höchsmann, Björn Voss, and Robert Giegerich.
Paper V
The Shannon Portal Installation: Interaction Design for Public Places

Luigina Ciolfi, Mikael Fernström, Liam J. Bannon, Parag Deshpande, Paul Gallagher, Colm McGettrick, Nicola Quinn, and Stephen Shirley
University of Limerick

The portal dolmen project at Ireland’s Shannon Airport tackled the challenges of a public exhibition and revealed the importance of focusing on situated activities as well as the crucial need for incorporating physical and aesthetic concerns into the design.

The Interaction Design Centre is involved in the Shared Worlds research project, funded by Science Foundation Ireland. This project investigates the design and deployment of interactive artifacts in public spaces.

Our approach views technology as a tool or mediator in human activities, thus requiring a careful observation and analysis of user activities as a prelude to concept design. Our work brings together aspects of the social and engineering sciences to inform and structure the design process. We undertake brainstorming and concept design activities only after extensive fieldwork at the site, with a view to understanding the various actors involved in the use of the public spaces and their interrelations as they unfold over time and space. Our iterative design process attempts to be as participative as possible in evaluating aspects of the prototypes we develop over the course of the project.

An interactive installation that we built for Shannon Airport, Ireland, let users select and personalize photographs of their own choice with annotations and drawings, then e-mail or upload them to a public image wall gallery projected in the airport transit lounge. Participants could use their own digital cameras to take pictures and then annotate and send them, or they could select an image from a corpus of public images on the system. Passersby could magnify individual images from the collage on the image wall by physically moving their bodies in front of the image wall. A computer-vision algorithm detected their movement, magnifying the image in front of that person. The image wall lens—which moved over the image wall at a rate determined by the movement of the visitors in front of the wall—provided feedback on the user’s locus of activity. The system supported both individual and collective activities, averaging the movements to determine the lens displacement.

This case study documents our efforts in designing and exhibiting a public installation in an airport setting. It presents some general reflections on the importance of using a combination of methods at different stages of the design process, emphasizing the need to understand the place in its entirety and to consider the system’s physical and material qualities when designing for public environments.

AIRPORTS AS PUBLIC SETTINGS

The challenge of studying public settings, where a broad range of activities go beyond the more traditional work-oriented scope of human-computer interaction and computer-supported collaborative work, is common to several ongoing research endeavors. For example, several current projects such as Fiasco,2 Familiar Strangers,3 and Equator’s CityWide4 deal with the augmentation of people’s activities within urban environments. Other public environments such as museums5,6 and exhibition spaces7 have provided the setting for explorations in ubiquitous computing technology that enhance people’s appreciation and interaction with them. Within this area of study, we focused on a
different and little-explored type of public space: transitional spaces, specifically airports.

Airports provide sites at which extensive research has already been done for improving the facility’s operational efficiency. However, little research has been carried out on how interactive technologies could support and enhance passengers’ experiences and activities in airports. Some technological explorations, such as SmartKiosk and WebWall, consider airports merely one potential setting, but unlike the portal installation in Figure 1 that we created for Shannon, they were not designed with the specific features of an airport—and the activities that take place there—in mind.

Arguably, the increasing number of passengers traveling by air and spending more time in airports offers scope for envisaging ways in which interactive technologies could improve passengers’ air travel experience. Our work on the portal aimed to explore these issues while designing an on-site intervention explicitly for Shannon Airport and its users.

SHANNON AIRPORT EXPERIENCES

We began Phase 1 of our project by conducting a series of field studies at Shannon Airport. These studies used a combination of qualitative inquiry methodologies, including naturalistic observations of people’s activities; video recording of specific areas of the space such as check-in desks and customs hall; semistructured interviews; and conversations with passengers, visitors, and staff.

Quantitative methods included surveys of flow and dwell time through the airport’s different areas. Many passengers fill the ample waiting time they commonly experience with activities such as talking to friends and relatives, working on laptops and talking on mobile phones, looking after children, reading, sleeping, and eating. At the same time, issues of airport security control that pervade their activities affect people’s experiences, sometimes making them anxious to comply with the airport’s rules of behavior.

