Sonified Interactions with Mobile Devices

Michal Rinott, Interaction Design Institute Ivrea

***Abstract*— this paper presents a framework of thought concerning interactions with mobile devices that involve manual input and auditory output ("audio-tactile" interactions). The potential of audio-tactility for creating rich interactive experiences is discussed via insights from the physical world and from the role of sound in film. An exploratory audio-tactile interface developed for the "Quikwriting" shorthand writing application is described in detail.**

***Index Terms*—Interactive Systems, Audio-Tactile Interaction, Mobile Devices, Sound Design**

I. Introduction

When we interact with physical things in the world, these interactions often 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).

In the world of digital technology the situation is usually very different. The digital landscape is often mute, or perhaps worse, dominated by flat, annoying sound. If our interactions produce auditory responses, they are usually abrupt, on-off phenomena. It seems that the richness of sound in everyday experience is lost in the digital domain, and people's natural expertise at learning from sound is unutilized.

This paper proposes that sound can be used in richer and more effective ways in digital objects *by strengthening the connection between action and sound*. This type of interaction, audio-tactile interaction, involves tactile (manual) input and auditory output. This paper proposes ways to enrich these interactions by turning them into carefully designed experiences with a strong and sophisticated relationship between action and sound.

Mobile devices are portable objects designed to be carried and used "on the move". Mobile devices include communication devices (e.g. mobile phones), information devices (e.g. PDAs), music players and more. These devices are designed to be compact and yet to provide complex functionality. Pironen et al [8] have recently described the challenges this creates for interaction: "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….Multimodal interaction is one way to enhance usability… flexible multimodal interfaces may allow users to interact in a more natural manner".

Mobile devices enable a wide range of functions – for example, browsing lists, inputting text, navigating menus, controlling playback. This paper will present an agenda for exploring some of the qualities of audio-tactile interaction through functions of mobile devices.

The paper begins by describing the qualities and the potential for richness in audio-tactile experiences. It proceeds to examine separately possible sources of richness in the auditory component and in the tactile component of this interaction. The final part of this paper describes the initial design of a sonified interaction for a text entry system on mobile devices.

Manuscript received December 15, 2003.

Michal Rinott is at the Interaction Design Institute Ivrea, Via Montenavale 1 Ivrea (TO), Italy. +39 0125 42211; fax: +39 0125 422101; e-mail: m.rinott@interaction-ivrea.it.

II. Audio-Tactility

Everyday experience with physical objects is multi sensory. We do not perceive our experiences as fragmented into different sensory channels, but as complete, unified wholes. Think of 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.

Film sound theorist Michel Chion [2] has made the following statement concerning sound in film: "there is no soundtrack". An extreme statement coming from a researcher of 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 Chions reasoning, I wish to suggest that "there is no soundtrack" for digital objects, or at least there should not be. Our perception of the sound of objects is linked to the actions that produce the sounds. Thus, as audio and vision are perceived in unison, so could audio and tactility. When we interact with digital objects, the sound they make should contribute to the holistic experience of the objects. Well designed digital objects should have the quality of making the sound inseparable from the interactions that invoke it.

Following Chion, I will refer to this quality of experience as "audio-tactility".

How can this ambitious goal be achieved? The next sections look at the two components – the auditory and the tactile, considering ways in which each can contribute to achieving unified audio-tactile experience through rich relationships between hand and ear.

III. THE AUDITORY COMPONENT

Which qualities of sound can contribute to the unification of the audio-tactile experience? This section points to several sources of inspiration: the qualities of sound in physical interactions, the use of sound in film, and scientific information sonification.

A. *Introduced sounds and Consequential sounds*

I refer to *Consequential Sound* as sound that is a by-product of the natural, physical interaction with an object: a meeting between physical materials. I refer to *Introduced Sound* as sound that is introduced into an object for communication purposes. An example of consequential sound is the click of the lock when closing a briefcase. An example of introduced sound is the "click" sound often used in digital camera interfaces in response to pressing the shoot button.

There is a central difference between these two types of sounds. Consequential sounds contain inherent variation. The locking sound of the briefcase will be slightly different every time, depending on the exact closing motion and on other factors in the environment. People perceive this variation and, I believe, enjoy their implicit expertise in disceming the correlation between physical action and the auditory response. Conversely, Introduced sounds typically contain no such variation. They are binary signals – fixed responses to fixed, defined conditions. It is my belief that this is a main reason for the dislike people feel towards many of the sounds that digital objects make. These sounds contain a low ratio of data-to-sound: they demand attention but reward it with a relatively small amount of information.

