Interaction Design Institute Ivrea, May 2004

Audio-Tactile

Audio-Tactile interactions with mobile devices

Michal Rinott
m.rinott@interation-ivrea.it

Thesis Committee

____

Michael Kieslinger
Associate Professor
Primary advisor

____

Jan-Christoph Zoels
Senior Associate Professor
Secondary advisor

____

Gillian Crampton-Smith
Director

____

Andrew Davidson
Chair of the Academic Programme

Audio-Tactile

Audio-Tactile interactions with mobile devices

Michal Rinott

Abstract

Our interactions with physical objects usually produce sound; this sound is tightly connected with our actions and the character of the objects with which we interact. This thesis is concerned with designing the same quality into mobile devices: interactions in which you ‘do things with your hands’ and get sound of both informative and expressive value. In three explorations into this relationship between hands, ears and mobile devices, audio-tactile interactions were designed for activities as diverse as text entry, sharing music between devices and composing the ring on a mobile phone. These designs explore the qualities of audio-tactility and test the idea that interactions of a functional nature can be situated on the border between task, instrument and game.

Table of Contents

Acknowledgments

This thesis would not be as it is without Michael Kieslinger
Jan Christoph Zoels gave invaluable advice and support
Edoardo Brambilla is the lord of the Keybong
Christian Palino created the SonicTexting graphic
Tal Drori helped with graphics and ideas all along the way
Hernando Barragan, Yaniv Steiner and Massimo Banzi taught me electronics
Yon Visel took my Max patch and skills some levels up
Stefano Mirti pushed forward at important moments
Giorgio Olivero gave the Keybong its name
Silvia Rollino gave the Keybong its voice.

Belmer, Davide, Daniele, Bernd, Oznur, Peggy, Ruth, Erez, Ricky and Manuela were there in moments of need, as well as many others from the IDII community.

Alexander Grunsteidl and Graham Pullin at IDEO London gave me a great summer start.
Gillian Crampton-Smith, Andy Davidson and the IDII faculty made me believe I could pull this off.

Special thanks to the people in my class:
walking this path with you has been an amazing experience.

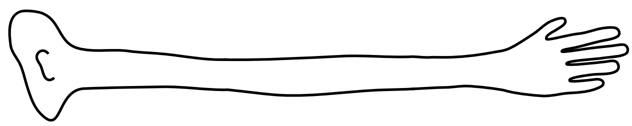
INTRODUCTION

At the onset of this work I posed the following question to a group of people: "which products, physical or digital, have a sound that you particularly like or dislike?"

Some people disliked the sounds of their mobile phones; others, the buzz at the end of the washing machine cycle and other beeps and bleeps. Some liked the sounds of musical instruments; others, their motorbike engine or the locking of their briefcase.

Two products especially caught my attention. One person loved the sound made when closing the cap of a certain bottle of perfume; another liked the clicking sound made with the rotating cap of a container of "Smarties" candies. I got those products and played with them: they were great. The perfume cap snapped so nicely when I closed it! And the Smarties box made really satisfying clicks when rotated, with the sound changing slightly according to the speed of movement.

The special qualities of the interaction with these products - the strong correspondence between movement and sound, the way the sound varies according to the action, the pleasure of working with hands and ears - are the main motivations for this thesis.



1 Ears and hands are connected

When we interact with physical things in the world, these interactions sometimes create sound. The nature of this sound is a combined product of our actions and of the physical attributes of the objects with which we interact – their form, materials and dynamics, as well as the surrounding environment. People possess a natural capacity for understanding information from sound.

In the world of digital technology our interactions sometimes produce sound. Quite often, though, this sound tends to be a more binary, on-off phenomenon. This thesis looks at uses of sound that have a more complex connection to actions. It proposes that digital interactions can be enriched by turning them into carefully designed experiences with a sophisticated relationship between movement and sound. This type of interaction is audio-tactile interaction, involving tactile input and auditory output: hands and ears.

The potential of audio-tactile interactions has two possible facets. Sophisticated audio-tactile connections are a way to create interactions that are rich in information and that tap into people's natural ability to extract information from sound. Another facet is enjoyment: the qualities discussed above can be a source of satisfaction or pleasure.

Mobile devices are portable digital objects designed to be used 'on the move'. This definition includes communication devices (e.g., mobile phones), information devices (e.g. PDAs), music players and more. Mobile devices are designed to be compact and yet to provide complex functionality.

I chose mobile devices as the ‘vehicle’ through which to explore audio-tactility in digital products. Mobile devices have strong auditory and tactile potential, while the eyes are often very busy in mobile situations. In this paper I present an exploration of the potential of hands, ears and mobile devices through the creation of audio-tactile interactions for selected functions of mobile devices. I ask - can these functional activities have qualities of enjoyment that situate the experience on the border between a functional object, an instrument and a game?

BACKGROUND RESEARCH

2.1. The Auditory

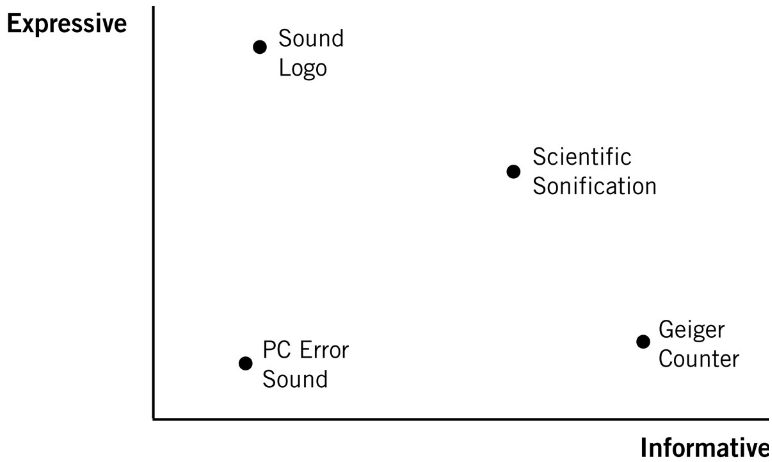
> "Sounds thicken the sensory stew of our lives, and we depend on them to help us interpret, communicate with, and express the world around us"
> • Dianne Ackerman

Sounds alert and inform us, speech connects us to others, and music is, as Ackerman defines it, "the perfume of sound". Sound is a powerful stimulus with a strong effect on our emotions. It is a stimulus that holds huge potential in the designer's palette.

2.1.1. Informative and Expressive Sound

Michel Chion, a researcher of film sound, has referred to two types of added value of sound in film: informative and expressive (5). I will use these categories, informative and expressive value, to discuss the use of sound in design. It should be noted that these categories are not mutually exclusive; in fact, I believe sound must be both informative and expressive in order to be used to its fullest in design. In the figure below a number of uses of sound are positioned according to these categories. The sound of a Geiger counter is, arguably, pure information. A sound logo: mostly expression; while it contains information, it is predominantly designed to create a feeling, an expression of, for example, a company's character.

2 **Expressive and Informative sound in design**



The Informative Value of Sound

Sound is information. As Ackerman notes, we use sound in everyday life to interpret the world around us. Through sound we infer the functioning of our car motor, our washing machine; in shared spaces we often maintain an awareness of the actions of the people around us through the sounds they make. Even subconsciously, we are constantly monitoring the environment through sound.

Using sound as a material in design is based on an understanding of the attributes of both sound as a phenomenon and human hearing as a processing system. Gaver compared looking and listening and pointed to the following contrast: “although sound exists in time and over space, vision exists in space and over time” (“SonicFinder”, 71). This distinction implies two major facilities of sound as an informational medium. Sound is well suited for displaying events with a beginning and an end - events that change over time. Also, sound’s omni-existence in space enables it to be perceived without requiring the listener to face its source. Since sound is caused by the vibration of materials, it can provide information that vision cannot: information about occluded, hidden things; about interiors of mechanisms.

People have sophisticated systems for perceiving and processing sound, and studies of auditory perception point to many advantages of sound for communicating information. To mention two examples:

- Humans are capable of parallel listening: monitoring and processing multiple auditory data sets. This is especially true if the different aspects of information have different auditory characteristics (Kramer).

- Humans have good abilities to discern relationships and trends in auditory material: to identify patterns. This is called ’Auditory gestalt formation’. Specifically, people are good at identifying co-variation between sound elements (two things changing in relation to each other) and at detecting small changes in continuous signals (e.g., the increase or decrease of the refrigerator hum).

Information sonification may well be the most abstract informational use of sound. Sonification is the use of sound to represent non-auditory data. This field has seen significant development in the last 10 years. Sonification was initially used for highly scientific purposes (e.g. exploration of seismic data through sonification), but is now being used also in more mundane contexts (e.g., sonifications of web site traffic, different educational purposes). Typically, multiple layers of information are mapped into different acoustic dimensions. Well-designed sonifications use mappings that maximize peoples’ perceptual abilities, such as the capacity for parallel listening and the auditory gestalt formation.

The Expressive Value of Sound

Sound is expression. It has powerful effects on our feelings and our mood. Film sound designer Randy Thom listed some of the ways in which sound creates expressive value in film: it determines the mood of a scene through music and external sounds; it can direct attention to, and from, different elements of the scene. Sound can affect the actual perception of the pace of a scene, rendering it faster or slower than it ’really’ is; it can give certain character to people, objects and abstract concepts within the flow of the scene through leitmotifs and other sound manipulations (Sider 127).

Today, products such as cars and motorcycles have their motor sounds tuned and modified to express qualities such as power and durability; even the materials of the doors and buttons are selected in order to produce specific sound character. Not only noise level but also sound quality is awarded attention in the design of 'noisy' products such as vacuum cleaners, keyboards, hair dryers and more.

The ability to create character through sound holds special potential for added value. It can bestow a product with a quality or 'feel' that might make the difference between an anonymous experience and a satisfying, desirable one. The design challenge this creates is significant: when sound is introduced into a product, how does one select the 'right' sound character for it?

