Sound in Mobile Multimedia: A Mobile Design Challenge

Ilpo Koskinen[1]

[1] School of Design, Industrial Design, University of Art and Design Helsinki, Hämeentie 135 C, 00560 Helsinki, Finland
ikoskine@uiah.fi // http://www.uiah.fi/~ikoskine/

Abstract. Sound is an important element of everyday life, but most research on the use of sound in mobile media is based on analyses of the Sony Walkman. This paper analyzes mobile multimedia devices that, unlike the Walkman, connect people to the surrounding soundscapes, making it not just observable, but also recordable and reportable for various purposes. I examine both foregrounded as well as ambient sound, the former defined as the most prominent sound element in a message and the latter as sound that is in the background of the message. Data is from a study of mobile multimedia conducted in Helsinki, Finland, in summer 2002. The analysis focuses on three things: the uses of foregrounded sound, how ambient sound makes place and social interaction available, and how recipients show their understanding of these sound types.

1 Capturing Sound with Mobile Multimedia

The most thoroughly researched area of portable sounds in the social sciences is the Sony Walkman [1]. In contrast to early public discourse, which tended to condemn the individualizing effect of the Walkman [2], early researchers argued that it reorganized the users' perception of the urban landscape. The Walkman distances the user's experience from the city around him; it was a liberating technology [3, 4]. It also acted as an interaction shield [5: 40]. This view was criticized for ignoring the uneven distribution of the Walkman and the fact that this is what the Walkman was designed for in the first place [2: 106-109]. Later, empirical research has focused on how the urban soundscape re-enters the Walkman experience by focusing, for example, on how people tune down the system to hear the soundscape around them when navigating the city [6]. In Bull's [1] evocative phrase, people are "sounding out the city" with personal stereos.

What about the opposite process, re-sensitizing ears to sounds? How does sound enter the digital experience through portable smart objects, capable of making re-

cordings and relating sounds to photographs? In some quickly spreading mobile mul-
timedia, the technical and use environment is different from the Walkman and digital
cameras. Mobile phones and PDAs (personal digital assistants) equipped with a cam-
era and either MPEG, MP3, or MMS technology offer an environment with many
types of technologies in one device: there is a camera as well as audio and sometimes
video capture functions. Moreover, people can easily annotate images with text,
bringing a discursive element into the process. There have been a few studies of how
mobile multimedia is used in ordinary life [7, 8, 9, 10, 11, 12, 13, a review is in 14].
Also, there are recent technological studies in which users have been given a promi-
nent status [15, 16, 17]. However, these studies have focused on photographs without
touching sound. Still, with new mobile multimedia devices, not just the visual, but also
the audible world something to be captured and shared. What kinds of problems audio
poses to design remains open.

2 Research on Sound in Mobile Multimedia

The best evidence of how people link photographs with audio comes from a series
of sociologically-informed studies in the context of technology development. In a
study on audiophotography, several UK families were given an "audiocamera" to use
on summer vacation; all had a multimedia PC [18, 19]. These audiocameras were
"experience prototypes" [20], consisting of an analog camera (Minolta AF 101R) and
a dictaphone (Lanier P-155) unit glued together. With this combination, the users
could record photographs and sounds in any combination, but not simultaneously, at
the capture time. However, audiocapture was not automatic, as in video camera; the
system forced users to think how to enhance photographs with sounds for research
purposes. Audio had many uses in the context of photography [18].

- Ambient (i.e. natural) sounds occurring just before, during, or after the photo-
  graph was taken added important context. 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 enhances memories: it revives memory better than a mere picture [21].
  Perhaps the most important message of this study was that sounds and photo-
  graphs work in two directions. 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 [19: 44-46] classifies interactions with audiophotographs in a model in
which the photograph, the subject, the photographer and the audience are players.

Various interactions between these parties can be classified into six types. Recognition involves the photographer reviewing his photos, while recollection and self-reflection is an activity of the subjects in the photos. The audience, which has no memory of the original situation, can interpret photos. When the photographer shows photos to an audience, Frohlich talks about storytelling in the first person, while storytelling in third person consists of the *subject* showing photos to an audience. In reminiscing, the photographer and the subject share content. Thus, the interpretation of photos by an audience is a special case in the home mode of photography [22]. In Frohlich distinguishes four types of sound: ambient sound, music, voiceovers, and storytelling attached to photographs. For example, music is usually entertaining and acts as a memory enhancer [19: 105-107], while ambient sounds typically assist personal recollection, audience understanding, and reliving the original situation by providing useful contextual information to assist interpretation [19: 76-77].