In conducting the informal interviews, the conversations followed such themes as “What was your best or worst experience in an airport?” “What about here at Shannon?” “If you were to recommend this airport to a friend or a family member, how would you describe it?”

We collected 21 interviews with passengers and visitors and eight interviews with staff members. Recounting their personal stories, the participants described the main issues that characterize their experience as air travelers and airport users and the main dimensions of the activities they perform in this context that the technology could support. Major findings include the following:

- In contrast with some of their previous experiences at other airports, people described Shannon overall as a welcoming environment because of the airport’s small size and the staff’s friendliness.
- Respondents expressed a need for constantly up-to-date information regarding flights, as well as the need to relax with activities such as reading and conversing. Overall, they expressed an interest in the presence of novel forms of entertainment and engaging activities at the airport if these were not too intrusive, demanding, or distracting.
- Several interviewees commented on the lack of a strong link between the airport and its surrounding area, West Ireland. Non-Irish passengers waiting for connecting flights noted that they had only a vague idea of Shannon Airport’s location and wanted to know more about the region. Overall, people would appreciate it if the airport had a more geographically rooted identity.

CONCEPT GENERATION

In Phase 2, we proposed and discussed several design concepts, based on analysis and reflection over observational and interview data. These discussions took place at a series of planning and design meetings and during several informal conversations among Shannon Airport personnel.

The first design session used keyword cards to trigger ideas and potential scenarios. The group subsequently discussed a smaller set of cards selected by the coordinator of the work at Shannon. These included keywords such as the following:
flow and paths: tracking and representation of people’s and planes’ movements in and around Shannon;
• luggage: what people bring with them as they embark on a flight; and
• postcards: a record of stories and episodes from the trip, to be shared with other passengers and travel companions.

The design team discussed the themes’ potential development in relation to the findings from field studies. As Figure 2 shows, after formulating short scenarios in relation to each keyword, the team selected a postcard’s theme and developed the initial idea for an installation that would let passengers and visitors produce a record of their trip for others.

We decided on a fixed installation form to respect people’s desire to avoid being engaged by a pervasive or mobile installation. This approach leaves people free to decide whether to interact with the installation or not. We also decided that an ambient component—such as a visual or auditory display—should be associated with the stand-alone piece to provide some entertainment for onlookers.

In summary, we focused on the following main requirements for developing the postcards scenario card:

• allowing people the freedom to engage with the piece at different levels, from active participation to onlooking;
• providing some entertainment in a space considered quite boring, especially for children;
• ensuring anonymity of the participants and respecting the airport security constraints;
• potentially involving airport staff members without interfering in their duties; and
• creating a link between the airport and its geographical area—in this case, West Ireland.

In a discussion of these ideas, the airport management staff provided useful feedback and suggestions for the scenario’s development. After further discussions, the participants decided that the proposed system would enable users to create e-cards of their own photos and annotate them, thus allowing for individual contributions to the installation.

Our initial discussions of the actual physical form the installation would take centered on the idea of an old-fashioned post office desk/counter—the post office being a traditional meeting and connection point in Irish society. We then discussed the possibility of a more timeless and less cluttered design such as the one shown in Figure 3.

Our ongoing discussions developed a radical concept, with strong cultural, historical, and geographical connections. As we developed the core idea further, it transformed into a modern-day, technologically enhanced portal dolmen. The people of Ireland’s Neolithic Stone Age built stone monuments, known as portal dolmen tombs, in the center of each community or tribe. These monuments served as a focal point and record of who their builders were. The highest density of portal tombs can be found in West Ireland.

DEVELOPMENT

Throughout the development phase, the design team further developed the idea of the portal as a link between present and past and as a community symbol rooted in the area’s history.

In Phase 3, we determined that the dolmen form fits appropriately with several concepts surrounding our design: It offers a linkage to local heritage and history, can be built in local material, represents the focal point of a community and its memento, and plays on the idea of a geographical and temporal boundary—which is intriguing when connected to the identity of the airport as a boundary space, in itself, between here and there, earth and heaven, and coming to the earth and departing from it. Moreover, it presented the opportunity for our Shannon Airport design to be unique and strongly linked to the place and its traditions. We based the installation’s final design on three spaces and interaction levels.