2.1.2. Symptomatic and Introduced Sound

Objects communicate with us through sound, in a range of different ways. One distinction I have found useful is between sound that is 'symptomatic' and sound that is 'introduced'. Symptomatic sound is the by-product of a natural operation of an object. Introduced sound is sound that is added to an object for communication purposes. Examples of symptomatic sound are the click of the lock when closing a briefcase or the hum of the spinning hard drive. Examples of introduced sound are the shutter-like click added to many digital cameras when pressing the shoot button or the ring of a mobile phone. As is apparent in the above examples, mechanical sounds are usually symptomatic and digital sounds are usually introduced.

Symptomatic sounds typically contain inherent variation, while introduced sounds typically do not. The locking sound of the briefcase will be slightly different every time, depending on the exact closing motion and on other external factors (e.g. the degree of humidity). We perceive this variation and, I believe, derive satisfaction from our implicit expertise in discerning the correlation between action and the auditory response. Introduced sounds are often binary signals – fixed responses to predefined conditions. This may well explain the irritation that people feel towards many of the sounds that digital objects make. These sounds contain a low ratio of data-to-sound: they demand our attention but reward us with a relatively small amount of information.

While sounds of digital products are typically introduced, this clearly does not mean they must be binary or single-layered. Introduced sound can be designed to contain richness, using more information and more variation in sound. It can have multiple layers of information, some immediately recognizable and others requiring more prolonged use and acquired expertise.

Bill Gaver's early SonicFinder interface is an impressive example of sophisticated introduced sound with multiple layers of information. The SonicFinder was a sound application built over the Macintosh Apple file system. In the SonicFinder, dragging a folder sounded differently according to its size and type. Also, dragging sounds differed according to the surface below: dragging over windows sounded different than dragging over the desktop. The SonicFinder demonstrated that auditory information for a product can be designed to be of different complexity, with some elements immediately apparent and some requiring more prolonged use and acquired expertise. Sound elements can have different relationships with the actions that invoke them: from binary signals to complex 'analog' sounds; from sound that corresponds directly to visual information to sound that reveals hidden information not otherwise present in the interactive experience.

2.2. The Tactile

> "By pointing, by pushing and pulling, by picking up tools, hands act as conduits through which we extend our will to the world. They also serve as conduits in the other direction: hands bring us knowledge of the world. Hands feel. They probe. They practice. They give us sense…"
> • Malcolm McCullough (2).

Tactility is defined by the Merriam-Webster dictionary as being "of or relating to the sense of touch". Tactility is addressed in this thesis in a restricted way - to refer to the use of hands. Furthermore, in the current context the tactile component is not a focus on its own but in its relation to the auditory component. I will present tactility briefly by focusing on two aspects of the use of hands: skill and gestures.

2.2.1. Skilled Hands

Skill, according to McCullough, is a source of satisfaction, and can "develop an intimate relation with certain contexts or tools, which makes it individual" (7). He quotes Octavio Paz in his discourse on craftsmanship with hands: "In the work of handscraftsmen there is a constant shifting back and forth between usefulness and beauty. This continual interchange has a name: pleasure. Things are pleasing because they are useful and beautiful" (9).

At the current state of interactions with technology McCullough sees hands as "not idle, just underemployed" (2). He promotes the potential for creating technological objects in ways that support the subtleties of the hand. A clear inspiration for this kind of skillful, subtle tactile manipulation is the way musical instruments are played, and, differently, the way gaming controllers are used. In both music performance and gaming, the player acquires skill in using an instrument – a tool – and this skill is a source of pleasure as much in it's process as in it's outcome.

Bill Buxton, a musician and computer scientist, has written extensively on the need for tactile richness and skill in common computer interactions, especially those involving long-term use and expertise. In 1999 he posited: "'Hands-on computing is a myth…..A professional violinist will spend more on a bow than computing professionals spend on a workstation. This is simply in order to be able to capture the full subtlety of gestural nuance that they are capable of producing" ("Human Input", 1).

2.2.2. Hand Gestures

Gestures are motions of the limbs or body used as a means of expression. More technically, gestures have been defined by Bongers as “a multiple degree-of-freedom meaningful movement” (5). Gestural control over computers can be a way to exploit the natural human propensity for manual skill and expertise. This thesis focuses on a significant subset of gestural expression: gestures produced by the hands and mediated by the use of some control mechanism.

Hand gestures have successfully been used as input for computer systems in many forms. The most common is probably the mouse. The gestural language of the mouse is quite simple due to its restricted gestural space and its limited ’vocabulary’ of actions which relates in a rather concrete way to the desktop metaphor.

In pen-based user interfaces such as those used with PDAs, commands are executed using gesture marks or strokes performed with a stylus. Long, Landay and Rowe have listed the advantages of using pen-based gestures: “commands issued with pens (i.e., gestures) are desirable because they are faster (because command and operand are specified in one stroke), commonly used, and iconic, which makes them easier to remember than textual commands” (1).

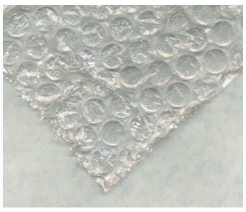
The creation of a gestural language is a significant design challenge. Some languages are designed to have a direct relationship between gesture and invoked action, while others are designed to be more abstract. In both cases gestures should have a natural relationship to the control mechanism with which they are performed and with the semantic meaning of the actions that they invoke.

2.3. Audio-Tactility

Everyday experience with physical objects is multi-sensory. We do not perceive this experience as fragmented into different sensory channels, but as a complete, unified whole. Imagine using a pencil: when we write, our manual actions produce visual and auditory traces. But we do not think of the pencil as having, for instance, sound. It is its unified ‘pencil-ness’ that we experience.

Michel Chion has stated, concerning sound in film, that: “there is no soundtrack”. An extreme statement coming from a researcher of film sound, Chion means that there is no way to separate the auditory and visual channels of a film. We experience them only through a unified sense which he terms “audio-vision”.

In line with Chion’s reasoning, I suggest that there should be no soundtrack for digital objects. In the same way that well-designed audio and vision are perceived in unison, so should well-designed audio and tactility. When we interact with digital objects, the sound they make should contribute to the holistic experience. They should have the qualities that make the sound inseparable from the actions that invoke it. I will refer to this quality of experience as “audio-tactility”.



3 Bubblewrap. A source of audio-tactile pleasure.

Hands and ears work together in coordination: an analog radio tuner is a good example of this. While there is a visual display on the radio, its visual resolution is extremely low; our hand and ear work in synchrony to fine-tune the exact position in which the sound is pure. Hands and ears work together in pleasure: "bubble wrap" is a good example of this. While it is difficult to define the exact sources of the appeal of puncturing bubble wrap, I believe that the strong synchronicity of action and sound is a central part of the pleasure.

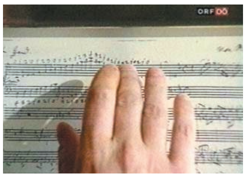
2.3.1. Audio-tactile inspirations

In this section are a number of examples of digital audio-tactile experiences, from different disciplines, that I have found inspiring for the purposes of this thesis.

4 **The "Sounding Object" project. Drinking sound synchronized to pulling the flap.**

The Sounding Object - Drinking lemonade

"The Sounding Object" (SOB) is a research project developing sound models that are responsive to physical interactions. SOB researchers Rochesso, Bresin and Fenström have presented a audio-tactile interaction using a 'pop-up' children's book – a book in which carton flaps are pulled to move elements in the page. One element in a book of this kind is a glass of lemonade that a cartoon creature is drinking with a straw. Pulling the flap down decreases the level of lemonade in the glass. In the SOB concept, a sensor and digital sound processor are connected to the flap so that pulling it will also create sound, by feeding the sensor data into a sound model of 'drinking from a straw'. This produces an auditory response that is tightly synchronized with the pulling action (45).



5 **Christian Moeller's "Fuge for Piano, op. 154". Plying the score by touching the written notes.**

Christian Moeller - Fuge for Piano, op. 154s

Artists who create interactive sound installations explore the relationship between sound and touch extensively. Christian Moeller's "Fuge for Piano, op. 154" is a touch-sensitive music installation. The original sheet music of Mozart's composition is exhibited beneath a touch-sensitive sheet of glass, fitted within a normal picture frame. By touching the glass above any given note of the music, the visitor is able to hear it played. In this way it is possible to play the composition by tracing the bars of music with one finger. The speed of movement determines the tempo of the playback.

6 **Leon Theremin playing the Theremin**

Instruments for Musical Expression

In 1920 Leon Theremin created the Theremin – an electronic instrument played by waving the hands in the vicinity of two antennas. In the 70's and 80's, Laurie Anderson used the violin and bow as an input device; outputs were audio palindromes (Tape Bow Violin), howling animal voices (Digital Violin) and more. Recently, the exploration of expressive tactile control for electronic instruments has been enhanced by the increased availability of inexpensive sensors. Many projects explore the mapping of virtually any input to any sound output; for example, in a recent course at Stanford University student developed input mechanisms such as an accordion-like object, a rocking horse and a turn-table to control midi-generated sound.

While the mappings of input to output can be almost arbitrary, I believe that good instruments capture some meaningful relationship between input and output.

The Knocking and Speaking Social Mobiles (SOMOs)

7 **The Speaking and Knocking Social Mobiles by Crispin Jones and IDEO**

The Social Mobiles project is a collection of concepts for phones created by Crispin Jones and the IDEO design firm. The Knocking SoMo is a mobile phone in which the user knocks on the phone to communicate the urgency of the call. The recipient hears the knock and can decide according to it how to treat the call. The creators speculate: “over time people would learn to recognize each others’ knocking mannerisms”. The Speaking SoMo is a concept for a mobile phone in which the ‘speaker’ of the phone uses a joystick to manually produce and intone expressive sounds.