If we extrapolate from this body of research, we can conjecture how people could use sound in mobile multimedia. Often, capturing ambient sound is an involuntary activity; the sound is there as if by accident. Frohlich and Tallyn [18] also show that users can annotate images with audio and text alike. They can also annotate sound with photographs and text. Finally, they may create constellations in which various multimedia elements can support each other, but also conflict with each other.

Useful as this conjecture is, it ought to be taken cautiously for at least three reasons. *First*, the technological and use contexts of audiophotograpy and mobile multimedia are different. In particular, mobile multimedia offers a small screen and a relatively poor quality audio track. However, this shortcoming is alleviated with the easy texting capacity offered by mobile phones and PDAs. *Secondly*, mobile multimedia is essentially a communicative environment, built to support communication regardless of place. *Finally*, mobile multimedia brings these technological capabilities to interaction: people can communicate their experience with mobile multimedia right away. In turn, distant others can ask if they do not understand something [cf. 14].

3 Foreground and Ambient Sound in Mobile Multimedia

Sound, like text, can be used to capture ideas, intentions, rational trains of thought and other forms of everyday impressions. In sending greetings, news, questions and sometimes humor, sound often works better than text in SMS or "voiceless" MMS. There are two types of audible sound in messages. First, there are "*foreground*" sounds: these are the most prominent sound elements in the message. Typically, the foreground is composed of one or two spoken utterances. People may also add color and meaning to messages by, for instance, making ironic or humorous statements next to the message [23]. Second, the microphone also captures "*ambient*" sounds: in the background of the message, there are sounds from the cityscape, people, animals, objects, music, street noise, sounds of traffic, people talking, and steps on sidewalks. These sounds are typically – but not necessarily – attached to messages unintentionally.

However, people interpret both of these sound types in terms of meanings, not just in terms of sound waves: people experience the sounds *of* things rather than mere sounds [24: 59]. What is involved in this work is what Garfinkel once called "the documentary method of interpretation," treating (here) sounds as "the document of," a presupposed underlying pattern, which is derived from individual documentary evidence. This individual documentary evidence, in turn, is interpreted on the basis of what is known about the underlying pattern. Each is used to elaborate the other [24: 78].

For example, talk can point to a café or restaurant, but just as well to the street. In a spoken audio clip, the recipient can study the speaker's voice to seek cues about what is going on. A weary or intoxicated voice provides a context that helps in understanding the message. Other elements in the message are then scrutinized to seek further cues about the pattern. Once the underlying pattern is identified, the sounds, still capable of many interpretations, are interpreted and acted on in terms of what people know about this place and action in it. Senders can use their knowledge of the recipients' understanding methodically in composing their messages. Each multimedia element in the message get a meaning from other elements in the message in a "reflexive" fashion [24: 1-11].

The claim of this paper is that people use and make sense of sound in mobile multimedia at several levels simultaneously. They take into account not just the prominent foreground, but also the ambient background in both devising and making sense of messages. Sound is a rich method of social action; it makes several things available simultaneously, though much of this typically escapes the users' and the analysts' attention.

4 Data and Methods

In the Radiolinja MMS Study (hereafter *Radiolinja),* 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 plugin 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. Exact numbers are confidential, but the following figures point to the scale of messaging in the pilot. 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 several assistants. The service was free of charge for the participants.

For this paper, we have treated these data in the following fashion. From the vast mass of *Radiolinja* messages, we have chosen a subsample consisting of 543 messages, sent by the 12th group (with 7 members) during the third and fourth weeks and by the 8th group (11 members) during the 4th week of the study. Participants knew that they were being studied, and were informed about the ethical procedures we used. In

particular, we told them how our data was produced, promised not to publish pictures without their consent, and promised to change details of images so that it would not be possible to identify them from our publications. In addition, we have followed standard academic and legal practice and have changed all names and details that could identify people or places.

Transcription of the audio data follows conventions from conversation analysis, though with a few differences [26]. (Table 1).

