On Sound in Mobile Multimedia

Ilpo Koskinen

Prof., Dr. (sociology)

School of Design, Industrial Design

University of Art and Design Helsinki, Hämeentie 135 C, 00560 Helsinki

+358-50-329-6021 (GSM) // +358-9-756-30345 (fax)

ikoskine@uiah.fi // http://www.uiah.fi/~ikoskine/

Acknowledgements. I would like to thank ** for comments, Radiolinja for data, and people I have studied for allowing me to use their messages as data. ((Print: 7/22/2004: 6576 wrds, 40 281 chrc))

Transcript Symbols

Audio files have been transcribed using a system that follows, but is slightly simplified, from the system developed by Gail Jefferson (1984).

(.)	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 period indicates a stopping fall in tone.
,	A comma indicates a slight fall in tone.
?	A question mark indicates a rising inflection.
?,,	A combined question mark/comma indicates 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, and the same in English.
.nff	Snuffling.
#that's true#	Creaky voice.
@what@	Markedly different tone than elsewhere.
$what's that$	Laughingly.
W(h)hat	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.

Abstract

Sound is an important element of present-day surrounding. However, most research on the use of sound is based on analyses of Sony Walkman. This paper analyzes mobile multimedia devices that, unlike the Walkman, connect people back to the surrounding soundscapes, making it not just observable, but also recordable and reportable for various purposes. The focus is on the foreground as well as on ambient sound, the former defined as the most pominen sound element in the 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 foreground sound, how ambient sound makes place available for recipients, and how it makes social interaction hearable through the multimedia device.

Key words

Mobile multimedia – sound – ambient sound – ethnomethodology – interaction

Soundscapes and Mobile Technology

As Corbin (1998) shows in his study of the soundscapes in 19th century French villages, church bells defined the village space, overriding most other sounds like animal noises and market cries. In mapping soundscapes in 16th century England, Smith (1999) noted a clear class division in ambient sound: upper classes distinguished themselves from sounds. A similar difference is described in the antebellum South, while in the North, the sounds of trade and industry were taken as signs of productive activity that, for this reason, was to be endured (Smith, M. 2001). Sound thus had moral qualities. Contrasting these descriptions with modern world, we see how the distinction between the rich and the poor has survived. Also, with the exception of deep countryside, people live in a thoroughly mechanistic sound environment of internal combustion and jet engines, crowds, trams, and ubiquitous air conditioning.

As these historical studies show, sound is a meaningful entity to people in many ways. People always live amidst of a multitude of sounds, making inferences and judgments from by it. Some sounds are pleasant, some not, and some prompt contradictory opinions. Sound also has commercial meaning, as a quick look at real estate advertisements or hotel listings tells. It also has institutional relevance: cities and states define appropriate sound levels for administrative and public health purposes.

However, by far the most thoroughly researched area of portable sounds in the social sciences is the Sony Walkman. In contrast to early public discourse, which

tended to condemn the individualizing effect of the Walkman (see du Gay et al. 1997: 89-93), early researchers argued that it reorganized the users' perception of the urban landscape. In effect, the Walkman deterritorializes the city by distancing the user's experience from the city around him; it was a liberating technology (Hosokawa 1984; Chambers 1990, quoted in du Gay et al. 1997). It also acted as an interaction shield (Goffman 1963: 40). This view was criticized for forgetting the uneven distribution of the Walkman and the fact that this is what the Walkman was designed for in the first place (du Gay et al. 1997: 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 in order to hear the soundscape around them in order to navigate the city (Thibaud 2003).

What about the opposite process, re-sensitizing ears to soundscapes? How sound enters the digital experience through portable smart objects, capable of making recordings and relating sounds to photographs? The first line of evidence comes from digital cameras. Most digital cameras have an audiocapture function. Typically, they are able to capture a few seconds of sounds when the switch is pressed. Still, the microphone is of low quality, but capable of capturing some elements from the soundscape. Furthermore, annotation is typically limited to file names and indexes, giving little flexibility to the system.