The SoMo project was motivated by a critical approach to the use of mobile phones in public spaces; this approach has sparked interesting audio-tactile interactions with phones.

3. CONCEPT

The goal of this thesis is to explore the potential of audio-tactile interaction with digital objects. Inspired by the audio-tactile qualities of musical instruments and sound installations as well as by the richness of audio-tactile interactions in the physical world, I wish to bring these qualities into the interaction with everyday digital objects – into functional activities performed on a routine basis. I chose mobile devices as the vehicle for this exploration.

3.1. Why mobile devices?

3.1.1. Mobile devices are auditory

Many of the functions of mobile devices are auditory in nature. Verbal communication (making phone calls) and listening to music are central ones, as well as the functions of alerting attention through sound (as in rings and other sound signals). Games, movies and other newer content have sound. The quality of sound in these devices is becoming increasingly high. Furthermore, devices such as the Nokia 3300 Music Phone are full music players designed for long-term listening. These developments make rich uses of sound in interaction both feasible and potentially pleasurable for the relevant user population.

3.1.2. Mobile devices are tactile

Mobile devices are held and manipulated with hands. In his analysis of the mobile phone as an extension of the body, Anthony Townsend wrote: “Individuals develop very personal relationships with mobile telephones... It should come as no surprise that in London, mobile phones recently became the most commonly left item on subway trains in the underground, replacing the umbrella. …Despite the fact that there is no service in the deep tunnels of the Underground, people are still losing their mobile phones because they are holding them in their hands” (6).

Mobile device manufacturers are constantly seeking to expand the input mechanisms of these devices. Among others, many phones now include small navigation joysticks; the Apple iPod boasts a touchpad with great tactile qualities; and the Sony minidisk comes with a clip-on remote control device.

Mobile phone users develop impressive skill in the use of their devices. The expertise that teenagers, especially, have developed in using the number pad for text-messaging demonstrates the potential for skilled interaction in this context – even if the interaction method is quite clearly sub-optimal.

Pressing on a miniaturized pad of buttons uses an extremely limited use of hands; they can potentially be used for more complex and expressive movements in controlling mobile devices. Harrison et al. explored different designs and uses of tactile user interfaces for mobile devices. They used tilting, handling and squeezing gestures to control a number of specific tasks. These experiments demonstrated that gestures can be successfully used to control mobile devices and can simplify input in a variety of situations.

3.1.3. Mobile device interfaces

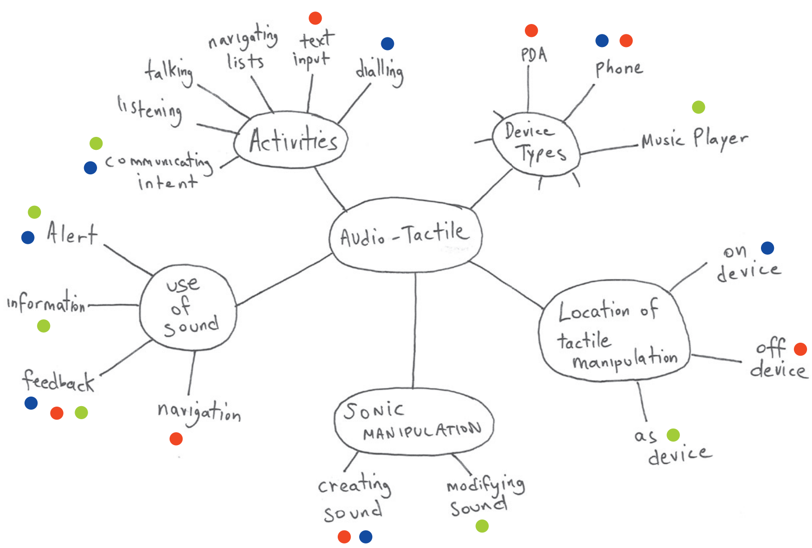
Mobile devices are portable objects carried and used 'on the move'. They are designed to be compact and yet to provide complex functionality. As these devices develop and in some cases converge, their functionality becomes increasingly more complex. Pirhonen et al have recently described the challenges this creates: "Designing interfaces for mobile computers… is problematic as there is a very limited amount of screen resource on which to display information and users' eyes are often needed on the environment rather than the interface (so that they can look where they are going)… so output is limited…." (291).

Wessel and Wright have defined the following aspiration for the design of musical interfaces: "low entry fee with no ceiling on virtuosity" (1). If interfaces are too easy to master, their expressive possibilities are quickly exhausted; if they are too difficult to get started with, they may be quickly discouraging. Clearly, interactions with mobile devices should not require the kind of expertise associated with musical instruments. But a certain degree of skilled interaction may lead to a richer interactive experience.

3.2. Mind-mapping opportunities

In the 'Mind-Mapping' diagram below I have laid out some of the dimensions of audio-tactile interactions with mobile devices to get a feeling for the breadth of opportunities in this combination.

The three explorations designed in this thesis are marked in within the mind-map in three different colors. These explorations: SonicTexting (in red), Shake (in green) and Ring (in blue) are presented in the design section below.



DESIGN AND IMPLEMENTATION

This thesis comprises one central project and two smaller explorations of audio-tactile interactions with mobile device activities. The central project is SonicTexting, an audio-tactile interface for text entry. The explorations are Shake and Ring. Shake is an audio-tactile interface for sharing music between portable music devices. Ring is an audio-tactile interface for real-time composing of the ring for a phone call.

This section describes the design and analysis of these in detail.

4.1. SonicTexting

> 'Can you write an SMS from your pocket?'

4.1.1. Overview

SonicTexting is an audio-tactile system for inputting text using gestures and sound. It can potentially be used with any device (e.g., Mobile phone, PDA) in which text input is needed. SonicTexting was designed with the goal of enabling 'texting' without vision - using only touch and sound. It uses generated sound for navigation and feedback purposes.

The writing method of SonicTexting is inspired by 'Quikwriting', a Graffiti equivalent for PDAs developed by Ken Perlin at NYU. In SonicTexting, smooth writing gestures are made using the Keybong, a joystick-like input device that fits in one hand. The joystick is manipulated using the thumb. The Keybong gestures produce tightly synchronized auditory responses.

4.1.2. Motivation – 'texting' and 'thumbing'

8 **James Trusler, texting world champion, in action**

Texting world champion, 30-year-old James Trusler recently typed:

> "The razor-toothed piranhas of the genera Serrasalmus and Pygocentrus are the most ferocious freshwater fish in the world. In reality they seldom attack a human"

into his phone in just 67 seconds

Researcher Sadie Plant finds that 'texting' has had a profound effect on the way teenagers use their thumbs (18). Because they are used to tapping out numbers and messages with their thumbs, they now point and even ring doorbells with their thumb instead of their forefinger.

Texting seemed a ripe area for exploring skilled audio-tactile interaction.

The SonicTexting working prototype was created as a series of design iterations. I will describe the current version and the design process that led to it.

4.1.3. The Sonic Texting prototype

Input Device

9 **The Keybong**

The Keybong is a one-handed input device that fits in the palm of the hand and is manipulated with the thumb. The Keybong's joystick naturally supports the common gesture pattern of SonicTexting – moving from the central location, through a specific path, and back to the center. The joystick is also a button: pressing it clears the entered text.

The Keybong contains a small eccentric motor that gives gentle vibro-tactile feedback in the writing process. This tactile layer accompanies and augments the sound layer. The Keybong shape is designed to ensure that it is held in a fixed orientation.

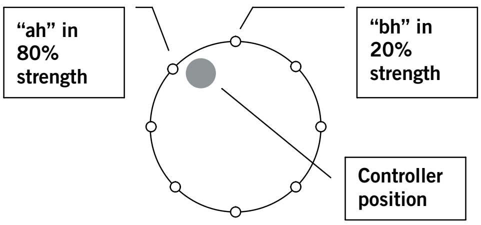
Sound

In SonicTexting letters are located using sound. Sound provides continuous feedback for navigation during the movement: an interactive sonification of the gesture path. There are two sound modes: beginner mode, which optimizes learning and memorization, and expert mode, which optimizes speed and effectiveness. Sound is also used after the gesture for a letter-by-letter 'readback' of completed words.

Beginner Mode - In this mode continuous, looping letter phonemes (e.g. – the sound "bh") are played in synchrony to movement. Loudness is a function of the distance between the controller and the eight equidistant points distributed on the periphery of the controller range.

When the controller is located between points, the phoneme sounds mix. When the controller is located close to a point, the single phoneme sound is pure.

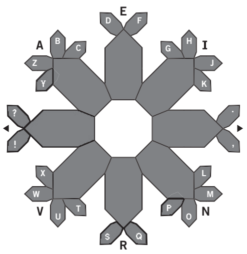
The letter phonemes are spoken-sung in different pitches according to their location. The pitch ascends clockwise, starting at the top-left corner note by note through one octave.



Expert Mode – In this mode discrete, synthetic percussion-like sounds are played in synchrony to movement. Each of the eight perimeter points has a different pitch in the same scale as in the beginner mode. The sound is played when the controller 'acquires' a point. The velocity of the movement is mapped to the loudness of the initial part of the sound (the attack), so that stronger movements create stronger sounds.

The beginner and expert mode require a different behavior for making use of the sound. Beginner mode is designed for slower movement and listening, and does not benefit from being used rapidly (since the looped phones may not be heard in a short time slot). Expert mode is designed to enable maximal speed in movement.

Gesture Map



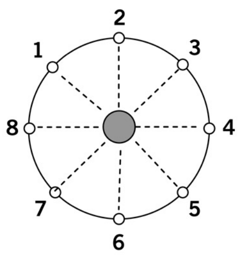
The gesture map is a static visual representation of the gesture path for each letter. It is a guide and means for learning the gesture paths.

