Thinking about Sound in Mobile Multimedia

Ilpo Koskinen
School of Design, Industrial Design
University of Art and Design Helsinki
Hämeentie 135 C
00560 Helsinki, Finland
ikoskine(at)uiah.fi

Katja Battarbee
IDEO
100 Forest Ave
Palo Alto, CA 94301
USA

ABSTRACT
In a series of studies, David Frohlich' has shown that sound can do many things in relation to pictures. For instance, if one shoots a blurry image, sound can "save" them by making them still interesting enough to be shared later in photo-talk. There are few existing studies on sound in mobile multimedia, but they suggest that in the future, we must pay attention not just to the fact that camera phones have made cameras ubiquitous, but they also have made microphones ubiquitous. This paper discusses the implications of this fact to mobile multimedia.

INTRODUCTION
As the Call for Papers for this workshop witnesses, practically all research on camera phones has focused on the camera function. Technological research has also largely focused on cameras. Reasons may be related to the powerful visual metaphors at the heart of Western cultures — and Eastern cultures alike.

However, mobile phones today are more than just camera phones. In essence, they are multimedia devices that have a capacity to process sound video and, increasingly, music. In a series of studies on audiophotography, several UK families were given an "audiocamera" to use on summer vacation [1,2,3]. Audio had many uses in the context of photography.

- Ambient sound were street noise, sounds of traffic, music, voices in the background, birds singing, animals, rain, water, and in family scenes, sounds of people walking and laughing. These sounds enrich photographs by adding mood, atmosphere, and humor.
- Sounds revived and save bad photos. In a bad picture of a marching band already gone by, the sound brings the band back to the foreground of experience. Without sound, the photo alone would be without value.
- Sound revives memory better than a mere picture. Sometimes people captured sound and *then* added a photograph to index it. In particular, this was the case with street music. As people learned to "listen" with their eyes, they got used to this practice. Audio and photography are reflexive; there is no clear-cut priority between them.

Frohlich's studies show that there may be something worth thinking in sound and audio, even though these aspects of multimedia phones have received scarce attention [but see 4].

This paper gives examples of some of the most important uses of sound in the Radiolinja MMS Study (hereafter *Radiolinja*), *in which* we followed three user groups in the Finnish mobile phone operator Radiolinja's (now Elisa) technology and service pilot, which took place in July 2002, and lasted about 5 weeks. Each user was given an MMS phone (either a Nokia 7650 with an integrated camera or a SonyEricsson T68i with a plug-in camera). Three mixed-gender groups with 7, 11, and 7 members were studied. Out of the Radiolinja pilot, we selected groups to take into account gender difference, terminal types, and the urban/rural axis. In all, users sent over 4000 messages, and over 2000 were unique (the rest being duplicates in group messages, or recycled messages). These data were produced through the Radiolinja system automatically, and compiled with the help of an assistant. The service was free of charge for the participants.

TYPES OF CONTENTS IN SOUND MESSAGES IN MMS
In the Radiolinja data, there were 190 messages with an audio file. As can be seen from Table 1, most MMS messages with audio are combinations of two or more types of content. The most common types of content were Sharing Mood and Picture Greeting, which were often combined with some other type of content. Messages contained up to four types of content: Greeting + Thank you + Sharing Mood + Good Night.

Informative. People share talk occasionally through sound files, not just through the mobile phone function. For example, in Message 1, Jaana is in a shop browsing cushion covers. She has promised to buy cushion covers for Anne, but cannot find the color they had agreed upon previously.

Table 1. Types of Content in MMS messages with sound (N=190, often many contents in one message)	
1. INFORMATIVE (42/190)	
Questions and answers	20
For Your Information	18
Test Message	4
2. GREETING (total: 68/190)	
Picture Greeting	55
Love Message	3
Congratulations	2
Thank You Note	2
Good Night	6
3. HUMOR (total: 92/190)	
Joke	3
Slur / Insult	7
Picture Puzzle	10
Situation Commentary	30
Media Spoof	3
Absurd Amusement	29
Serial Fire	8
Role play	2
4. EMOTIVE (total: 71/190)	
Sharing Mood	62
Baby and Pet Messages	9