Table 1. Transcript Symbols

Symbol	Explanation
(.)	Micropause, or interval of 0.1 second in talk
(0.4)	An interval of 0.4 seconds
'n [she sa]id 　　[But th-]	Overlap begins and ends
=[[I'm saying 　[[But no::	Utterances start simultaneously
Wha:t	A colon indicates an extension of the sound it follows. Each colon is about 0.1 seconds
.	A stopping fall in tone
,	A slight fall in tone
?	A rising inflection
?,,	A slight rising intonation
/ \	Rise and fall in intonation
Wha:t	Underlining indicates emphasis
WHAT	Loudly
what	Quietly, or in whisper
hhh .hhh .nhh	Outbreath, inbreath, and inbreath through nose respectively. Each "h" is about 0.1 seconds
(what)( ) say	Single parentheses indicate transcriber's doubt or best guess
((door slams))	Double parentheses indicate various features of the setting or transcriber's comments
.mt .pt	Click or a smack of tongue
.nff	Snuffling
#that's true#	Creaky voice
@what@	Markedly different tone than elsewhere
$what's that$	Laughingly
W(h)at	Within words, (h) is a laughter token
he HEH HEH hah	Laughter tokens
wh-	Cutoff of a word
And th( )<	The speaker halts some unit in progress
>she said<	Quickly

In all, there were 72 audio messages, i.e. files with an audio component using the AMR file format. However, most of these sound files were sent as copies: only 14 are

original; the rest were like the baby example analyzed later in this paper, which was sent to 10 persons. The length of the audio clip ranged from 4 to 24 seconds. With one exception (a graphic), there was always a photograph in the message. Three messages were sent without text. There were 28 distinguishable ambient audio elements in these data. In ten cases, these elements were ambient noise: sounds of shopping malls, the street and bars, but also wind, echo and a radio in the background. In 18 cases, human sounds were in an ambient role. In 12 cases, there was indistinguishable talk in the background, in five cases laughter, and in one case a crying baby.

We selected a subsample of 543 messages. In analyzing data in detail, the first phase was an unmotivated search for similarities and recurrent issues, and the second was creating a series of hypotheses from data in group 8. This interpretation was treated as a working model, which was "tested" with all data. Thus, the analysis followed analytic induction: if a previous model works in new data, it provides a sufficient explanation. If not, the model is changed until the model accounts for all cases. This procedure creates a model that describes what is going in *these* data. It does not generalize to other data; local circumstances and research design therefore restrict the applicability of these results to other data.

5 Talk in the Message Foreground

Multimedia messages in these data typically consist of a photograph, text, and sometimes sound. In some messages, sound is a dominating element in the message. Typically, this element is talk. What does talk do in mobile multimedia messaging? What is its main value in relation to phone calls?

By far, the most important thing done with foreground sound is greetings of various sorts. These ranged from birthday greetings to "have a nice day" messages. The role of photographs and text varied in these messages. Thus, in one example, Niko send a "Happy shopping" message to Anne, who had sent a picture of her baby from a local shopping mall five minutes before. In this message, which included no text and an image that was apparently unrelated to its content, Niko sent greetings to Anne from his home suburb. In the message, he describes his home, currently under rennovation, as a "work camp," comparing his environment to Anne's more pleasant one with this joke.

Another example of a greeting is the following message (Fig. 1), a birthday greeting for Markku who is turning 30. This message is essentially a singing postcard. In it, Anne and her husband sing the clichéd "Happy Birthday" (in Finnish) to Markku, also adding a picture of flowers and a textual greeting to the message. We also hear how two people react to their performance. He is initially out of tune and rhythm, which she acknowledges with laughter tokens while singing. The pair accounts for their flawed performance with the signature "honeytones," which makes the message jocular: this is a simple tune, but so clearly out of rhythm and melody that they account for these flaws with this self-ironic item.