However, in quickly spreading mobile multimedia, the technical and the use environment is different from Walkmans and digital cameras. Essentially, this means mobile phones and PDAs (personal digital assistants) equipped with a camera and either a MPEG, MP3, or MMS technology. The devices typically have a multimedia component. Many kinds of observational devices coexist in one accessory: there is a camera as well as audio and sometimes video capture functions. People can

furthermore annotate images with text easily, bringing in a discursive element into the process (see Koskinen et al. 2002; Daisuke and Ito 2003; Ling 2004; Scifo 2004). With these devices, the audible world becomes observable and reportable for practical – and other – purposes. This work may vary from simply capturing sound to explaining it, and transforming it into a travesty.

Sound and Mobile Multimedia

The best evidence on how people link photographs with audio comes from a series of sociologically informed studied in the technology development context. In Frohlich and Tallyn's study (1999), four UK families were given an audiocamera to use on summer vacation; all had a multimedia PC. These "audiocameras" were "experience prototypes" (Buchenau and Fulton Suri 2000), 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.

In photograph-supported interviews, Frohlich and Tallyn (1999) learned that audio has many uses in the context of photography. *First*, ambient (i.e. natural) sounds occurring just before, during, or after the photograph was taken. 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. *Secondly*, such 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. *Finally*, sound enhances

memories: it revives memory better than a mere picture. (See also Frohlich et al. 2002).

Perhaps the most important message of this study was that sounds and photographs work in two directions. Sometimes people captured sound and *then* add a photograph to index that. In particular, this is the case of 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 in them.

If we extrapolate from this study, we get a conjecture on 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 also show that users can annotate images with audio and text alike. They can also hear sounds and annotate them with photographs and text. Finally, they may create constellations in which various multimedia elements can support each other, disagree with each other, conflict with each other, and so forth.

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 of audio track. However, this shortcoming is alleviated with 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 takes these processes to interaction: people can communicate our experience with mobile multimedia right away. In their turn, distant others can ask if they do not understand something (compare to Koskinen 2004).

Foreground and Ambient Sound in Mobile Multimedia

In going about their lives, people not only listen to their environment, but also make sense of it. They get ideas and memories from sounds, sights, smell, and other sensory features of the environment. Sound, like text, can capture these 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. People may also add color and meaning to messages by, for instance, making ironic or humorous statements next to the message (see Battarbee and Koskinen 2004). These make the “*foreground*” of the sound environment: they are the most prominent sound elements in the message, elements that can hardly go unnoticed.

However, the microphone also captures other sounds, often inadvertently. In the background of the intended message, there are sounds from the cityscape, people, animals, objects, music, street noise, sounds of traffic, people talking, and steps in the sidewalks. These are “*ambient*” sounds: elements in the background, typically but no necessarily attached to messages unintentionally: the point of the message lies in the image, in text, or in talk. Of course, people can change places to get certain sounds, or to get away from certain other sounds. Still, the world captured by mobile devices is far from the carefully construed soundscape of radio, advertising, or commercial audiovisual culture.

Within the message, ambient sound is typically something in the background, often barely hearable and unspecific, but people can do various interpretations from it anyway. What is involved in this work is what Harold Garfinkel once called “the documentary method of interpretation.”

> The method consists of treating an actual appearance as “the document of,” as “pointing to,” as “standing on behalf of” a presupposed underlying pattern. Not only is the underlying pattern derived from its individual documentary evidences, but the individual documentary evidences, in their turn, are interpreted on the basis of “what is known” about the underlying pattern. Each is used to elaborate the other. (Garfinkel 1967: 78).

When people hear ambient sounds, they treat them as evidence of a pattern. Thus, ambient talk can point to a café or restaurant, but just as well to the street. 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. For example, when callers hear that the recipient is in noisy surroundings, they typically ask whether the recipient can talk, or start to talk louder.