Message 1. Cushion Covers. Jaana to Anne. July 24, 14:10 (18 sec)

Text:

((No text))

In audio, Jaana says:
Here I am, Anne (.) D'you see the m<u>i</u>ddle one among these cushion covers (.) the c<u>o</u>lors are shown pretty badly?, (.) But ehm h (.) this bl<u>ue</u> is quite pr<u>e</u>tty?, /But ehm .h <u>send</u> me a mess<u>age</u> ((intercom in the background)) I'll take it w<u>i</u>th me:::,h (.) so do <u>you</u> w<u>a:</u>nt it or not (.) /Bye h
Quiet ambient sound all along: crowd noise, echo.

Greetings. However, in many cases, people also sing to the phone. In fact, this was a surprisingly common act. Message 2 is almost a cliché-like example of how a multimedia message can be transformed into a singing greeting card.

Message 2. Happy Birthday

Text:

May you have a sunny 30th birthday! Br. Honeytones

Audio: Woman and man singing "Happy birthday to Markku." He is out of tune and rhythm, making her laugh.

Humor and jokes. In many cases, people sent other human sounds with their phones. Most typical of these sounds were cries and laughter. Message 3 is an example of the latter type.

Message 3. Fly Fishing Gear. Jan to Arne. June 26, 12:08 (6 sec)

Text:

I bought fly fishing equipment.

01 Man   HEH HEH HEH HEH HEH
02   <HE HE HE HE HE HEE::e:: hoe:::.>
Ambient sound all along: strong echo, apparently alone in a room.

Emotive. The most common message content type in this study was Sharing Mood. Seven out of eight Sharing Mood messages had picture, sound and text, using the entire spectrum of communication. Often these messages were "celebrations of life" where there was no clear message or greeting, instead the messages were momentary glimpses into what was going on at that particular time. The duration of the pilot coincided with the summer holidays, so free time activities were widely featured. Other emotional messages were focused around babies and pets – a continuous source of interest to the people themselves as well as to other involved parties: other parents, relatives, aunts, uncles, godparents and other pet owners.

SOME FUNCTIONS OF SOUNDS
Obviously, people do several things in these sound files. The most important actions are the following.

Micro-coordination. In a few cases, the sound file was used in micro-coordination [5] instead of a call or a piece of text. In one message, Oona sent a picture of her face and upper torso, and attached a long sound file (30 sec), in which she invited her friend, who was pregnant to join her and another friend for a lazy summer afternoon in her home However, she opened the message by saying: "I got tired on writing, so listen instead..." Another typical case is related to errands, as in Message 1.

Rituals of friendship and other relationships. Good examples of ritualistic speech acts are greetings and other rituals. Although a singing postcard, like Message 2, is not a speech act, it uses a well-known symbolic form to reinforce a friendship.

Fun through parody. Many humorous sound files were set especially early on in the study when the camera phone was a new phenomenon, and people experimented with its

functions. In particular, one man, Mara, became an expert in sending odd cries and sounds to others, including mimicking hens, pigeons, ostriches (!), and pigs, as well as introjections of various sorts. A slightly more sophisticated usage built in existing cultural forms. For example, Anna sent a series of pictures of her sailing trip in the South Coast of Finland to her friends. In one example, there was a fuzzy picture from a marina. In sound file, Anna whispered like a stalker how the picture is about how two well-known actors have just been exchanging bottles of cognac.

Maintaining Co-presence. In a series of papers, Okabe and Ito have elaborated the notion of "co-presence" [6]. People often send all kinds of materials to the recipients just to share a sense of their current whereabouts. These messages are often about insignificant things. Such practice takes place between family and couples, but also among close friends especially if they have been drinking. Good examples are sound samples from bars, rock joints and jazz concerts.

TWO POSITIONS OF SOUND

Koskinen [5] has divided these into foreground and ambient sounds. Talk and greetings are typical examples of foreground sounds: they are the point of the sound file. Echo, streetcar sounds, and quiet, barely audible crowd talking in the background are examples of ambient (or background) sound. Typically, they add context to the foreground sound, but are not framed as key parts of the communicative act.