Interactive dolmen

The dolmen’s functionality resembles the increasing numbers of “make your own prints” machines commonly available today. Unlike other systems, users can annotate
their photos by, for example, using an electronic stylus to draw on the uploaded image. They can then submit their annotated images to the portal’s public gallery. The images smoothly leave the dolmen’s screen interface to reappear on the image wall, sliding into place, as Figure 4 shows.

The activities the dolmen supports resemble those featured in other installations, such as CommPose, a project that lets mobile camera phone users send their annotated pictures to a public screen. However, CommPose does not allow for handwritten annotations and e-mail facility, important communication features allowed through “the portal.” Further, these features can only be used by participants who possess a camera phone.

**The image wall**

The dolmen display projects parts of the gallery on a nearby image wall. The projection provides a window into the gallery space. At any given time, a collage of uploaded images, with their annotations, displays. As Figure 5 shows, with the image wall lens, a virtual magnifying glass controlled by their gestures in front of the projection, users can navigate the overall image wall.

To further improve users’ feeling of control when manipulating the virtual magnifying glass on the image wall, we added an auditory display dimension. Given that an airport is quite noisy, we decided to use an inharmonic Shepherd tone illusion that provides left/right information through panning, and up/down information through the direction of pitch change. We chose the inharmonic series to fit the site’s general noise spectrum while still being easy to segregate.

The image wall allows ambient, implicit, and subtle interaction, following the terminology of Daniel Vogel and Ravin Balakrishnan. When not in use, the display shows a selection of uploaded photographs and a short video about how to interact with it. If a person approaches the image wall, the display comes to life both visually and sonically, giving the user feedback on its active state. Once at the dolmen, a user can actively navigate the image archive by controlling the magnifying glass through body movements.

Much research has been conducted on using large interactive displays for public places. However, these explorations usually take place in semipublic areas, such as offices and labs, and target a particular group of users—mainly workers in a specific organization—rather than the general public. For example, the Palimpsest installation, which features a gesture-controlled display similar to the portal image wall, was entirely deployed within the research lab itself. Even its physical design did not take into account any requirements for public exhibition, such as safety and robustness.

**Web image wall**

We have made a similar window to the gallery space simultaneously available at www.shannonportal.ie/gallery. The availability of free Internet access at Shannon Airport meant that this third component of the installation could be accessed both on- and offsite.

After reflections on the choice of appropriate material, scale, and location, the components were sketched, technically drawn and specified, and built. Two Shuttle PCs (www.shuttle.com) were used in the system, one for

---

**Figure 3. Postcards concept and implementation.** (a) Early post office sketches for Postcards, a design intended to echo the local post office, a traditional meeting and connection point in Irish society; (b) and (c) first cardboard prototype for the Postcards installation.
the image wall, the other for the interactive dolmen. In the dolmen, the PC connected to a Wacom (www.wacom.com) Cintiq interactive pen display and a multi-format flash card reader. We implemented the software for the dolmen in Macromedia Flash (www.adobe.com) and PHP (www.php.net) running on Ubuntu Linux (www.ubuntu.com).

Throughout our prototyping, we had experimented with using both Microsoft Windows and Apple OS X, but handling issues with flash cards from the users’ cameras and mobile phones revealed that we needed full control over the mounting and unmounting processes. Hence, Linux proved to be the preferred option.

The image wall PC connected to the dolmen PC via a 100BASE-T Ethernet, a Logitech Web camera mounted above the user space in front of the image wall projection, and a Barco (www.barco.com) iD R600 projector configured for back-projection on a Glimm (www.glim.nl) Blackfire screen. We implemented the image wall PC software in Microsoft C/C++ and Windows XP, using Intel’s OpenCV computer vision library (www.intel.com/technology/computing/opencv) to detect user movements in front of the projection screen and OpenGL to generate the dynamic visualization of the images users uploaded.

We implemented the interactive sonification generating the Shepherd tones in Pure Data (http://puredata.info). When users uploaded images or annotated and e-mailed them, the system copied this data to a Web server running Redhat Linux 9 (www.redhat.com) and Apache 2 (www.apache.org). Figure 6 shows an overview of the portal’s technical infrastructure. The dolmen screen interface and the image wall went through several usability evaluation phases involving test users in our lab.