The second prominent use of foreground sound is for sound samples [18]. Typically, samples were sent of babies (see Fig. 3 below) and friends. The third prominent

use was imitating human or animal voices. For example, in one message, there was a woman sleeping in a car. The audio mimicked loud snoring. In another case, the object of imitation was an ostrich; the sender had visited an ostrich farm. The fourth usage was more rational. As Fig. 2 below shows, voice could be used instead of a call to assist decision-making. Jaana sent a message to Anne about cushion covers she had seen in a shop. However, in these data, this is the only case in which audio files were used to coordinate actions in commercial or institutional contexts. The final, and quite prominent, way to use sound in the message foreground was to use it as an "emotion enhancer": sound described the sender's feelings. Interestingly, this usage has a syntax-like format. Most typically, there was a picture of the sender's face, with some kind of yell or other emotionally loaded sound added. For example, Arne once sent a message evaluating his recent cruise by noting in the text that he is disappointed and ready for further adventure. The sound was a loud "Bla::::::,h::::," which leaves little doubt about how his cruise had gone. Interestingly, explanatory uses of voice – annotating images with words – is non-existent in these data [18, 21].

Fig. 1. Sent to Markku. 091_0725_0812_08_272Anne_to_Markku.psd (14 sec)	
	Text: Have a sunny 30th birthday! Br. Honeytones
01 Woman ((singing)) Happy bir[thday to you, happy 02 Man [*day to you* ((joins 03 Woman =[Birt(h)(h)hday to you, happy birthday to Markku,h 04 Man [((Singing, first out of tune and rhythm - - - - - - - - - -)) 05 Woman =[Happy /birthday] to you ((at the end, out of tune)) 06 Man [((- - - - - - - -))] Ambient sound: no recognizable ambient sound.	

6 Hearing Place and Interaction in Ambient Sound

The soundscape in mobile multimedia is considerably more complex than the foreground. By listening to the foreground only, recipients not only hear the message, but also get an idea of what the sender has intended to say, his mood, how he has assessed his experiences, his sense of humor, and also his company. However, when recipients

turn their attention to a background pattern rather than the evidence in the foreground, they get access to a host of other, more contextual aspects of what is going on.

In the next example (Fig. 2), it is easy enough to understand what the message is intended to do. Jaana has promised to check a cushion cover in a shop, and has promised to buy it, given certain specifications by Anne. In the shop, Jaana learns that the color Anne wants is unavailable. She captures an image and sends it, with an audio clip asking whether she should buy it anyway. Interestingly, Jaana also corrects an element in the image with audio [18] when she notes that the colors are reproduced badly in the photograph. She asks for a quick reply before closing the message with a goodbye.

Fig. 2. Sent to Arne. 029_0724_1410_08_897Jaana_to_Anne.psd (18 sec)	
	Text: ((No text))

01	Woman	.pth Here I am, Anne (.) D'you see the middle one among
02		these cushion covers (.) the colors are shown pretty
03		badly?,
04		(.)
05		But ehm h (.) this blue is quite pretty?, /But ehm .h [send me a
07	Amb.	[((intercom))
08	Woman	message, I'll take it with me:::,h (.) so do you wa:nt
09		it or not (.) /Bye h

Quiet ambient sound all along: crowd noise, echo, intercom.

Notice that there is no mention of *where* Jaana is shopping. It could be any commercial venue – a shopping mall, department store, or a shop. As the "here I am" at the opening of the message implies, Jaana and Anne have discussed Jaana's going to the mall previously, so that there is no need to identify the place anymore. Still, ambient sound manages to situate the message in a shopping mall, for two reasons. First, the background noise tells from the very beginning that the place is large. There is a strong echo in the sound, and the talking crowd is sizeable, thus ruling out a local shop or a boutique. When the intercom starts after 12 seconds, the remaining alternatives are either a shopping mall or a large department store.

This case shows how audio and other multimedia elements work together to situate action in a certain place. People may infer what others are doing *based on* their knowledge of the place, and they may infer place from activity. When people hear that someone is in a shopping mall or a bar, they can figure out with good confidence what

he is doing there, and also his mindset (see [27] for how places and geography functions in talk; a more formal treatment, see [28: 96-106]). Even a minimalist ambient soundscape may do important work in mobile multimedia.

Ambient sound may also make social action available for the hearer. In some cases, hearing the ambient sound tells us about social action. In Message 3, we saw how two people took each other's orientations into account while singing a birthday song. In the following message (Fig. 3), there is a picture of a baby against blue water, suggesting that the image was taken in a swimming pool. This is confirmed by the text, which situates the message in the "East Centre," a large shopping mall roughly 10 km east of downtown Helsinki. There is a swimming pool near the mall which arranges baby swimming. At the top, the whole message is offered as a "Greeting" sent to ten persons, and the sound of a laughing baby. Sharing a delightful experience also justifies the message.