Sounds of digital systems are predominantly introduced sounds, "artificial" phenomena. But does this mean they must be simple, binary signals? The answer is definitely negative.

Bill Gaver's early "SonicFinder" interface [3] is a strong example of sophisticated introduced sound that contains multiple layers of information. The SonicFinder was a sound application built over the Macintosh Apple file system. For example, in the SonicFinder a dragged folder sounded different according to its size and type. Sonic information layers can be of different complexity, some immediately recognizable and some requiring more prolonged use and acquired expertise. Layers 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 present in the direct experience of the interaction (as the file size sound layer in the SonicFinder).

B. *The added value of sound - Informative and Expressive*

Michel Chion has analyzed the *added value* of sound in film. In this analysis, Chion refers sound as having both *informative* and *expressive* added value.

I feel that these categories of informative and expressive are beneficial in the analysis of the use of sound in interaction design. It should be noted that these categories are non-exclusive; in fact, I believe interactive sound must be both informative and expressive in order to be effective.

Informative Value – the field of scientific information sonification demonstrates perhaps the most sophisticated use of the informative value of sound in design. In information sonification, multiple layers of information are mapped into acoustic dimensions. Well designed sonifications employ mappings that maximize peoples' perceptual abilities, such as the capacity for parallel listening and the ability to discern relationships and patterns in auditory material [5].

Expressive Value – film sound creates expressive value in extremely sophisticated ways. Sound designer Randy Thom listed some of them: sound creates the mood of a scene through music and external sound; sound can direct attention to, and from, different elements of the scene. Sound can affect the perception of the pace of a scene, rendering it faster or slower than it "really" is; it can give character to people, objects and abstract concepts within the flow of the scene through leitmotifs and other sound manipulations. [9].

The ability to give character through sound holds special potential for added value. It can give an interaction a "sound and feel" (akin to the familiar "look and feel") that might make the difference between an anonymous experience and a satisfying, desirable one. The challenge this creates is significant: how does one select the "right" character for a sonic interaction?

Even more complex relationships between visual display and sound are employed in film sound design. For example, sound does not always conform to the visual display - it can sometimes even be used in a subverted way, to confuse or alter the perception of what is seen. In summary, can the sounds designed for digital objects employ some of the principles of film sound to achieve degrees of richness and sophistication?

IV. THE TACTILE COMPONENT

Which qualities of tactility can contribute to the unification of the audio-tactile experience? This section points to two sources of inspiration: the use of the natural human propensity for manual skill and expertise, and the selection of gestural languages that have a natural relationship with the actions that they invoke.

A. Handicraft

According to Malcolm McCullough [6], our hands are currently underrated and underemployed in the use of everyday computing. Clearly, buttons, pointers and even the mouse represent a small and limited range of what our hands can do. McCullough sees the potential in using technological objects in ways that support the subtleties of the hand.

In line with McCullogh I wish to suggest that creating rich interactions in tactile terms – that make use of the skills of our hands – can make the interactive experience a more desirable one. Skill, according to McCullogh, is a source of satisfaction, and "develops an intimate relation with certain contexts or tools, which makes it individual".

A clear inspiration for this kind of tactile manipulation can be found in the way hands are used in playing musical instruments, and, though differently, in the way gaming controllers are used. These are situations where the skill in using an instrument – a tool – is a source of pleasure, both from the outcome and from the process. McCullough quotes Octavio Paz on craftsmanship: "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".

Clearly interactions with mobile devices should not require the kind of expertise associated with musical instruments. But perhaps a certain degree of skilled interaction can lead to a richer interactive experience.

B. Gestural Lanuage

Gestures are motions of the limbs or body used as a means of expression. This paper considers only a subset of gestural expressions - gestures produced by the hands and mediated by the use of some control mechanism.

Hand gestures have been used as input for computer systems in many forms. The most common is probably pen-based user interfaces for mobile devices, in which gesture marks or strokes performed with a stylus cause commands to execute.

Harrison et al [4] 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 demonstrate that more natural gestures can be used to control mobile devices and can simplify input in a variety of situations.

In an elegant study, Pirhonen Brewster & Holguin [8] used "metaphorical" gestures on a touchpad to control a mobile music player. For example, the 'next track' gesture was a sweep of a finger across the screen left to right and the 'volume up' gesture was a sweep up the screen, bottom to top. The beauty of this interface is in the very intuitive gestural language mapping between action and invoked function.