LAB EVALUATION

In Phase 4, experts from our research group who were external to the project evaluated initial versions of the dolmen interface. They conducted usability inspections on the interface that uncovered its most significant usability flaws. We also involved students and postgraduate researchers in trials of the image wall, discussing and evaluating alternative designs for the photo display and exploring ways of controlling the magnifying glass. We applied criteria such as visibility, feedback, and learnability, then sketched suggestions and alternative ideas regarding the image wall on a shared SMARTboard.
system to visualize them on a scale similar to that of the image wall itself.

Once we implemented the suggested changes and developed the installation further to allow more extensive trials, we undertook cooperative evaluation sessions using the thinking aloud technique, which involved participants from the broader university community and from our personal social networks. Fifteen people agreed to participate in the cooperative evaluation of the portal, which took place in our building’s lobby. Participants from Ireland, Italy, Spain, and the US took part in sessions based on a task list for exploring the dolmen along with more informal trials of the image wall in which researchers asked participants to comment on the display’s feedback and visibility.

The interface still presented some small technical glitches, which the design team took into account when conducting the evaluation. We invited all users to bring their own digital cameras, and four did. We briefed the users on the project’s nature and goals, but not on the portal’s functionality, withholding this information intentionally to assess how easily they could understand the portal’s function and evaluate the effectiveness of a set of video instructions continuously showing on the dolmen screen when not in use.

Following the introduction, we asked all participants to perform two series of tasks on the interface. The first dealt with sending an image from the image gallery, the second with sending another image from a digital camera or memory card. In general, all participants found the installation intriguing and engaging, and considered it potentially very successful within the airport. They all stated they would use it if they encountered it at the airport, and some participants said they would encourage their children to use it.

The dolmen shape was well liked. However, one participant said it gave her the impression the installation would provide her with historical information on the local area. Following up on this comment, we decided to design and print leaflets to be distributed in the transit lounge that clearly described the installation’s goals. Usability issues that emerged from the cooperative evaluation sessions resulted in the redesign of elements such as the dolmen’s initial screen layout, card reader feedback, and sketching interface.

The portal, installed in the transit lounge at Shannon in July 2006, remained available to the public for three weeks. During that time, approximately 1,500 people interacted with it in some form: Specifically, 432 photographs were uploaded to the image wall and a total of 535 e-mails were sent.

PUBLIC EXHIBITION OF THE PORTAL

During Phase 5—the Portal’s public exhibition—we conducted an extensive evaluation in situ. We collected most of the observational data as video and audio recordings, using two cameras and two microphones located in the piece’s surroundings. Naturalistic observations and note taking integrated this process. We also conducted conversations and informal interviews with both passengers and airport staff to obtain more explicit comments regarding the installation’s features.

A brief overview of the main issues that emerged from the portal’s public testing highlights how our design has led to interaction patterns in situ. The data analysis underscored the following important issues regarding people’s interactions with the portal:

- The installation supports multiple levels of engagement, facilitating a variety of behaviors that range from onlooking to active participation and interaction with the piece’s different components.
- The portal engendered and supported collaboration and social interaction, including collaborative annotation and drawing, joint exploration of the image wall, and exploration of the piece by two or more users who gave directions to each other.
- Participants appreciated the material and aesthetic qualities of the piece and its distinctiveness. They used the portal’s artificial grass area for sitting, playing, and posing for photographs, and hugged the dolmen while exploring it from every angle.
- Although using auditory display in a public environment always poses a challenge, in this design and environment the auditory display element worked well both for grabbing initial attention and contributing to users’ feelings of directly controlling the system—no users found the sound annoying or intrusive.
- Participants appreciated the possibility of being able to contribute their own content, while being able to choose whether to keep it private or make it public.

Privacy and anonymity are sensitive issues within the airport environment, and the portal’s design let people be in control of the content produced and its form of display—by e-mail only, on the image wall, or both.