The gesture map is read as follows: to write an ‘axis’ letter (i.e. the letters on the main axes, here printed in black) the controller is moved to that location and returned to the center. Upon return to the center the letter is written. To write a nested letter (all other letters) the controller is initially moved to the axis location, then to one or two positions according to where the letter is written relative to the axis letter, and returned to the center. For example:

- To write an “A” the controller is moved to the upper-left corner, then back to the center.

- To write a “B” the controller is moved to the upper-left corner, then moved one position to the right, then back to the center.

Software Implementation



This is a short description of the algorithm implemented in the MAX/MSP software to create continuous sound according to position (see appendix B).

MAX receives x,y data from the controller. For every position input, the distance of this position from eight equidistant points on the perimeter of the x,y range is calculated. Each of these points is a sound source (SS). The distance from a source determines the loudness of the sound coming from that source.

Each sound source is set to a certain letter-sound depending on the movement path; specifically, the first direction taken by the user sets all sound sources into that mode. For example, when the controller is moved to A, SS1 is set to “ah” SS2 is set to “bh” and SS3 is set to “ch”. Returning to the center resets all sound sources to the default set.

Mental model & Instructions

The basic instruction given to people is “move in the direction of the letter; use the sound; try to find the pure sound”.

This achieves the goal of making the task perceived as predominantly auditory: an auditory space that is being navigated with the controller.

The guidance material for SonicTexting is a video in which the Keybong is shown in use simultaneously with the map. The map contains a dynamic visual indication of the current direction of the controller. The letter being written is displayed below and the written text builds up dynamically.

Hardware Implementation

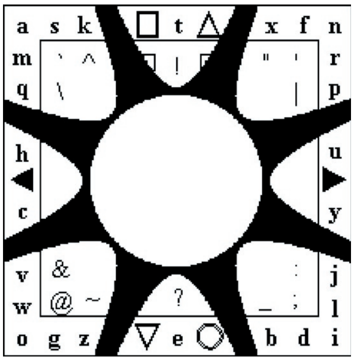
A BX-24 microcontroller was used for two-way serial communication between the

Keybong and the MAX software. Normalized X, Y and button data was sent from the Keybong to MAX, and on–off commands were sent to the Keybong motor from MAX. See the SonicTexting circuit diagram in Appendix A.

4.1.4. The process

The starting point for the SonicTexting project was Quikwriting. Quikwriting is a shorthand designed for use on stylus-based devices. It enables inputting text through stylus gestures without picking the stylus up off of the surface of the PDA screen. Quikwriting is based on a visual scheme of 9 regions. Writing a single character involves starting from a central region, moving between one, two or three peripheral regions, and returning to the center. The character is written upon the return from the last region to the center.

10 **The Quikwriting gesture scheme**



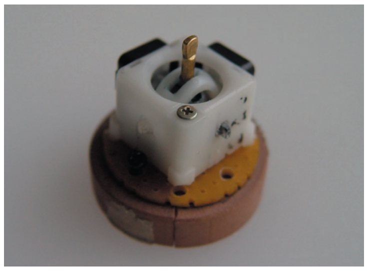
Quikwriting has an interesting gestural ‘vocabulary’. It is a complex task with high requirements, in terms of feedback and orientation, which needs to be performed in mobile situations. I decided to develop an audio–tactile text input interface using the Quikwriting gesture scheme as a starting point.

Input device

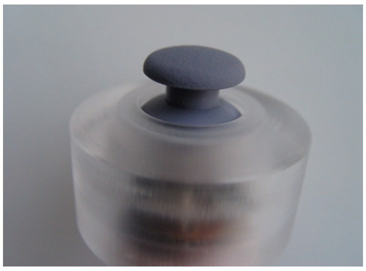
The goal in designing the input device was to create a one-handed controller to support the writing gestures. However, I was careful not to make this complex product-design challenge the focus of my thesis. Therefore, the process of designing the Keybong was not a thorough, ergonomic one but rather a search for the simplest form that would be enable performing the gestures comfortably with either hand.

Below is a series of images from different stages of this process.

11 Initial form idea for one-handed controller



12 Hacked joystick from a Sony Playstation (thanks to Graham Pullin for the idea)



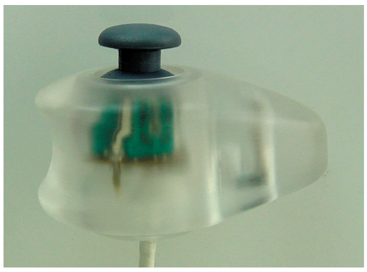
13 First working prototype – Keybong ver 0.1

14 Form explorations for preventing rotation

15 Keybong ver 0.2 with groove for the index finger

16 'Eccentric' motor added for tactile feedback

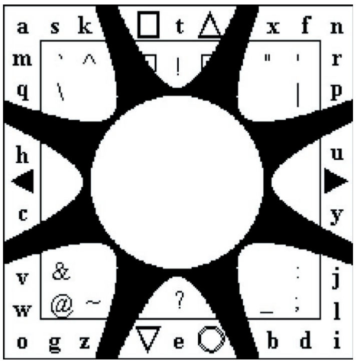
17 Keybong ver. 0.3

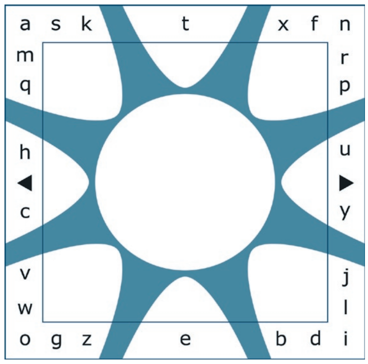


18 Alternative form with rounder bottom; better for large hands

Gesture Map

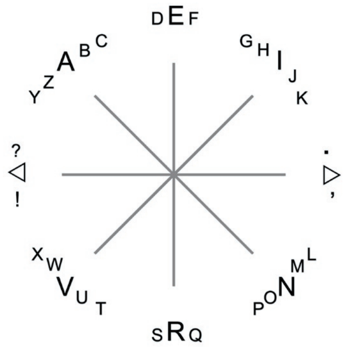
There were three main stages in the design of the gesture map: simplification of the Quikwriting template; transition to a round representation to correspond with the controller shape and an alphabetical letter order; and design of fractal-inspired map to communicate the nested nature of the model. An alternative representation model was tested (Version 4 below) in which the letter was written in the final location and color was used to indicate the path. It had advantages in instruction but was much harder to memorize since it required remembering two layers of information: color and position.

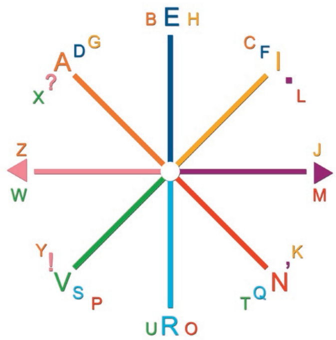




19 Version 1 – original Quikwriting map

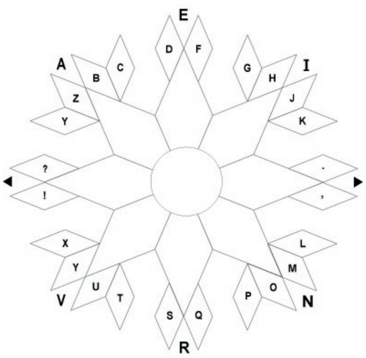
20 Version 2 – modified Quikwriting map with extra characters removed; font, colors and lines treated

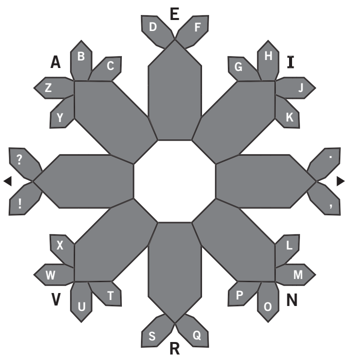




21 Version 3 – circular representation; ‘point and line’ model; alphabetical ordering; four punctuations added.

22 Version 4 – alternative representation model: letters appear in the destination position (not the origin),color indicates the outbound direction.





23 Version 5 – fractal-inspired design duplicating and rotating a basic element to communicate the ‘nested’ nature of the writing model.

24 Version 6 – design adjusted to also represent the direction of movement for the nested letters

Sound

Below are descriptions of the SonicTexting sound at different stages of the design process. Click on the ear icons to hear the word "sonic" written in each stage.

Vocal 1- a letter phoneme is played when the controller enters a region (e.g. – the sound "bh"). Upon return to the center, the complete letter is sounded (i.e. - the sound "bee").

Physical 1 - an attempt to sonify the 'physicality' of the Quikwriting template. Contains two elements: a 'border' sound that is invoked when the controller moves over a border between the regions and a 'friction' sound that is played constantly with the loudness and tempo corresponding to the speed of Keybong motion.

Musical 1 - a note is played while the control is within a region. The note remains on until the user returns to the center. In this way 'legal' letter paths create harmonic chords, while sequences of letters create note sequences – little tunes.

Musical 2 - ascending note order that corresponds to the new, alphabetical letter order. In this note order the notes ascends when moving through a-b-c. The musical version builds on peoples' proficiency in memorizing and recognizing musical tunes. Errors can be identified by noticing that the tune is wrong.

Physical 2 - breath sounds added: inhale on exiting the center and exhale on returning to the center. This was a way of strengthening the physical aspect of the gesture template by enhancing the transition between interior and exterior parts and relating it to the start and finish of the character input.

Vocal 1b - a percussive sound element was added to the Vocal 1 sound.

Vocal 3 - the 'Daniele' version – continuous, looped phoneme sounds with the amplitude determined by distance. Uses an Italian male voice.

Vocal 4 - the 'Silvia' version – continuous, looped phoneme sounds using an Italian female voice. Phonemes are sung in different pitches according to their location in the circle circumference.

Vocal 4b - extra character sounds added: space, backspace, period, comma, question and exclamation marks.

Vocal 4c - 'readback' feature added – when a whole word is written the inputted letters are read back in succession. This is the final 'beginner mode'.