How such understanding evolves depends essentially on how sound develops (see Nyíri 2002 for a similar point regarding animations; for video, Francis and Hart 1997; Ihde 1976 gives a phenomenological statement). When listening, the listeners are situated in a lived organization that existed at the time of the capture. Depending on how sound develops, action is available from ambient sound in many ways. For instance, people can get an understanding of what is going on by hearing how people participate in it. Hearing an order, price, handing in the money and thank yous tells about a routine commercial transaction. Baby talk followed by laughter tells about interacting with a baby. Thus, although any element in a multimedia message is indexical and can be heard in many ways, this is not typical to ordinary action. Each

element in the message – including ambient sounds – get a definite meaning from other elements in the message in a "reflexive" fashion. (See Garfinkel 1967: 1-11).

The claim of this article is that people use and make sense of sound in mobile multimedia at two 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, as much as this typically escapes the users' and the analysts' attention.

Data and Methods

In the Radiolinja MMS Study (from now on, *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 11-20, 2002, and lasted about 5 weeks. Each user was given a MMS phone (either Nokia 7650 with an integrated camera or 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 city-countryside axis. Exact numbers are confidential, but the following figures point the scale of messaging in the pilot. In all, users sent over 4000 messages during the pilot. Over 2000 were unique (the rest being duplicates in group messages, or recycled messages). These data were produced through the Radiolinja system automatically. The service was free of charge. This study was done with industrial designers Esko Kurvinen and Katja Battarbee, and several assistants.

For this paper, we have treated these data in the following fashion. From the vast mass of *Radiolinja* messages, we have chosen a subsample that consists 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 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 (Jefferson 1984), but with a few differences. Since most ambient sound, the focus of this paper, are too unclear for exact transcription, we have timed them, and placed them to separate "tracks" in the transcription to get better at their mutual relations.

In all, there were 72 audio messages, i.e. files with an audio component. However, most of these were send as copies: only 14 are original, the rest like the baby example analyzed later in this paper, sent to 10 persons. The length of the audio clip ranged from 4 to 24 seconds. With one exception of 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 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.

The analysis proceeded in the following fashion. We selected a subsample of 543 messages. In analyzing data in detail, the first phase was unmotivated search for similarities and recurrent issues, the second 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 restrict the applicability of the results to other data.

Sound in the Message Foreground

In each message in these data, there is a dominating sound element that is available to the recipient, and apparently meant to dominate the message. What do these foreground sounds do in mobile multimedia messaging? What is their 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 thee 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 with no text and an image that was apparently unrelated to the content of the message, Niko sent his greetings to Anne from the suburb of Mellunmäki, telling that he is in a “work camp,” describing his workplace in less flattering terms.

To given an example of a greeting, the following message is a birthday greeting for Markku, who turns 30. This example is essentially a singing postcard. In it, Anne and his boyfriend/husband sing the cliché–like Happy Birthday (in Finnish) to Markku, add a picture of flowers, and a textual greeting to the message. The only flaw in this performance is that the man is first out of tune and rhythm when he joins Anne. She acknowledges his problems with laughter tokens while singing. Later in the text, they account for the flaw in describing themselves as “honeytones” in the underwriting.

<table>
<tr><td colspan="3">Message 1.</td></tr>
<tr><td></td><td>091_0725_0812_08_272Anne_to_Markku.psd</td><td></td></tr>
<tr><td>14 sec</td><td>  </td><td>Text:
May the hero have a sunny 30th birthday!
Br.
Honeytones</td></tr>
<tr><td rowspan="2">Timeline (sec)</td><td colspan="2">Audio elements</td></tr>
<tr><td>Woman (above) and man (below)</td><td>Ambient</td></tr>
<tr><td>2</td><td>((singing)) Happy bir[thday to you, happy
[*day to you ((joins</td><td rowspan="3">((no recognizable ambient sound))</td></tr>
<tr><td></td><td>birt(h)(h)hday to you, happy birthday to Markku,h
singing, first out of tune and rhytm - - - - - - - - - -</td></tr>
<tr><td></td><td>Happy /birthday] to you ((at the end, voice is too high))
- - - - - - - - -))]</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