Borrowing Chions terms into the tactile component, it seems that tactile control has the potential for both informative and expressive value. To summarize, richness in tactile interactions can arise from skilled hand actions, such as those associated with musical instruments and game devices; and from finding natural, intuitive mappings between gestures and the functions they perform.

V. AUDIO-TACTILITY IN MOBILE DEVICES

A large body of research has been produced by Dr. Stephen Brewster and his colleagues at the University of Glasgow [e.g. 1, 8] on multi-modal interaction, with a special focus on audio-tactile interaction with mobile devices. This research has shown that the use of audio/gesture based interfaces for mobile devices can be preferable over visual/pen-based interfaces [8].

As mobile devices develop and in some cases converge, their functionality becomes increasingly more challenging. I believe recent developments make mobile devices even more suited for both the audio and the tactile components of the interaction. I make this case for both the auditory and tactile components below.

A. Mobile devices are auditory

Many of the functions of mobile devices are auditory in nature. Verbal communication and listening to music are the central ones, as well as the functions of alerting attention through sound (as in rings and other sound signals), gaming and more. The quality of sound in these devices is currently becoming very high. Furthermore, devices such as the Nokia 3300 Music Phone are full fledged sound devices, designed to be listened to for long periods of time. These developments make rich uses of sound in interaction both feasible and potentially pleasurable for the relevant user population.

B. Mobile devices are tactile

Mobile devices are objects that are constantly handled and manipulated. The skill that teenagers have developed in using the number pad for text-messaging – "thumbing" – shows both the potential for skill and the motivational need for interfaces that support fast, natural tactile manipulations of the devices.

Mobile devices are also expanding their input mechanisms. The Sony Ericsson T300 phone now contains a tiny navigation joystick; the Apple iPod boasts a touchpad with superb tactile qualities; and the Sony minidisk comes with a separate, dedicated control device.

I believe that achieving designs that use audio-tactility to its highest potential can create interactions that are more fluent, rich and natural than the ones currently afforded by the developments described above.

VI. The Quikwriting Exploration

Quikwriting [7] is a shorthand designed for use on stylus-based portable devices (currently on Palm Pilot and Pocket PC). Designed by Ken Perlin of NYU, Quikwriting enables inputting text through stylus gestures without picking the stylus up off of the surface. It is based on a visual template 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. Thus each character is a unique gesture of two or three directions, and writing whole words creates combined, longer gestures.

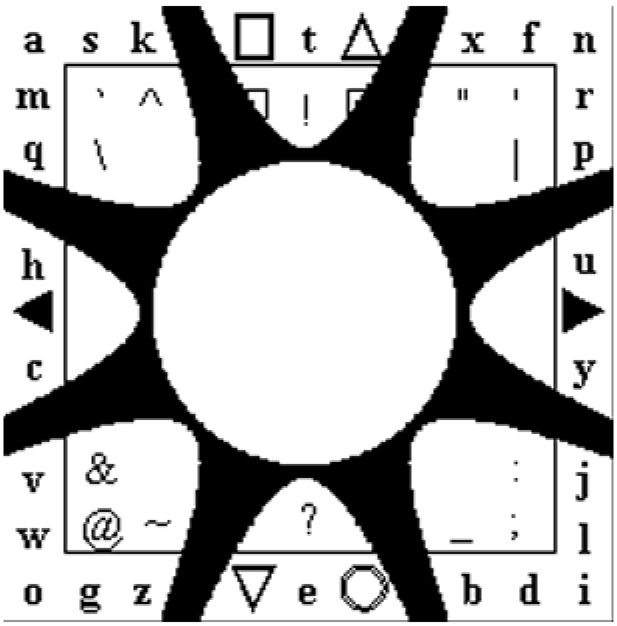

Figure 1 – The Quikwriting template

Quikwriting has a rich gestural language, but a very bleak visual display to support it. It is a complex task with high requirements in terms of feedback and orientation, which needs to be performed in mobile situations. It seems a good test case for adding an audio layer and creating an audio-tactile interaction experience. The crux of an audio-tactile experience I propose for Quikwriting is: moving your hand in the gesture, and getting sound that corresponds to and enriches the gesture.

A. The Tactile component

The standard stylus system enables the Quikwrting gestures, but it does not foster them naturally in its form. For increasing the natural feeling of the task, a prototype physical controller was created for Quikwriting – playfully named the Keybong. Keybong is a small joystick device that fits into the palm of the hand and is manipulated with the thumb.