Fig. 3. Sent to 10 persons. 006_0724_1237_02_272Susan_to_Arne.psd (26 sec)	
	Text: Greetings from East Centre, br. Zoey

```
01   Baby    ((Quiet noise, attempts to talk))
02           (2.0)
03   Woman   ((Laughs))
04   Woman   ((Talks to other adults in the background, words
05           inaudible))
06           (1.0)
07   Woman   Oh look, here it (comes)(   ) ((talks to other adults))
08           (1.0)
09   Baby    ((Talks louder))
10   Woman   ((Talks, followed by laughter, inaudible))
11           (1.0)
12   Baby    ((Cries loud, almost starts to scream))
13   Woman   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.

But this is the foreground sound only. An analysis of how the ambient sound develops and is directed gives us a detailed idea of what is going on. We first hear a baby attempting to talk, and see a young man in the background (who he is remains unclear). Next a woman laughs near the microphone. It is at this moment that we first

hear a key element of the social organization: it is the baby's mom [29, 30]. After approximately 7 sec., the mother talks again near the microphone. After 11 sec., she turns her head away from the microphone and talks to other adults, who respond to her. By now, all participants are known. However, a second phase starts when Zoey begins her baby talk again (13 sec.). In response, the mother talks to other adults, and laughs briefly with them. After two seconds, Zoey starts to cry, this time loud. The mother takes this as a sign of hunger, and starts to feed her (24 sec). Even though this is not mentioned anywhere in the message, we hear how the mother keeps Zoey in her arms and feeds her.

Figure 3 shows how people can use common–sense knowledge of social structures to make sense of what they hear [25: 76-103]. This scene tells not only about how the mother and baby interact, but also about their company and its mood: they are with other adults, who share the mother's joy rather than remain merely bystanders. We can also follow the mother's orientation from the sound even though we have no idea of her interlocutors or how they are placed around her. Recipients of the message not only hear several voices, but they also hear an on-going social organization: how the mother juggles in an attempt to allocate her involvement between the baby and her social surroundings.

7 Discussion: Towards New Design Challenges

Unlike portable music devices such as the Walkman that people use to restructure their sonic experience while on the move [1, 2, 3, 6], mobile multimedia gives people new means to observe and report their activities and experiences with the environment [for example, 7, 9, 11, 12, 13, 15, 17]. This paper has focused on one aspect of the "hiptop" multimedia environment only: the uses of audio capacity. The focus has been first on how people use sound in the foreground of the message and second on how they use ambient sound. The analysis of this paper confirms the main result of *Audio-photography* [18, 19], that is, audio "augments" images in many ways: it adds life to images, allowing people to communicate more fully non-tangible qualities of action and context. It would be hard to communicate all these meanings with text alone, as in SMS, or even with photographs augmented with text.

Much as the camera, the audio feature in mobile phones roots our experience tighter to our physical and social environment [see 13]. With sound, we are able to capture a far livelier impression of what is going on around us and share it with recipients. They, in turn, get a rich access to the sender's situation, and are able to relive parts of it, rather than just construct it from a picture and a piece of text [7]. They hear how things unfold in time, and are able to connect what they see in picture and read in text to a more general, meaningful co-experience, as the sharing of experience has recently been called [23].

Given that sound has an impressive expressive and interpretive potential, why are there so few audio messages? We may present a few cultural conjectures. First, voice mail culture has not been a standard part of life in Helsinki. Second, sound is not a part of the considerably more lively text messaging culture [31, 32]. Third, the sound

world in mobile multimedia is largely self-evident and taken for granted. Here the difference to Frohlich [19] is illustrative: for example, in his data, ambient sounds were typically from trips to unfamiliar places, or places dense with memories. The Radiolinja data is from places with a familiar soundscape that was too unremarkable to guarantee interest. Finally, it may be that people associate sound with video rather than still pictures.