The second prominent use of sound in the foreground is for sound samples. Typically, sampled were babies (see Message 4 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 Message 3 below shows, voice could be used

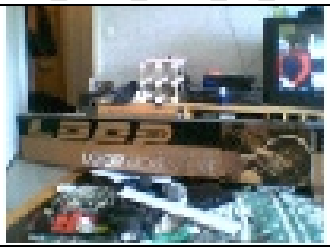
instead of a call to assist decision-making. In that message, Jaana sent a message of cushion covers she had seen in a shop to Anne to learn whether she wants them or not. However, in these data, this is the only case in which audio files were used to coordinate actions in commercial or institutional 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, added with some kind of yell or other emotionally loaded sound. For example, Arne sent once a message evaluating his recent cruise by noting in 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 (see Frohlich and Tallyn 1999; Frohlich et al. 2002).

The final thing worth noting about sound in the foreground of the message is that with two exceptions, it was always used in a "first position," that is, it was used to initiate action rather than to respond to it. In the simpler of these two cases, the audio message was used to return a greeting. In a more complex case, Tonya sent her "Have a nice holiday" greetings to Arne. Earlier that day, Arne had sent a message in which he told that his holiday had just started, and added a loud, happy howl to the message. He had continued to advertise his holiday mood with the ostrich imitation mentioned earlier, and a mystical rhetorical question. In consequence, Tonya's response took these signs of his mood into account. She wished him a "jaunty" holiday, and a graphic image of a cocktail glass. With these items she, then, showed what she thinks Arne is doing – and what is going to make his holiday "jaunty."

Hearing Place in Ambient Sound

The soundscape in mobile multimedia is considerably more complex than the foreground. This is not to say that sound in the foreground is a simple thing. By listening to the foreground only, recipients get an idea of what the sender has intended to say, his mood, how he has assessed his experiences, his wits, and also his company. Furthermore, there are many kinds of "performatives" at work in the foreground. In particular, audio may used to elicit responses from the recipient. However, when recipients turn their attention to ambient sound, they get access to a host of other, more contextual aspects of what is going on. Next we turn to these.

In the following, there are two minimal cases, one with non-human, another with a human ambient sound. In Message 2, there is an obscure item in the photograph. The text tells that there is a fly fishing gear in the image. Laughter in the foreground is so loud and exaggerated that the sender's joy cannot go unnoticed. The sound may tell either about the long waiting before the purchase, or about the future days spent in fishing. Most likely, it tells about both things simultaneously.

<table>
<tr><td colspan="3">Message 2.</td></tr>
<tr><td></td><td>153_0726_1208_08_240Jan_to_Arne.psd</td><td></td></tr>
<tr><td>6 sec</td><td>  </td><td>Text:
I bought fly fishing equipment.</td></tr>
<tr><td rowspan="2">Timeline (sec)</td><td colspan="2">Audio elements</td></tr>
<tr><td>Man</td><td>Ambient</td></tr>
<tr><td>0.3</td><td>HEH HEH HEH HEH HEH</td><td>((strong echo,</td></tr>
<tr><td>2</td><td><HE HE HE HE HE HEE::e:: hoe:::>.</td><td>alone in a</td></tr>
<tr><td>5</td><td>((silent))</td><td>room))</td></tr>
</table>

The ambient sound is strong echo typical to a room that is close to empty. The message is situated inside rather than outside. There are no sounds of the wind or the street; instead, the message is compiled in a quiet environment. The sender is also apparently alone in the room: nobody laughs with him. Given the picture that shows consumer electronics in a messy room, a good guess is that the picture is taken at Jan's home.