In these two positions, sound has several uses in multimedia communication. For instance, as Koskinen [7] has argued, one can follow how interaction evolves from a sound file by listening to it. Through this interaction, the listener may also get an idea of the social organization of the setting without any explanation in text. For example, in the following example, the foreground sound is clearly the sounds of the baby that are designed to delight Arne (the message was sent to nine other people too). However, ambient elements clearly situate the baby into the middle of a small group of adults talking about something, and having fun with the baby. As the sound file goes on, this organization becomes "hearable": Arne does not have access to it at the outset.

Message 4. Baby Talk. Susan to Arne. Sent to 10 persons. July 24, 12: 37 (26 sec)

Text:

Greetings from East Centre, br. Zoey

01 B:	((Quiet noise, attempts to talk))
02	(2.0)
03 W	((Laughs))
04 W	((Talks to other adults in the background,
05	words inaudible))
06	(1.0)
07 W	Oh look, here it (comes) ((talks to adults))
08	(1.0)
09 B	((Talks louder))
10 W	((Talks, followed by laughter, inaudible))
11	(1.0)
12 B	((Cries loud, almost starts to scream))
13 W	And then we take the bott— ((sound cut off))

Ambient sound all along: Quiet talk by many people, a café—like sound that disappears when the baby starts to talk and cry, or the mother talks loud.

Message 5 offers a slightly more complex example. In this example, Tiina is in the bar of a cruise ship ("Silja" refers to Silja Line, a cruise company operating mainly in Finland, Sweden, and Estonia). In addition to greetings, she sends picture, and builds a riddle on it. In the sound file, she offers a sample of music she is hearing at the moment, and identifies the artist by name.

Message 5. What's in the Mug? Tiina to Thomas

Text:
Greetings from the Silja pub. A riddle: what's in the travelers' mugs? Have a nice Sunday!

Audio:
Foreground: Female voice saying: "And the artist of the night, Jari Mäki. A sample comes here."
Ambient: restaurant noise; background music: Who'll Stop the Rain.

Message 6. (Response) Thomas to Tiina

Text:
Our guess is red wine, but we could not make sense of the sound.

Audio:
A man sings an old commercial jingle (transl. IK):
Always drink Juice Cat, it always tastes great, Juice Cat

As the reply shows, the riddle was easy and playful enough to be answered with a jocular reply. However, although Thomas also comments about the bad quality of the sound sample, he also replies with another well-known song, which gives the reply a jocular, summer-drink related tone. Apparently, he was able to pick up a part of the message, and build on that even though the sound quality was bad in Thomas's opinion

AN INITIAL CLASSIFICATION OF SOUND IN MOBILE MULTIMEDIA

Some of these audio elements are planned by the sender, while some are accidental. If we cross-tabulate the distinction between foreground and ambient sound by whether an element was intentional of unintentional, we get Table 2, which breaks sounds in the Radiolinja data into four main types.

Table 2. Four types of sounds in multimedia messages: an initial typology

	Intentional	Non-intentional
Foreground	Singing in Messages 2 and 6; laughter in 3; talk in 1 and 5; and baby talk in 4.	The woman's laughter in Message 2; talk in 4 (lines 7, 13).
Ambient	Music sample in Message 5.	Ambient talk in Message 4 (lines 3-4, 10); intercom in 1; echo in 3; talking crowd in 5.

As Table 2 suggests, there is practically always an intentional element in messages. In most cases, this is in the foreground, but it can be in an ambient position too. In contrast, it is perfectly possible that a sound message does not contain non-intentional elements. Typically, they are just remnants of what was going on then the foreground elements were recorded.

Given these variations in aims, types and positions, it is not surprising that sound does lots of things in multimedia messages. It enlivens and enriches pictures in many ways. It makes it possible for people to coordinate their comings and goings. It conveys a sense of place. As examples in this paper show, people do not have to "design" sounds in the way they did in Frohlich' studies, where they deliberately told stories about images, and added pieces of music to their shots. Even "accidental" sounds provide samples of life and can change the meaning of the image.