Figure 2 – The “Keybong”

The Keybong design has two main attributes making it natural for Quikwriting.

The first is that it has a natural start point location at its center – to which it returns with light spring force at the end of a gesture. Therefore, performing the gesture for a character is simplified into a movement into the regions for each character, without having to purposely execute the center-outward and center-inward gesture components.

The second attribute is its physical size and form – it is designed to be used with the thumb in gestures that are a real form of “thumbing” and can be performed, literally, in the pocket.

B. The Auditory component

Creating the auditory component of the audio-tactile interaction with Quikwriting has two intertwined goals. The first is functional – sound provides feedback as to the location within the template, and as to the writing progression. The second is expressive – the choice of sound, coupled with gesture, gives Quikwriting a certain character.

Three exploratory sound sets were designed for Quikwriting, to test three different sound directions.

Sound Set 1: Vocal - In this set, a letter phoneme is sounded 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”).

Sound Set 2: Physical – This set is an attempt to sonify the “physicality” of the Quikwriting template. It contains two elements: a “border” sound that is invoked when the controller moves over a border between the regions and a “movement” sound that corresponds to the speed of movement.

Sound Set 3: Musical - In this set, 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 (when more than one region is entered before returning to the center), while sequences of letters create note sequences – little tunes.

VII. Next Steps

The audio-tactile interface for Quikwriting is currently under development. The next steps involve performing an evaluation of the interface and its different sound sets.

VIII. Conclusions

This paper proposed a framework for thinking about audio-tactile interaction with mobile devices. It raised, and made only an initial attempt to explore, different questions concerning the quality of audio-tactile interaction:

- Can **natural mappings** between sound, gesture and task be found? Do such mappings enrich the experience of the task?

- Can the association between gesture and sound be designed create a **unified** audio-tactile experience? Can the interaction achieve the informative and expressive potential of **consequential** sound?

- Can a digital object be given a specific **character** through the selection of sound and gesture? Can sound and gesture that are "right" for the task at hand render it more pleasurable?

- Can the use of continuous gesture and sound create an interactive experience that is more **expressive** than common digital experiences?

Acknowledgment

I wish to thank Michael Kieslinger, Edouardo Brambilla, Jan-Christoph Zoels, Yon Visell, Alexander Grunsteidl and Graham Pullin for various forms of inspiration and help.

References

[1] Brewster, S.A. "Navigating Telephone-Based Interfaces with Earcons". In Proceedings of BCS HCI'97, Bristol, UK: Springer, 1997, pp. 39-56.
[2] Chion, M. *Audio-Vision*. New York: Columbia University Press, 1994.
[3] Gaver, W. "The Sonic Finder: An Interface That Uses Auditory Icons". In Human Computer Interaction, 1989, (4) pp. 67-94.
[4] Harrison, B. et al "Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces" in Proceedings of the SIGCHI conference on Human factors in computing systems, New York: ACM, 1998, pp. 17-24.
[5] Kramer, G. et al, "Sonification Report: Status of the Field and Research Agenda". Prepared for the NSF. Available: http://www.icad.org/websiteV2.0/References/nsf.html
[6] McCullogh, M. *Abstracting Craft: The Practiced Digital Hand*. Massachusetts: MIT Press, 1998.
[7] Perlin, K. The Quikwriting shorthand system. Available: http://mrl.nyu.edu/projects/quikwriting.
[8] Pirhonen, A., Brewster, S. and Holguin, C. "Gestural and Audio Metaphors as a Means of Control for Mobile Devices" in Proceedings of the SIGCHI conference on Human factors in computing systems, New York: ACM, 2002, pp. 291–298.
[9] Sider, L. Freeman, D and Sider, J. (Eds.) *Soundscape - the School of Sound Lectures 1998-2001*. London: Wallflower, 2003.

Michal Rinott is a graduate student of interaction design at the Interaction Design Institute Ivrea, Ivrea Italy. She received a Masters degree in cognitive psychology from Tel Aviv University, Tel Aviv Israel, in 2000 and a Bachelors degree in psychology and economics from Tel Aviv University in 1996. Her current research focuses on use of sound in interaction design, with a special interest in audio-tactile interaction.

Ms. Rinott has worked as an interaction designer at Pamam - Human Factors Engineering (1997-2002) and at IDEO London (internship, 2003).