In the next example, the action is easy enough to understand from the point of the message. Jaana has promised to check the cover for a cushion 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 making querying whether she should buy it anyway. Interestingly, Jaana also corrects an element in the image with audio (see Frohlich and Tallyn 1999) 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; we do not have access to the reply that, if it came, was probably a call. This message is a part of a more extended action that has a history and a future course.

Here the ambient sound consists of background noise and a commercial announcement. Notice that there is no mention of *where* Jaana is shopping. It could be any commercial venue – a shopping mall, department store, or just a shop. As the "here I am" in opening of the message implies, Jaana and Anne have discussed about 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 to a shopping mall, for two reasons. First, the background noise tells from the very beginning that the place is large. There is strong echo in the sound, and the talking crowd is sizeable, thus ruling out a local shop or a boutique. When the commercial announcement comes in after 12

seconds, the remaining alternatives are either a shopping mall or a large department store.[1] In Message 2, no similar stepwise development took place in the ambient sound.

Message 3.		
	029_0724_1410_08_897Jaana_to_Anne.psd	
18 sec.		((no text))
Timeline (sec)	Audio elements	
	Woman	Ambient
1	.pth Here I am, Anne (.) There's the middle one among the cushion covers (.) the colors can be seen pretty badly?, (.)	((background noise,
9	But ehm h (.) is is a quite pretty blue?, /But Ehm .h send me a message	talking crowd, strong echo;
12	I'll take it with me:::,h (.) so do you wa:nt is of not (.) /Bye h	After 12 sec: a commercial announcement in the background))

These cases show how audio and other multimedia elements work together to situate action into certain place. Importantly, place tells about action: people may infer what others are doing *from* knowledge of the place. In the second example, the sender is in a shopping mall; apparently, shopping is the default activity and mindset there. In the first, the sender is in a calm, peaceful setting; what takes place there is more ambiguous. However, people may also infer place from activity. In the shopping mall case, there are no identifiers of place, but still we manage to hear it that way. Importantly, certain places are linked to certain activities (compare to Schegloff 1972: 102-105). 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 Drew 1978 for how places and geography functions in talk; a more formal treatment is in Schegloff 1972: 96-106). In this sense, even minimalist ambient soundscape does important work in mobile multimedia messaging.

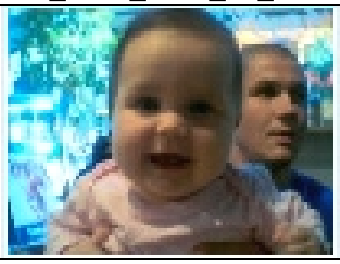
Hearing Interaction

More happens in ambient sound than mere situating to a place, and things that follow from this bind. In particular, ambient sound may make social action available in a considerably more nuanced manner. In some cases, hearing the ambient sound tells us vividly what is going on regardless of text or the fact that in *Radiolinja*, all messages contained only still images.

In the following message (Message 4), there is a picture of a baby against blue water, suggesting that the image was taken in a swimming hall. This is confirmed by the text, which situates the message to “East Center,” which is a large shopping mall area roughly 10 km east from downtown Helsinki. In the premises of the mall, there is a swimming hall, which also arranges swimming for babies. On the top, the whole message is offered as a ”Greeting,” sent to ten persons.

The camera has captured a laughing baby. The sound tells more about what is going on in the scene. At first, we hear the sounds of the baby and a few adults. The baby “Zoewy” (nickname) is bathing with her mother,[2] and keeps cheerful noise all along. Adults nearby are laughing, and apparently amused by the baby. At the end of the message Zoewy gets food from the bottle. For the sender, the message becomes a vehicle for sharing a delightful, which also justifies sending the message.