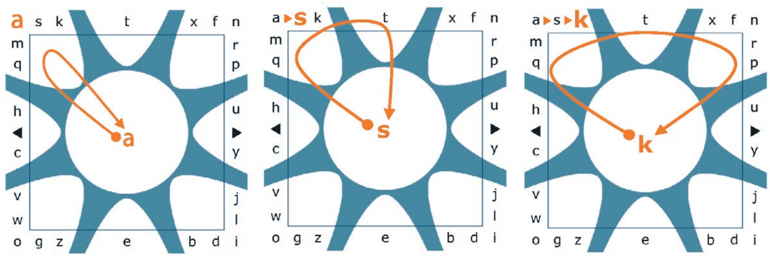
Musical 3 - percussive sound elements with a the same pitches as in beginner mode. This is the final 'expert mode'.

Mental model & Instructions

The inherent complexity in the Quikwriting / SonicTexting writing model created the need for ways of representing and explaining the model. The instructions and the initial exposure to the system determined the mental model users develop.

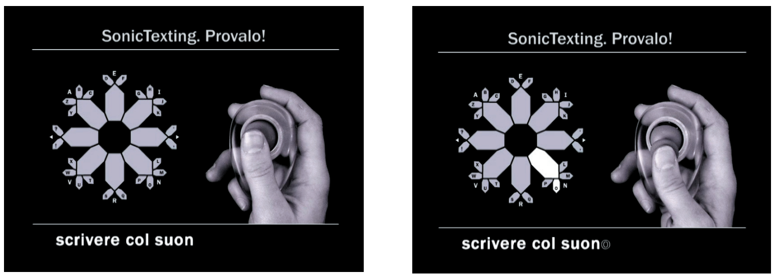
Instructions version 1 – gesture representations

This version was designed to strengthen the perception of the movement as a gesture path, as opposed to a discrete movement between regions.



Instructions version 2 – synchronized video display

This version was an attempt to create a complete, self-explanatory sequence that simultaneously presents gesture, visual map, sound and written letters. The synchronized animated elements on the graphic display were designed to visualize the connection between gesture and map.



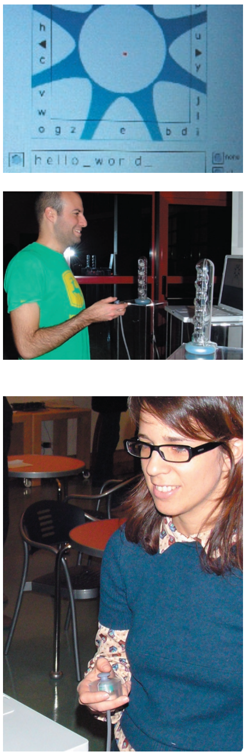
Instructions version 3 – Verbal demonstration

In this 'live' explanation method, the user was instructed by demonstrating a gesture with the Keybong, referring to the visual map, and – most importantly – directing attention to the relationship with the sound.

A typical explanation would be:

> "To write an A, move the controller in the direction where A is written and listen to the sound [pausing and listening together to the sound]. When you hear the strong and clear "ah" let the controller return to the center. The A is written."

User Observations



25 First user observation

First user observation - Interaction-Ivrea, January '03.

The project was installed in the institute gallery and tried by about 20 people. The following main insights emerged:

- Most people enjoyed trying SonicTexting.
- People liked the feel of the Keybong.
- The writing model was not easy to grasp.
- There were varied levels of success in using SonicTexting. Some managed to write well while others did not and preferred to play with the sound and movement.
- The round Keybong tended to rotate in people's hands. The groove indicating the top was not sufficient to maintain a consistent grasp.
- The visual feedback – a red dot indicating the position of the controller on the display – proved very dominant. It seemed to attract people's attention over the auditory.

26 Second user observation

Second user observation - Bu.Net bar in Turin, January '03.

The project was installed in the Bu.Net bar as part of an Interaction-Ivrea exhibition. It was tried out by about 60 people. The following main insights emerged:

- People enjoyed trying SonicTexting. Some tried and returned to try it again or to show friends.
- The explanatory diagram helped in explanation; still, some found the writing model difficult to grasp.
- The diagonal directions were relatively hard to reach; people often reached the horizontal or vertical points when aiming for the diagonals.
- People commented on the nice feel of the Keybong. The new form dramatically improved grasp consistency – the controller no longer rotated during use.
- The exact orientation of the Keybong in the hand sometimes needed to be corrected so that the Keybong 'top' calibrated with the display top.
- The continuous nature of the sound created a more coherent sound experience; however it seemed that the male voice was perceived as less pleasing than the previous, female voice.

Third user observation - 'Salone Del Mobile', Milan, April '04

The project was installed in the Triennale museum as part of an Interaction-Ivrea exhibition at the Salone Del Mobile. It was tried out by an estimated 1000 visitors! The following main insights emerged:

- Visitors liked SonicTexting. In this fair-like context, many treated it as a pure 'fun' experience.
- The understanding of the writing model increased dramatically. The vast

27 **Third user observation. About 70% of the people wrote their name with SonicTexting; others wrote "Ciao" or longer sentences**

majority understood the model after the basic verbal explanation. Others understood though watching the video, without further explanation.

- The degree of success in using SonicTexting increased dramatically. Most people managed to write their name; some composed whole sentences.
- The dependence on sound increased; this is probably due to the enhanced sound quality. Sound was played though a suspended 'sound dome' that created directional sound from above. This helped in focusing attention on the sound.
- People often made errors and corrected them. In most cases the errors were enjoyed because of the sound associated with the backspace function.
- People enjoyed the pitched vocal sounds and their tactile control over them.
- Some perceived the project as a sound installation.

4.1.5. Analysis

The work on SonicTexting spanned five months, multiple iterations and user observation phases. Though I had originally planned as a much shorter exploration, I found myself very motivated to make it really work: to bring it to a level in which people would both be able to, and enjoy, using. I feel that the current version of SonicTexting is very close to this goal.

Many different people have tried SonicTexting during this time: friends, exhibition visitors, teachers and external critics. It has been exciting to see that different people relate to it in different contexts: as a realistic and desirable mobile device feature, as an instrument or performance object, as a sound installation, as an accessibility project, and as an educational application.

In the following sections I analyze separately the different components of the work on SonicTexting.

The Keybong

Form. According to users the Keybong form is aesthetically pleasing and has a good feel n the hand. Most people, though not all, found it comfortable. The groove for the index finger and the oval form succeed in keeping the controller aligned to the top. However, the keybong was not ergonomically designed and has a number of problems:

- The main problem is the motion of the thumb. Moving the controller in the outward and sideways directions is comfortable, but moving it inward (to the bottom-right corner) requires the thumb to move backwards, a motion that is less natural.
- Not all users understand immediately how the Keybong should be held – the form does not directly imply this. While some grasp it well, some hold it in ways that make the circular thumb motion uncomfortable.
- Pressing the button felt a bit wobbly in positions in which the thumb was not directly above it.

As a next step, collaboration with an industrial designer can solve these issues by improving the ergonomic design of the Keybong.

Function. The main cause of letter errors in SonicTexting is imprecision in accessing the initial 8 directions or axis letters. This was improved, but not solved, by adjusting the algorithm to make the diagonal points more easily reachable.

As a next step, this problem can be tackled from different directions: through creating structural guidance within the Keybong form, tuning the sound feedback to clarify the initial stages of movement, using a joystick with better resolution and enhancing the software processing through gesture recognition algorithms.

The vibration feedback was designed to be the minimal perceptible strength. It is difficult to assess its importance. It is quite subtle, yet it gives the object a dynamic feel, and some users reported enjoying this very much. It is difficult to estimate how much users come to rely on it. More serious testing of different vibration lengths and behaviors would be necessary.

An interesting feature emerged serendipitously in one of the Keybong versions. The fastening of the motor to the Keybong interior was not completely firm; therefore a part of the vibration was translated into noise. This accidental feature resulted in a sound with a ‘symptomatic’ feel: a mechanical response to the acquisition of a point/letter. Given more time I would explore this further.

Sound output

Working on the SonicTexting sound was a long learning process. The most significant realization was that the real effort in the sound design was to edit out sound elements and achieve the most elegant and economical sound that would convey the information. At the end of the process, the beginner mode contained only one sound layer (phonemes) in which loudness, degree of overlap and pitch are used to represent the additional information elements. The expert mode’s discrete percussive sounds are an even more minimal representation of this information in terms of sound.

People’s use of sound. There were extreme differences between people in the degree to which they made use of the sound output. In general, the more people relied on the sound, the better they succeeded in using SonicTexting.

Gesture Map

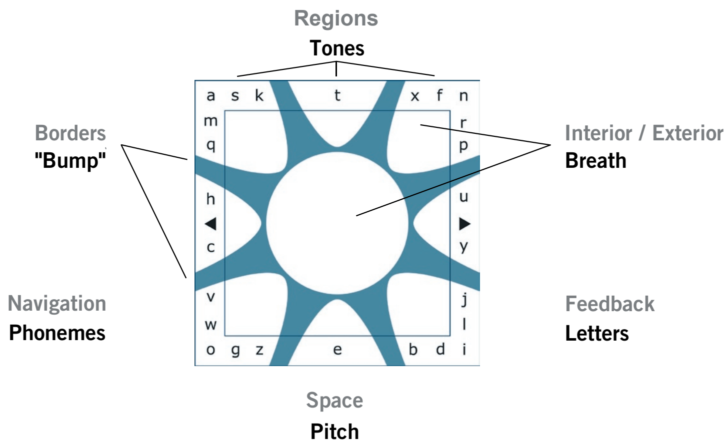
The Quikwriting writing model used in SonicTexting is not completely intuitive. Understanding the connection between map and gesture requires grasping an abstraction in the representation of the nested letters. The smaller area in which these letters are written does not conform to the size of the movement to be made: the actual movement needs to be bigger, spanning over the adjacent circle points.