Aside these cultural issues, the reason may be closer to design: using all the media available may be just a bother. What can designers do to help people use sound in mobile devices? With further research pending, at least these things need attention:

- *Improve capture*. The technical quality of microphone should be improved. People in Radiolinja often did not hear what was going on in audio files without going into a quiet place. Technical level enough to capture talk for a call is not always enough for meaningful and fun capture of ambient sounds.

- *Improve the ability to construct messages*. In phones used in Radiolinja, it was possible to construct multimedia messages with three separate files: in addition to text, a picture file (JPG), and an audio file (AMR). 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?

- *Improve user interfaces*. Implementing a "timeline" type user interface feature to the small screen of a mobile phone requires research and innovation. Should one want to send a message with several sound files and connect them into a soundscape, one needs either a timeline or another similar UI component. However, for first-timers, timelines are difficult to learn on the large screen in PC-based multimedia programs. How about in the small screen, with no mouse and a limited keyboard? Are there alternative, easier metaphors that could replace the timeline?

- *Improve sharing audio*. Systems aimed at easing sharing mobile multimedia [17] ought to support audio, not just visual content and text. A related, but separate question relates to music. Compared to *Audiophotography* [19], little music was sent in Radiolinja (only a couple of sound samples of live bands were shared), probably largely because sound quality was not good enough. There ought to be an easy way to build and use personal sound libraries for mobile phones and share their contents with other people. How could the Internet be used in this function – without lapsing into crass commercialism?

- *Improve methods of design research*. For design researchers, sound poses a need to develop new methods. Interviewing aside, current design research methods – such as the probes and the procedures of contextual inquiry – are largely visual. How to take sound into account in designing a distributed environment typical to mobile technology?

- *Improve designers' ability to think about sound as experience*. Another issue for design research is conceptual. Designers typically think visually, but to analyze sound, we need new, non-visual ways of thinking. Designers typically use concepts from physics to understand sound, but they do not help in understanding how sound relates to experience. Building on early ethnomethodology [25], with

affinities to phenomenology [24], this paper has suggested one framework to act as a prism for observing how people use sound. Are there alternatives designers are better familiar with?

Does sound enhance the quality of social links between people? On purely practical grounds, sound files add little to mobile messaging. With it, people do things they have done with text messages for years: greetings, questions, and even gossip. However, although people may develop distinct styles of writing, sound is a more easily personal instrument than text for most people. For example, in Figure 2, Jaana not only makes a question, but also gives a sound sample to Anne. Furthermore, this sound sample contains ambient elements that make it easy for Anne to situate Jaana into certain surrounding. Over a series of messages like these, it is likely that people develop a richer understanding of their fellows and where they move. In this sense, sound enhances sociability, much like photographs in "mute" mobile multimedia [see 7: 84-85; 12]. However, a proper answer to this question requires another study that would relate mobile multimedia to social organization rather than start from messages as such.

In all, the observations in this paper represent a call for arms for designers to start to pay attention to the possibilities of sound in mobile computing environments. Sound is a rich source of fun and enjoyment in mobile multimedia, even when used by ordinary people with no training in creating, analyzing, interpreting, or planning audio clips. Consequently, sound represents a remarkable design potential for mobile devices. However, it also represents significant challenged for engineering, user interface design, and for design research. The aim of this paper's excavation of minuscule practices that are taken for granted and go largely unnoticed in everyday life has been to open up our eyes to these design challenges.