But this is the foreground sound only. An analysis of how the ambient sound develops gives us a nuanced idea of what happens in the message. We first hear a baby attempting to talk, and see a young man in the background (who he is remains unclear). Then after 4 seconds, a woman laughs near to the microphone. It is at this moment that we first hear a key element of the social organization: it is the baby's mom. Located next to the microphone, these two become hearably units of a pair (see Sacks 1972a, 1972b). After 7 seconds, the mother talks again near to the microphone. After 11 seconds, she turns her head away from the microphone and talks to other adults, who respond to her. By now, all participants are known. However, the episode gets to the second phase when Zoewy starts her baby talk again (13.3–16 seconds). In response, the mother talks to other adults, and laughs briefly with them. After two seconds, Zoewy again starts to shout, this time much louder. 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 Susan keeps Zoewy in her arms and feeds her.

<table>
<tr><td colspan="4">Message 4. Sent to 10 persons.</td></tr>
<tr><td></td><td colspan="2">006_0724_1237_02_272Susan_to_Arne.psd</td><td></td></tr>
<tr><td>26 sec</td><td colspan="2">  </td><td>Text:
Greetings from East Centre, t. Zoewy</td></tr>
<tr><td rowspan="2">Timeline (sec)</td><td colspan="3">Audio elements</td></tr>
<tr><td>Baby</td><td>Woman</td><td>Ambient</td></tr>
<tr><td>1-2</td><td>quiet noise, attempts to talk</td><td></td><td></td></tr>
<tr><td>4</td><td></td><td>Laughs</td><td></td></tr>
<tr><td>7</td><td></td><td>Oh look, here it (comes)( )</td><td></td></tr>
<tr><td>11</td><td></td><td></td><td>the woman talks with other adults</td></tr>
</table>

13.3-16	Baby shouts louder		
17		(talks, inaudible)	((laughter))
19-23	Baby cries loud		
24		And then we take the bott- ((sound cut off))	

Message 4 provides a nice example of how people can use common–sense knowledge of social structures (Garfinkel 1967) to make sense of what they hear. This scene tells not only about the mommy's and the baby's mood, but also that the pair is with other adults, who share the mother's joy, thus being relevant partners in the scene, not just bystanders. Recipients of the message not only hear several voices, but they also hear a key feature of on-going social organization. At first, the mother tries to get the baby to talk to the microphone. When she is successful, she is able to direct her attention to her adult company for a moment before the baby catches her attention again by getting hungry. Here, the ambient sound introduces a social dimension to the message, tells about its character, and tells about how the mother and the baby allocate their involvement between each other and the social surroundings.

Conclusions and Discussion

Unlike portable music devices such as the Walkman that people use to restructure their sonic experience while on move (for instance, Hosokawa 1984; du Gay et al. 1997; Thibaud 2003), mobile multimedia gives people new means to observe and report their activities and experiences with the environment (see Koskinen et al. 2002, 2004; Daisuke and Ito 2003; Ling 2004; Scifo 2004). 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. In the foreground, audio files are used for various purposes, ranging from greetings to sending sound samples, imitations of animal and humans, and to communicate emotions. It could also be used as a surrogate for a phone call or a voice mail message. Typically, audio was used to initiate action rather than to respond to on-going lines of action.

By ambient sound, the paper has meant those sound elements that are outside the main intention of the message: street noise in the back of an invitation, or laughter in a greeting card-like multimedia message. What have we learned about the uses of ambient sound in the *Radiolinja* data? Audio is not the most common element in multimedia messaging, but when it is used, it typically has ambient elements. In 14 messages analyzed for this paper, there were 28 ambient audio elements. Roughly 2/3 of these ambient elements consisted of human sounds, the rest being from other sources. Throughout the analysis, we have seen that ambient audio is a complicated element in mobile multimedia. It interprets images and text, but is also explicated by them. However, although it has various functions in mobile multimedia messaging, it is not an invention of the moment; there are a few main uses of ambient sound, and people use these in a methodic fashion in their messages.