The final iteration of the gesture map was the most successful in both aesthetic and explanatory terms. Many people could figure out the abstraction by trying and understanding the mapping via the sound.

Still, the most common problem in the interpretation of the map remained the literal reading of it, in which users tried to perform small movements to reach the nested letters. A quick verbal explanation was needed to remedy this.

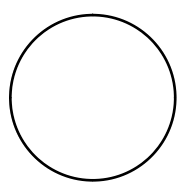
Mental model & instructions

My first approach to creating the audio-tactile interaction for the Quikwriting letter scheme had been to identify all the elements composing the Quikwriting scheme and to assign each of them a sonic element.

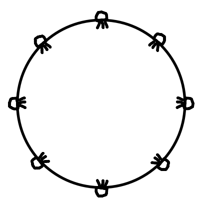


The SonicTexting system works in this way: when a main / axis letter is acquired, the system enters into a mode in which all points are set to new letters. Reaching the nested letters is done by moving to the adjacent point positions. Having implemented the algorithm in this way, I thought that this is the model to be communicated to the user.

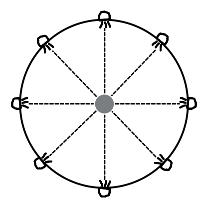
Interestingly, this is not true. Talking with some users, I realized that they did not have this mental model: their model was sound-based! They moved to the axis letter, and then rotated until they heard the right sound. They did not think of the position they reached as being in the visual space, but in the auditory space. According to this important insight I fundamentally changed my design of the sound behavior and the instructions for SonicTexting. The new model is dissociated from the visual scheme – it is an audio-tactile model. The model is described through six principles below.



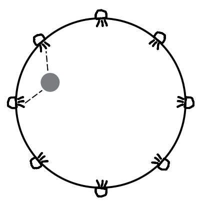
It's an Auditory Space



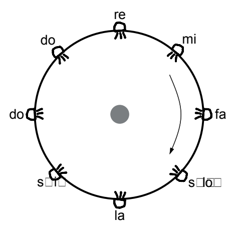
With Eight 'Letter Machines' Around It



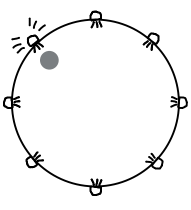
The Closer You Are, the Louder



In-Between, There Is overlap



Every L-Machine Has a Voice (Pitch)



Letter Machines 'Stir' When Touched

Lessons from User observations

Incidental users. SonicTexting was used at the Salone Del Mobile for a whole week, and most visitors succeeded in writing after a brief explanation. People's enjoyment was evident by their verbal reactions, the time they spent and their desire to make sure their friends try it. Some used it to play and flirt; one couple wrote a teasing dialog: "Erika, don't bug me"; "Claudio, you are free".

Most fun was watching children use SonicTexting. Eleanor, aged 8, wrote her name with relative ease, then proceeded to ask for, and write, my first and last names, the name of my sister and more. All children of reading age succeeded in using SonicTexting. Pre-reading-age children played with the Keybong and responded with pleasure to the sounds it made.

It was exciting to see that a number of my colleagues who were presenting at the same exhibition used SonicTexting once in a while as a kind of a pause or play. Since the written text was projected in large scale type, it was sometimes used to write messages for the staff at the other side of the exhibition space (e.g. "We are hungry!").

Longer-term users. Naturally, I am the person who has used SonicTexting most extensively. When focused on it, I can use it almost without error, at a quick pace. I have become very proficient at using the sound. The only problem I encounter is an occasional error in accessing the 8 axis directions. I have not completely memorized the letter scheme and continue to refer to the map when I have it. When I use SonicTexting without the map I work with the sound to find the letters. The pitch progression proves to be a significant memory aid: I remember the sequence in which the letters are sung, and infer the required movement direction from it ("Nnn, ohh" are sung with a one-note progression, meaning 'O' is further in the clockwise direction). A friend who used SonicTexting quite often shows the same patterns and problems.

Experience with a blind user. I was very happy for the opportunity to get an opinion on SonicTexting from Emmanuelle, a blind person who was interested in the idea. I was eager to both see how he would use it, given his strong dependence on the audio-tactile connection, and get his advice concerning the possibilities of creating a completely non-visual version.

Emmanuelle spent about two hours trying the final SonicTexting version and discussing it. He was extremely enthusiastic about it. However he was only partially able to use it in this time frame. In general, all problems with the system were intensified by this non-visual use. Specifically, difficulties were created by the lack of a non-visual gesture map, the need to align the controller top and the lack of feedback on letter writing in this version.

During the trial we developed the following strategy for finding the letters: he would say the letter he wanted; I would tell him the axis direction (top-left) and then the direction to move (move right).

The most important feedback from this experience concerned the mental model. Emmanuelle was very convinced that he would not want a Braille map of the gesture patterns (“what for?” he asked). He said that given some time with SonicTexting he would learn the directions by sound. This was for me a strengthening of the notion of an audio-tactile mental model. I hope to have the opportunity for Emmanuelle to use SonicTexting over longer time and to better understand his mental model concerning the audio-tactile space.

4.2. Additional Explorations

4.2.1. Shake (with Mathias Dahlström)

Overview
Shake is an audio-tactile system for sharing music between music devices. It uses gestures to perform the social interchanges involved in sharing music with others. We have mapped these emerging sharing functions into a gesture vocabulary performed with the device. The gestures involve arm motions such as shaking and sweeping. The gestures are accompanied by sound – real-time modifications of the played music in synchrony with the gesture - exploring the natural connection between music and body movement. The modifications are achieved by applying filters and effects over the music in ways that at once provide feedback for the gestures and enhance the expressive experience of performing them.

Motivation - Mobile Music Sharing
It seems mobile music sharing is the ‘next thing’, fundamentally transforming the music player from a disconnection device to a connection device. The Media Lab Europe has developed the Tuna project. Tuna is a “mobile wireless application that allows users to share their music locally through handheld devices. Users can ‘tune in’ to other nearby tunA music players and listen to what someone else is listening to”. Mobile music sharing is moving from the research labs to commercial product: the latest Samsung MP3 player has a built-in FM transmitter, so you can broadcast what you’re listening to. Similarly, the Apple iPod has a plug-in FM transmitter called iTrip. It seems our music players will soon be communicating with each other. But how will we actually perform the interactions involved in sharing music?

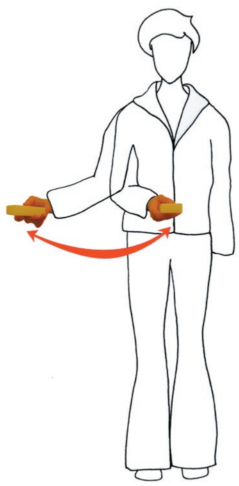
We developed an exploratory working prototype of Shake. Because of time constrains this prototype was not fully tested and is presented as a work in progress. I describe both the prototype and the design process that led to it below.

The process

Gesture Map. In our exploration we imagine a public space where people have a line of sight between them and use the device in a ‘flirtatious’ way, to communicate their intentions and negotiate the sharing of their music by gesture. After short sessions of experience prototyping between us and with fellow students, we identified seven different activities in this sharing vocabulary.

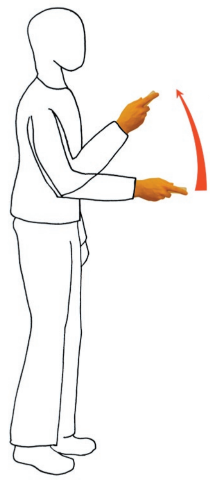
Sweep - Scanning the player space for good music
Alert - Getting someone's attention
Ask - Asking someone to share what they are listening to
Offer - Offering someone to share what you are listening to
Give - Giving music
Receive - Receiving music
Lose - Getting rid of music someone gave you

For each of the elements described above we have defined a gesture archetype.

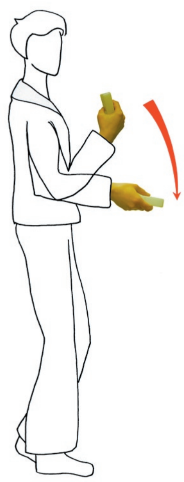


The 'Sweep' gesture

The 'Alert' gesture

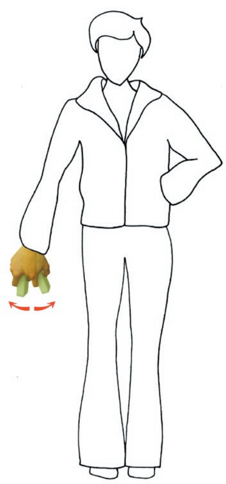


The 'Ask' gesture



The 'Offer' gesture

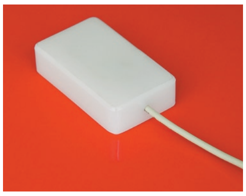
The 'Give' gesture



The 'Lose' gesture

Sound. For each of the gestures described above we defined a sound metaphor and behavior.

Gesture	Metaphor	Sound behavior	Implementation
Sweep	Tuning in to others	Moving between stations – noise and signal	White noise and music mixed according to pointing direction
Alert	Nagging, like a tap on the shoulder	'squeaky' up and down nag	Notch filter corresponding to tilt
Offer	Sound trying to escape from its container	Sound collecting in the container	Lower frequencies enhanced as device is tipped back
Give	Pouring /rolling / shooting out	Sound shooting out of its container	Looped playback of sound segment with decreasing reverb
Lose	Shaking something off	Sound dies out	Volume falls in 'waves'
Ask	Sound 'making space' for another	(not implemented)	(not implemented)
Receive	Filling up	(not implemented)	(not implemented)



28 The Shake input device – abstract controller with iPod dimensions

Controller. We created a physical object in the form factor of the iPod music player, to be held in the hand. Inside it we fitted a 2-axis accelerometer to detect acceleration/tilt. We indicated the top of the device; otherwise it was a completely blank, abstract device.