Acknowledgements

References

1. Bull, M.: Sounding Out the City. Berg, Oxford (2000)
2. Du Gay, P. et al.: Doing Cultural Studies: The Story of the Sony Walkman. Sage, London (1997)
3. Hosokawa, S.: The Walkman Effect. Popul. Music 4 (1984) 165-180.
4. Chambers, I.: Border Dialogues. Journeys in Postmodernity. (Partly available in pp. 141-143 in Du Gay, P. et al.: Doing Cultural Studies: The Story of the Sony Walkman. Sage, London (1997). Original (1990)

5. Goffman, E.: Behavior in Public Places. The Free Press, New York (1963)
6. Thibaud, J.-P.: The Sonic Composition of the City. In Bull, Michael and Les Back (eds.): The Auditory Culture Reader. Berg, Oxford (2003)
7. Koskinen, I., Kurvinen, E., Lehtonen, T.K.: Mobile Image. IT Press, Helsinki (2002)
8. Kurvinen, E.: Emotions in Action: a Case in Mobile Visual Communication. Proceedings of the 3rd International Design and Emotion Conference D+E'02 (2002).
9. Kurvinen, E.: Only When Miss Universe Snatches Me. Teasing in MMS Messaging. Proceeding of Designing Pleasurable Products and Interfaces DPPI'03, June 23-26, 2003, Pittsburgh, Pennsylvania, USA (2003)
10. Okabe, D., Ito, M.: Camera Phones Changing the Definition of Picture-Worthy. Japan Media Review. Annenberg School for Communication, Univ. Sth. Calif. (2003). www.ojr.org/japan/wireless/1062208524.php. Accessed April 15, 2004.
11. Ling, R., Julsrud, T.: The Development of Grounded Genres in Multimedia Messaging Systems (MMS) among Mobile Professionals. In Nyíri, K. (ed.): A Sense of Place. Passagen-Verlag, Vienna (2005)
12. Scifo, B.: The Domestication of the Camera Phone and MMS Communications. The Experience of Young Italians. In Nyíri, K. (ed.): A Sense of Place. Passagen-Verlag, Vienna (2005)
13. Koskinen, I.: Seeing with Mobile Images: Notes on Collaborative Seeing in MMS. In Nyíri, K. (ed.): A Sense of Place. Passagen-Verlag, Vienna (2005)
14. Koskinen, I., Kurvinen, E.: Mobile Multimedia and Users: The Domestication of Mobile Multimedia. *Telektronikk,* 2005, (forthcoming).
15. Kindberg, T, Spasojevic, M., Fleck, R., Sellen, A.: How and Why People Use Camera Phones. Consumer Applications and Systems Laboratory. H&P Laboratories Bristol, HPL-2004-216, Bristol (2004). Accessed November 26, 2004.
16. Berg, S., Taylor, A. S., Harper, R.: Mobile Phones for the Next Generation: Device Designs for Teenagers. Proceedings of Computer-Human Interaction CHI 2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA (2003)
17. 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)
18. Frohlich, D., Tallyn, E.: Audiophotography: Practice and Prospects. Proceedings of CHI'99, May 15-20, ACM (1999)
19. Frohlich, D.: Audiophotography. Bringing Photos to Life with Sounds. Kluwer, London (2004)
20. Buchenau, M., Fulton Suri, J.: Experience Prototyping. Proceedings of DIS 2000. New York: ACM, (2000) 424-433
21. Frohlich, D. et al.: Requirements for Photoware. Proceedings of CSCW'02, Nov. 16-20, New Orleans (2002)
22. Chalfen, R.: Snapshot Versions of Life. Bowling Green State University Press, Bowling Green, Ohio (1987)
23. Battarbee, K., Koskinen, I.: Co-Experience: User Experience as Interaction. CoDesign Journal 1 (2004) 5-18
24. Ihde, D.: Listening and Voice. A Phenomenology of Sound. Ohio University Press, Athens, Ohio (1976)
25. Garfinkel, H.: Studies in Ethnomethodology. Prentice-Hall, Englewood Cliffs, NJ (1967)
26. Jefferson, G.: Transcript Notation. In Atkinson, J. M., Heritage, J. (eds.): Structures of Social Action. Studies in Conversation Analysis. Cambridge University Press, Cambridge (1984)
27. Drew, P.: Accusations: The Occasioned Use of Members' Knowledge of "Religious Geography" in Describing Events. Sociology 12 (1978) 1-22

28. Schegloff, E. A.: Notes on a Conversational Practice: Formulating Place. In Sudnow, D. N. (ed.): Studies in Social Interaction. The Free Press, New York (1972) 75-119.
29. Sacks, H.: On the Analyzability of Stories by Children. In Gumperz, J. J., Hymes, D. (eds.): Directions in Sociolinguistics: The Ethnography of Communication. Holt, Reinhart and Winston, New York (1972a)
30. Sacks, H.: An Initial Investigation of the Usability of Conversational Data for Doing Sociology. In Sudnow, D. N. (ed.): Studies in Social Interaction. The Free Press, New York (1972b)
31. Kasesniemi, E.-L.: Mobile Messages. Young People and a New Communication Culture. Tampere University Press, Tampere (2003)
32. Lorente, S.: Special Number on "Youth and Mobile Telephones". Revista de estudios de juventud 57 (juny) (2002)