As Frohlich and Tallyn (1999) suggest, audio “augments” images in many ways: it adds life to images, allowing people to communicate more fully. However, we have also seen that it gives recipients a more nuanced access to place and interaction, and allows a host of inferences about what is going on in the message. It would be hard to communicate all these meanings with text alone, as in SMS, or even with photographs augmented with text. The audio capability is a significant addition to the methodic repertoire of people using mobile multimedia.

Data poses some obvious limitations to an attempt to generalize this analysis to future mass-market phenomena. First, the study is from the very first months of mobile multimedia on the market. Secondly, the number of cases analyzed for this paper is small. Third, this analysis has focused on just one aspect of sound, ambience, not on every use of sound in mobile multimedia messaging. However, these are not crucial issues in terms of the validity of the present analysis. Although the data is from the early days of mobile multimedia, users were ordinary, not professionals in information technology, sound production, or photography. The small number of cases limits analysis, but does not make the observations of this paper wrong: the paper describes ordinary practices that no doubt will be found in larger data sets in the future. Larger data will in all likelihood reveal more nuances of audio use, but this does not threaten the validity of the analysis presented in this paper.

Thus, the analysis shows just how rich a method audio is, even when used by ordinary people with no training in creating, analyzing, interpreting, or planning it. The analysis, of course, is ultimately an attempt to explicate ways audio is used methodically in ordinary life. The final objective of this study has been excavating minuscule practices that are often so simple that they are taken for granted and consequently, go largely unnoticed in everyday life and in sociology alike. For design, such negligence is not possible. This is the final point to make: note that as these data comes from design studies, they tell about first explorations with technology rather than about mature technology. By the end of 2004, mobile multimedia is still too small a phenomenon to be studied with statistical means, or even with focus groups or ethnographies (see Ling 2004). However, since the focus is on the uses of mobile multimedia, this analysis will probably provide an idea of more mature uses as well.

People do not change their methods of social action as technology matures, although some of these methods may get more elaborate over time.

References

Battarbee, Katja and Ilpo Koskinen 2004. Co-Experience: User Experience as Interaction. *Co-design*, Vol. 1(1).

Buchenau, M. & Fulton Suri, J. (2000). Experience Prototyping. *Proceedings of DIS 2000*. New York: ACM, pp. 424-433.

Bull, Michael. 2000. *Sounding Out the City*. Oxford: Berg.

Bull, Michael and Les Back (eds.) 2003. *The Auditory Culture Reader*. Oxford: Berg.

Chambers, Ian 1990. *Border Dialogues. Journeys in Postmodernity*. London: Routledge. (Partly available in pp. 141-143 in Du Gay, Paul et al. 1997. *Doing Cultural Studies: The Story of the Sony Walkman*. London: Sage.)

Corbin, Alain 1998. *Village Bells: Sound and Meaning in the Nineteenth-Century French Countryside.* New York: Columbia University Press.

Daisuke, Okabe and Mitzuko Ito 2003. Camera Phones Changing the Definition of Picture-Worthy. *Japan Media Review.* Annenberg School for Communication, USC. www.ojr.org/japan/wireless/1062208524.php. Accessed April 15, 2004.

Drew, Paul 1978. Accusations: The Occasioned Use of Members' Knowledge of "Religious Geography" in Describing Events. *Sociology* 12: 1-22.

Du Gay, Paul et al. 1997. *Doing Cultural Studies: The Story of the Sony Walkman*. London: Sage.

Francis, David and Christopher Hart 1997. Narrative Intelligibility and Membership Categorization in a Television Commercial. In Hester, Stephen and Peter Eglin

(eds.) *Culture in Action. Studies in Membership Categorization Analysis.* Washington, D.C.: International Institute for Ethnomethodology and Conversation Analysis and University Press of America.

Frohlich, David and Ella Tallyn 1999. Audiophotography: Practice and Prospects. *Proceedings of CHI'99*, May 15-20, ACM.

Frohlich, David et al. 2002. Requirements for Photoware. *Proceedings of CSCW'02*, Nov. 16-20, New Orleans.