Instructions. The initial prototype of Shake does not contain gesture recognition; a user interface enables the user to select the gesture, then to perform it and experience the sound feedback. The interface contained a sketch of the selected gesture as a simple form of instruction.

Software Implementation. A microcontroller processed the accelerometer readings into values which were fed into Max/MSP. Each gesture was associated with a different filter combination in Max according to the defined metaphors. The accelerometer values were fed into the filters to produce real-time modification effects.

Analysis. The work on Shake lasted about 4 weeks. From this short process we concluded that there is significant potential for developing the concept of gestural music-sharing with sound modification further. Given the limited time and technical challenges - the analysis of the accelerometer data and the implementation of the real-time filters in MAX – the quality of our prototype was restricted. Since the audio-tactile experience depends on the synchronicity and richness of the sound-gesture connection, the prototype of Shake was only partially useful as a test of the Shake concept.

In our implementation of the Shake effects we sought to create sound modifications that were subtle, that affected the music but maintained a certain ‘respect’ for it. We chose not to use ‘cartoonish’ sound effects; we modified the sound but did not to insert external elements.

No formal user testing was performed with the Shake prototype due to the limitations mentioned above. However, a number of colleagues and teachers tried it and provided feedback. In general they were enthusiastic about this experience of connection between gesture and sound. Once people discovered their effect on the sound, they found it very enjoyable and tended to play with the effects in ways that matched to the beat of the music.

One problem was the subtlety of the effects – some were not perceived strongly enough. It seems that the trial situation, in which we used external audio and not headphones, required stronger modifications and perhaps using more ‘introduced’ sound elements.

Shake has only scratched the surface; future work will need to address issues on many levels. On the audio-tactile level, it will be important to test the existing and additional sound modifications. The extent to which different modifications can be effective over varied types of music is a central question to tackle.

4.2.2. Ring

Overview

Ring is an audio-tactile system for composing the ring to be heard on the recipient’s phone when making a call. It uses gestures to produce a real-time sound that can contain both information and expression, making the ring a unique, non-verbal communication piece. Ringing gestures are made with the fingers using a touch-pad surface on the exterior of the phone device. Touching the pad at different locations produces different sounds. The pad space is mapped to a sound ‘palette’. As in IDEOs knocking SOMO, the interesting aspect of Ring is in the potential for the development of personal ‘languages’ of ringing.

Motivation - Ringtones and personalization

In 2003 consumers globally spent somewhere between $1 and $3.3 Billion on ringtones. People’s desire for personalization is such that some even use world famous speech excerpts from John F. Kennedy, Nixon, Reagan and others as their ringtones.

The interest in personalizing phone sound can be seen in the latest mobile craze, RingBack tones. RingBack tones replace the boring “bring-bring” sound every time the call connects with the user’s choice of music track. According to the site Ringtonia.com, the RingBack service has seen the fastest take up of any wireless service to-date.

Some mobile phone users currently take the effort to couple a specific Ringtone with a specific person. This creates an enhancement of both the informational (identifying the person through the sound) and expressive (associating the person with a certain sound) aspects of the phone ring.

Ring is a conceptual exploration; due to time constraints, it was not developed beyond a preliminary prototype. The process and preliminary prototype are described below.

The process

Concept explorations with users

29 **Notebook given to users for documenting mobile communications in Ring**

Step 1 – documenting communications. An initial exploration was performed to create a documentation of the kinds of communications people make with their mobile phone. I asked three people, all young Italian people from Ivrea, to document every call or SMS they made over a period of three days in a small notebook I gave them. For every entry they answered three questions: “When did I call?”, “What was the purpose of the call?” and “How did I feel?”

The results showed that the types of communications could be very generally divided into two main categories that I named ‘coordinations’ and ‘friendlies’. Coordinations are communications such as “I’ll be late”; “when is the train?”, “do we meet after work?”, “where are you?”. Friendlies are communications to say “just hi” “I love you” or “how are you?”.

Step 2 – exploring audio tactile communications. In this step I asked two of the people that had recorded their communications to imagine how they would use an audio tactile ring interface for these communications. For each entry in the notebook I asked them to show me, with their finger and their voice, the way they would use the Ring system if it existed. The finger was moved over a vertically placed white surface; the voice was to be used freely except for the request to use only non-speech sounds.

The users found the activity a bit unstructured, and had some initial hesitancy in ‘singing’ out loud. This subsided after a few entries. As they went through their notebook, both developed patterns in their use of sound and of the pad space. Each developed a different, but internally consistent, gesture language.

Gesture Map. Differently from the other explorations, the gesture map in Ring is developed by the user. The pad provides a constrained space in which the users create their own language.

30 **The Ring input device - phone with embedded touchpad**

Controller. I used an old, rather large mobile phone for the preliminary prototype of Ring. A USB touchpad was embedded into the exterior of the phone (image). The phone in this prototype was not functional as a communication device; it was used to achieve an authentic feeling when performing the ring gestures.

Analysis. The concept explorations of Ring showed a clear potential for the creation of audio-tactile languages for communicating information and expression with the Ring. This is a very encouraging first step; however, much prototyping and testing is needed to determine whether, and how, people would want to use Rings. These future steps will enable an analysis of Ring which is now premature.

KEY FINDINGS

Looking back, this thesis has followed two layers of investigation. One concerns the specific qualities of auditory interaction and the audio-tactile connection, these were studied most deeply in SonicTexting; the second concerns the concepts of audio tactility in mobile devices, this was addressed by the combination of the three explorations.

I will now present the key findings from these two layers.

5.1. The audio-tactile connection

Informative sound
Sound was successfully used in SonicTexting to provide complex information about navigation, orientation and the outcomes of actions. People could process and use the multiple aspects of sound for this purpose.

Expressive sound
Throughout the work on SonicTexting I struggled over the question: "What is the sound of SonicTexting? What is its sonic character?" I searched the nature of the activity, the Keybong form and other aspects for answers.

In parallel, the sound design of SonicTexting was a constant process of reduction and editing to find the sound combination that would contain all information layers while creating an enjoyable and not-too-dense sonic experience. In retrospect, I find that this process of 'editing out' provided a large part of the answer to the question of the SonicTexting sound character. Essentially, I found that the informative aspects of the sound defined - to a large extent - its expression.

Audio-tactility
I think that both SonicTexting and Shake have the qualities of audio-tactility. People like using them and the connection between gesture and sound s a large part of this. The experience is perceived as continuous and synchronous, the sound and gesture are a naturally unified entity. As one SonicTexting user put it:

> "Sound is a natural part of using the Keybong. It rounds the experience, makes it harmonic".

Dependence on sound
SonicTexting showed that people can successfully and enjoyably use sound as a source of information; one user described the change in experience after the visual feedback was removed: "a breakthrough of pleasure". However, using sound it

is not people’s first tendency. If presented with parallel visual information, this modality will dominate, even if the auditory information is of a better quality.

It seems that the more the sound is central in the way the experience is constructed, the easier it is to cause people to depend on it. Instructions also have a large role in getting people to use the sound. A combination of manual demonstration and explicit reference to the sound proved the most effective instruction method for the SonicTexting audio-tactile interaction.

Audio-Tactile mental models

SonicTexting exists in three spaces: the visual space of the gesture map, the tactile space of the Keybong motion and the auditory space of the sound experience. As described above, the sound design and implementation were built thinking predominantly about the auditory space. It is my initial impression that users created a mental model of SonicTexting that is at least partially audio-tactile. As described above, people came to depend on their hand-ear coordination to find the nested letters, as opposed to the hand-eye coordination of the visual map. One user compared between using the Keybong and using a keyboard:

> “With a keyboard, the space is laid out in front of you; with the Keybong it is more abstract: the space is in the head, not in the Keybong”.

Mastery

SonicTexting created an aspiration for mastery in some of its users. A number of people studying with me said they would like to ‘take it home’ and use it for a longer period of time. When installed for a week in the Milan exhibition, some co-exhibitors ‘practiced’ SonicTexting every day, reporting to me about their progress. Future work is needed to see how well people master SonicTexting with long-term use and how much they continue to enjoy it.

5.2. Audio-tactility in mobile devices

This thesis suggests audio-tactile interactions for three diverse mobile device activities. I have not tested these interactions in a real context of use (e.g. writing real text messages). The questions I have asked in this thesis are: can an audio tactile-interface be designed for this activity? Can it feel good? And does it ‘make sense’? The first two questions were discussed above; I will now discuss the third.

Audio-Tactile and Mobile?

People’s responses to SonicTexting indicate that it made sense. I was often asked such pragmatic questions as whether SonicTexting would work with the current phone joysticks, whether it would be used with headphones, and most importantly, when it would be available. The concept of ‘texting‘ from the pocket appealed to many people. Responses to the concepts of Ring and Shake were also favorable.

More generally, it seems that many people (though clearly not all) are open to, and seeking, new ways of using mobile devices. As one visitor to the Milan exhibition put it:

> "Looking at our devices you could think we have only one finger. It's time that devices became more interesting".

Functional, instrument, game

The variety of responses to SonicTexting and Shake: as real products, as musical devices and as play-things, show that these projects succeeded in being designed 'on the border' as I had intended. I feel that their state of being 'on the border' implies a certain quality or richness.

Feasibility & potential

In my opinion audio-tactile interactions with mobile devices, such as the ones presented in this thesis, are feasible and interesting in business terms. The potential for enjoyment, for long-term mastery and for play suggest that audio-tactile products could have market success.

It is outside the scope of this paper to discuss the process in which these explorations could become products. However, it is important to note that there is much flexibility in the way devices could become more audio-tactile. Taking SonicTexting as an example, this functionality could be implemented in a variety of ways. To name three:

- The Keybong, as designed, could be an accessory object that communicates via Bluetooth with a current mobile phone
- The Keybong joystick could be integrated into mobile phones of the current form.
- The Keybong could be a complete 'texting' device in itself, with the communication functions embedded within it. In this case the writing feedback would be only auditory.