In terms of the visitor interaction with the Shannon installation, several issues arose, and we are still analyzing an extensive corpus of video data. We are also writing up a more extensive account of the prototype evolution, from paper sketches through to the deployed installation, as well as mining the video material for...
more microlevel analyses of visitor movements and interactions around and through the installation.

Besides producing an installation that visitors found successful, the design process gave us valuable insights about the need for integrating different methodologies and blending them seamlessly into the design process. Facing the challenges of a public exhibition revealed the importance of focusing on situated activities as well as the crucial need for incorporating physical and aesthetic concerns into the design.

Acknowledgments

We acknowledge the support of Science Foundation Ireland for the Shared Worlds project. Thanks to Joe Buckley, Claire McEnery, Dorothy Quinn, and everyone at Shannon Airport for their support and to Marcin Ziemnicki, the Sculpture Factory Cork, and all at the IDC for making the installation possible. Thanks to John McCarthy for his contribution to the project and for his useful comments on previous versions of this article.

References


Luigina Ciolfi is a postdoctoral research fellow and director of the Digital Media Design program at the Interaction Design Centre, University of Limerick. Her research interests include interaction design, ubiquitous computing for public spaces, computer-supported cooperative work, and augmented physical environments. She received a PhD in interaction design from the University of Limerick. Contact her at luigina.ciolfi@ul.ie.

Mikael Fernström is the director of the master’s program in Interactive Media at the University of Limerick, Ireland. His main research interests are interaction design, sound, music, and electronics. Fernström received an MSc in human-computer interaction from the University of Limerick. Contact him at mikael.fernstrom@ul.ie.

Liam J. Bannon is the director of the Interaction Design Centre and professor of computer science in the Department of Computer Science and Information Systems at the University of Limerick. His research interests include improving the usability and desirability of computational artifacts. He also examines alternative conceptual and methodological frameworks for understanding human activities and their mediation by various technologies. Bannon received a PhD in experimental psychology from the University of Western Ontario, Canada. Contact him at liam.bannon@ul.ie.

Parag Deshpande is a senior faculty member with the National Institute of Design in India. He is currently pursuing a PhD with the Interaction Design Centre, University of Limerick, where he works on the Shared Worlds project. Deshpande helped establish the postgraduate
program in software and user interface design at NID, the first in India. He received an M.Arch. from the School of Planning and Architecture, New Delhi, India. Contact him at parag.deshpande@ul.ie.

Paul Gallagher is a PhD student in the IDC, currently working on the Shared Worlds project and on a tracking toolkit (www.idc.ul.ie/toolkit). His research focuses on the design, implementation, and evaluation of software tools and interaction techniques that more seamlessly integrate the physical and virtual worlds. Gallagher received an MSc in computer science from the University of Limerick. Contact him at paul.gallagher@ul.ie.

Colm McGettrick provides technical support in the Interaction Design Centre. He received a BS in physics and maths from NUI Maynooth. McGettrick’s move into the computer field has principally resulted from running his own Web development company for four years. Contact him at colm.mcgettrick@ul.ie.

Nicola Quinn is a research assistant in the Interaction Design Centre. Her research interests include haptics, artistic interfaces, multimodal interfaces, the design of interfaces, digital art, and creating technology for art. Quinn received an MSc in computer science from the University of Limerick. She has worked on the European project INDeT and Shared Worlds. Contact her at nicola.quinn@ul.ie.

Stephen Shirley is an MSc student in the Interaction Design Centre at the University of Limerick. His research is currently part time, as he is also employed by Google. His research interests include tactile feedback and simplifying sensor integration for designers. Shirley received a BEng in computer engineering from the University of Limerick. Contact him at diamond@skynet.ie.

Looking for accessible tutorials on software development, project management, and emerging technologies? Then have a look at ReadyNotes, another new product from the IEEE Computer Society. ReadyNotes are guidebooks that serve as quick-start references for busy computing professionals. Available as immediately downloadable PDFs (with a credit card purchase), ReadyNotes sell for $19 or less. www.computer.org/ReadyNotes

Here now from the IEEE Computer Society

IEEE ReadyNotes

Available as immediately downloadable PDFs (with a credit card purchase), ReadyNotes sell for $19 or less. www.computer.org/ReadyNotes