PROBLEMS TO BE THOUGHT

I take it that things above have shown that sound is a rich resource for action in multimedia. How to use this resource and design for it now that MP3-capable phones are quickly spreading in the market, making it possible to not just receive and (possibly) send music files and samples of music, but also recording sound and sharing either to support photographs, or to provide the background upon which people design their photos?

Sound quality: the microphone and the sound-packing software. The main limit is perhaps still the microphone and the software used in packing sound into files that are small enough for sharing over still fairly slow networks.

User interfaces. The first generation of multimedia phones offered a limited set of tools for managing sound. In essence, it was only possible to record sound, and attach it into messages. Inevitably, some kind of timeline is needed for creating more complicated sound scapes. How to make them simple enough so that people learn to use them in the limited context offered by the mobile phone? Are there better metaphors to begin with?

Supporting sampling/collages. It ought to be possible to capture several audio files and blend them into one message. Such feature would make it possible to send ambient sound, add sound effects, and blend these with a separate foreground talk track. Such features exist for simple visual effects. How about sound?

Sound libraries. There ought to be an easy way to build and use personal sound libraries for mobile phones and share their contents with other people. Koskinen has heard about a musician who uses his phone to collect rhythms: when he invents a new one, he drums it to his phone immediately, building an ever-growing catalogue with the phone in the same manner as artists carry their notebooks.

Integration to the Web. How could the Internet be used to support creating and supporting sound — without lapsing into crass commercialism? A good socially aware example of how sound can be used to enhance experience is the moblog in Tokyo's Shibamata district [8]. With camera phones, people could download sound files from the server created by researchers at Keio University. These sound files typically had not just ambient sounds from the streets, but also stories about particular scenes depicted in postcards created by researchers and students of Keio out of this traditional, picturesque neighborhood. Systems aimed at easing sharing mobile multimedia [9] ought to support audio, not just visual content and text.

Music. One main type of sound is music. We need a separate study of it. Compared to *Audiophotography* [1], little music was sent in the Radiolinja study, probably largely because sound quality was not good enough.

DISCUSSION

This paper has been a plea for analyzing sound and integrating this highly expressive and rich function into the research and development agenda of pervasive and ubiquitous computing. Mobile video is in the marketplace, but we do believe that the Radiolinja data shows that there is a place for supporting and creating sound files that are not synchronized with images.

REFERENCES

1. Frohlich, D. (2004). *Audiophotography. Bringing Photos to Life with Sounds*. London: Kluwer.

2. Frohlich, D., Tallyn, E. Audiophotograpy: Practice and Prospects. *Proceedings of Computer-Human Interaction CHI'99*, May 15–20, 1999. New York: ACM Library.

3. Frohlich, D., Kuchinsky, A., Pering, C., Don, A., Ariss, S. Requirements for Photoware. *Proceedings of*

Computer-Supported Collaborative Work CSCW'02, November 16–20, 2002, New Orleans, Louisiana, USA.

4. Koskinen, I. How People Use Sound in Mobile Multimedia. *Proceedings of DPPI 2005: Designing Pleasurable Products and Interfaces*, TU/Eindhoven, Oct. 24-28, Eindhoven, the Netherlands, 2005.

5. Ling, R. (2004). *The Mobile Connection.* San Francisco: Morgan Kaufmann.

6. Okabe, D., Ito, M. Everyday Contexts of Camera Phone Use: Steps Toward Technosocial Ethnographic Frameworks. In Höflich, J., Hartmann, M. (eds). *Mobile Communication in Everyday Life: An Ethnographic View.* Berlin: Frank & Timme, 2006.

7. Koskinen, I. (2007). *Postcards from the Mobile Phone Generation*. Transaction Publishers, New Brunswick.

8. Kato, F. Seeing the "Seeing" of Others. Conducting a Field Study with Mobile Phones/Mobile Cameras. *Proceedings of Seeing, Understanding. Learning in the Mobile Age. Communications in the 21st Century: The Mobile Information Society*, Budapest, Hungary, April 28–30, 2005.

9. Van House, N., Davis, M., Takhteyev, Y., Good, N., Wilhelm, A., Finn, M.: From "What?" to "Why?": The Social Uses of Personal Photos. Proceedings of CSCW'04, November 6–10, 2004, Chicago, Illinois, USA, 2004.