Finally, text entry is a very complex activity; this project has convinced me that if this task can be a successful audio-tactile interaction, there is real potential for other audio-tactile mobile device activities at different levels of complexity.

CONCLUSIONS

The work on this thesis has taken me on a multidisciplinary path through sound design, mental models, electronics, software, product design, information design and more.

Due to this nature of the thesis, I feel that the findings presented in it have relevance to many different fields. Within interaction design, there are implications for auditory interfaces, for multi-modality, for the concepts of direct manipulation, for accessibility, and, of course, for mobile device interfaces. There are also possible implications or ideas for other fields: information sonification, interactive sound experiences and to some extent education and new musical instruments.

I started this thesis with an interest and enthusiasm for audio-tactility. I am finishing it with a strong belief in the power of audio-tactility as an experience and as a form of interaction with digital technology. I am impressed with the way people can use sound as information in complex and sophisticated ways and think that there is space for objects that require skill and mastery even for everyday tasks.

I feel that this thesis only touched on the questions of real audio-tactile interactions for mobile devices, due to a relative focus on depth (SonicTexting qualities) over breadth (the three explorations as a whole). However, I am convinced that the potential for implementing audio-tactile mobile devices is strong. Next steps on this path will involve more research, more prototyping and longer-term testing of the current ideas – especially of Shake and Ring. These should be tested in real use scenarios and fine-tuned to specific user populations.

On the more sensual level of audio-tactility, I am excited by the quality of the audio-tactile experiences it was possible to reach in this project given the limited level of software and hardware prototyping. Playing with the Keybong and exhibiting it over the last months, many ideas have come to my mind and have been proposed by others who tried it:

- What would it sound like in different languages?
- How would it feel in languages with sounds I can not make with my mouth?
- What other forms could the Keybong take, and how would that affect the sound design for it?
- How could the vibration and the sound complement each other in more complex ways?
- SonicTexting was used extremely well by children. How could this experience be used in an educational in relation to the process of teaching reading and writing?
- SonicTexting is a way to create vocal sounds using the hand. Could it be used – as many who tried it pointed to – not SonicTexting but "TactileSpeaking": a way of speaking through gesture?

These are only a few of the possible directions that the audio-tactile experience of SonicTexting has sparked. I hope that this potential of audio-tactile experiences may also inspire other designers to work with the connection between hands and ears.

REFERENCES

Ackerman, D. A Natural History of the Senses. Random House, 1991.

Anderson, L. The Record Of The Time – Sound In The Work Of Laurie Anderson (2003) Mazzotta.

Bassoli, A. Cullinan, C., Moore, J., Agamanolis, S. “Tuna: A Mobile Music Experience to Foster Local Interactions”. (Poster), Presented: Ubicomp 2003 Fifth International Conference On Ubiquitous Computing, Seattle, 12 - 15 October 2003.

Bongers, B, Van Der Veer, G. “Tactual Articulatory Feedback and Gestural Input”. Paper available at: http://www.cs.vu.nl/~bertbon/Downloads/ArticulatoryFeedback.pdf.

Brewster, S.A. “Navigating Telephone-Based Interfaces with Earcons”. Proceedings of BCS HCI’97, Bristol, UK: Springer, (1997): 39-56.

Buxton, W. Gaver, W. Bly, S. “Auditory Interfaces: The Use of Non-Speech Audio at the Interface”. Unpublished Book Manuscript. Available at: <http://www.billbuxton.com/Audio.TOC.html>

Buxton, W. (in progress). Human Input to Computer Systems: Theories, Techniques and Technology. Unfinished book manuscript. Available at: http://www.billbuxton.com/input01.Introduction.pdf

Chion, M. Audio-Vision. New York: Columbia University Press, 1994.

Gaver, W. “The Sonic Finder: An Interface That Uses Auditory Icons”. Human Computer Interaction 4 (1989): 67-94.

Gaver, W. “Auditory Interfaces”. In: M. Helander. T Landauer, & P. Prabhu (Eds.), Handbook of Human-Computer Interaction, 2nd ed. (1997): 1003-1042. Amsterdam: Elsevier.

Harrison, B. et al “Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (1998): 17-24.

iPod <http://www.apple.com/ipod>

iTrip <http://www.griffintechnology.com/products/itrip/>

Jones, C. and IDEO. "Social Mobiles" Available: http://www.ideo.com/case_studies/social_mobiles/menu.html

Kramer, G. et al, "Sonification Report: Status of the Field and Research Agenda". Prepared for the NSF. Available at: http://www.icad.org/websitev2.0/references/nsf.html

Long, Jr., A. C., Landay, J. A., Rowe, L. A. "Implications for A Gesture Design Tool". Human Factors in Computing Systems (SIGCHI Proceedings). ACM, ACM Press, (1999): 40-47.

MAX/MSP <http://www.cycling74.com>

McCullogh, M. Abstracting Craft: The Practiced Digital Hand. Massachusetts: MIT Press, 1998.

Nokia 3300 Phone <http://www.nokia.com/nokia/0,,5819,00.html>

Perlin, K. The Quikwriting Shorthand System. <http://mrl.nyu.edu/projects/quikwriting>.

Pirhonen, A., Brewster, S. And Holguin, C. "Gestural and Audio Metaphors as a Means of Control for Mobile Devices". Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York: ACM, *2002): 291–298.

Plant, S. "On the Mobile - The Effects of Mobile Telephones on Social and Individual Life" Report prepared for Motorola. Available at: http://www.motorola.com/mediacenter/industry/background/1,1083,,00.html

Ring Back tones <http://www.textually.org/ringtonia/archives/003081.htm>

Rocchesso, D., Bresin, R., and Fernström, M. "Sounding Objects" IEEE Multimedia, 10(2): 42-52.

Sider, L. Freeman, D, Sider, J. (Eds.) Soundscape - The School Of Sound Lectures 1998-2001. London: Wallflower, 2003.

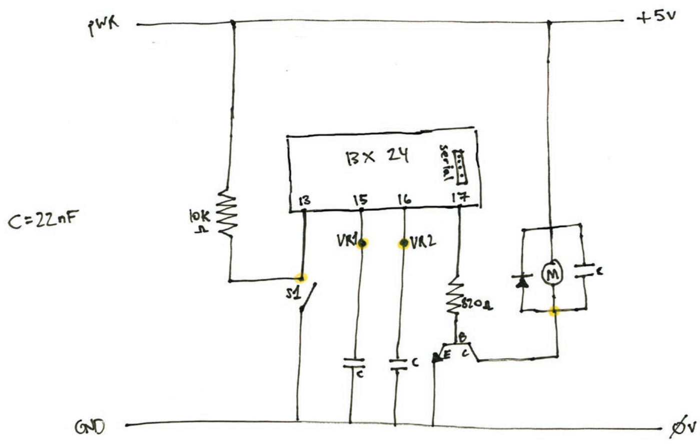
Townsend A. M., "Life in the Real-Time City: Mobile Telephones and Urban Metabolism". Journal of Urban Technology. 7 (2000): 85-104.

Wessel, David, Wright, Matthew, "Problems and Prospects for Intimate Musical Control of Computers" ACM CHI 2001 Workshop on New Interfaces For Musical Expression (2001) Seattle, USA.

Verplank, B., Sapp, C., Mathews, M. "A Course on Controllers", ACM CHI 2001 Workshop on New Interfaces For Musical Expression (2001) Seattle, USA.

APPENDICES

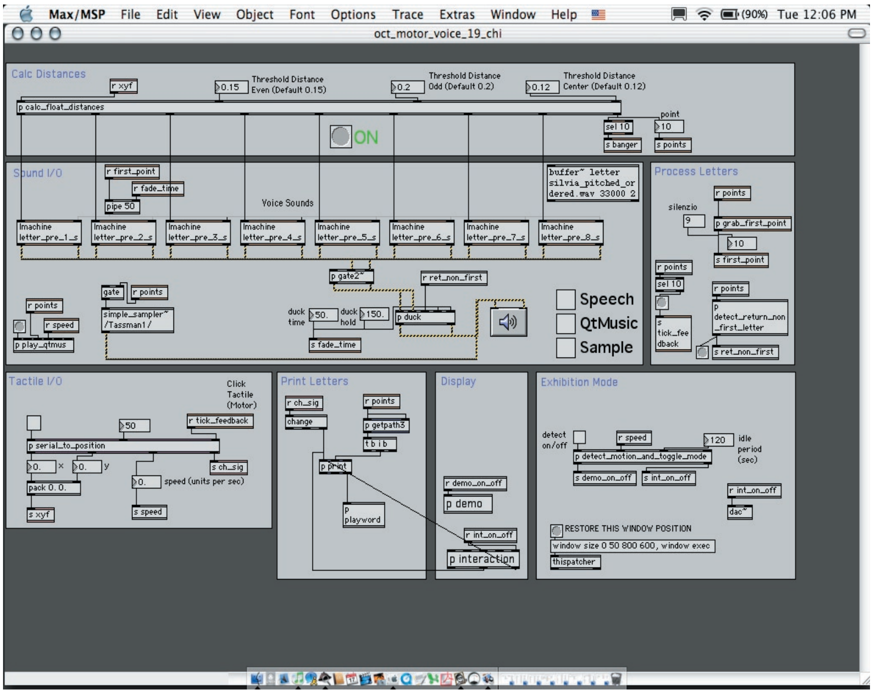
Appendix A – Circuit diagrams for SonicTexting



SONIC TEXTING-KEYBONG – CIRCUIT

Appendix B – Max/MSP implementation of SonicTexting

31 Max/MSP patch with the eight 'letter machines



32 'A 'letter machine' patch playing an 'R' - showing the curve that describes the volume of the letter machine as a function of distance