Garfinkel, Harold 1967. *Studies in Ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall.

Goffman, Erving 1963. *Behavior in Public Places.* New York: The Free Press.

Hosokawa, Shuhei 1984. The Walkman Effect. *Popular Music* 4: 165-180.

Ihde, Don 1976. *Listening and Voice. A Phenomenology of Sound.* Athens, OH: Ohio University Press.

Jefferson, Gail 1984. Transcript Notation. In Atkinson, J. Maxwell and John Heritage (eds.) *Structures of Social Action. Studies in Conversation Analysis.* Cambridge: Cambridge University Press.

Koskinen, Ilpo 2004. Seeing with Mobile Images: Notes on Collaborative Seeing in MMS. Presented at *COMMUNICATIONS IN THE 21ST CENTURY: The Mobile Information Society*, Budapest, June 10–12, 2004.

Koskinen, Ilpo, Esko Kurvinen, and Turo-Kimmo Lehtonen (2002). *Mobile Image*. Helsinki: IT Press.

Ling, Rich 2004. The Development of Grounded Genres in Multimedia Messaging Systems (MMS) among Mobile Professionals. Presented at *COMMUNICATIONS IN THE 21ST CENTURY: The Mobile Information Society*, Budapest, June 10–12, 2004.

Nyíri, Kristóf 2003. Pictorial Meaning and Mobile Communication. In Nyíri, Kristóf (ed.) *Mobile Communication. Essays on Cognition and Community*. Vienna: Passagen Verlag.

Sacks, Harvey 1972a. On the Analyzability of Stories by Children. In Gumperz, John J. and Dell Hymes (eds.) *Directions in Sociolinguistics: The Ethnography of Communication.* New York: Holt, Reinhart and Winston.

Sacks, Harvey 1972b. An Initial Investigation of the Usability of Conversational Data for Doing Sociology. In Sudnow, David N. (ed.) *Studies in Social Interaction.* New York: The Free Press.

Schegloff, Emaniel A. 1972. Notes on a Conversational Practice: Formulating Place. In Sudnow, David (ed.) *Studies in Social Interaction.* New York: The Free Press. pp. 75-119.

Scifo, Barbara 2004. The Domestication of the Camera Phone and MMS Communications. The Experience of Young Italians. Presented at *COMMUNICATIONS IN THE 21ST CENTURY: The Mobile Information Society*, Budapest, June 10–12, 2004.

Smith, Bruce R. 1999. *The Acoustic World of Early Modern England. Attending to the O Factor.* Chicago: University of Chicago Press.

Smith, Mark M. 2001. *Listening to Nineteenth-Century America.* Chapel Hill: The University of North Carolina Press.

Thibaud, Jean-Paul 2003. The Sonic Composition of the City. In Bull, Michael and Les Back (eds.). *The Auditory Culture Reader.* Oxford: Berg.

Notes

[1] Of course, more fine nuances on what Anne and Jaana are doing here are available in this message. Notice the pivotal structure in her characterization. ”you see the colors kinda badly?, (.) but ehm h (.) it is a fairly pretty blue?,” Here she shows that she knows what Jaana prefers, and that she empathizes with her because her color is not there. She also shows her own judgment, which she offers as an indirect suggestion for buying the second best alternative, but without taking issue with Jaana’s taste. She does not say ”you would like it”, but keeps the talk within her own authority. However, she may be using a question intonation here, though, to soften the judgment that might be heard as too direct in other ways. Should she suggest immediately that the covers are beautiful, though not in the color preferred by Jaana, she would make her action understandable to Jaana as impatient: ”Anne wants to get out, she does not care about by preference, but just puts her own intentions before my wishes.”

[2] However, note that nothing in the message tells that Susan is Zoewy’s mother. We just take it for granted that the woman who accompanies a small baby and picks her up to cuddle her and to satisfy her needs is her mother. – See Sacks (1972a,b) for an elaboration of this point.