Scribe4Me: Evaluating a Mobile Sound Transcription Tool for the Deaf

Tara Matthews[1], Scott Carter[1], Carol Pai[2], Janette Fong[2], and Jennifer Mankoff[2]

[1] Berkeley Institute of Design, CS Division, University of California, Berkeley, CA, USA
{tmatthew, sacarter}@cs.berkeley.edu
[2] Human Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, USA
{magicpai, janette.fong}@gmail.com, jmankoff@cs.cmu.edu

Abstract. People who are deaf or hard-of-hearing may have challenges communicating with others *via* spoken words and may have challenges being aware of audio events in their environments. This is especially true in public places, which may not have accessible ways of communicating announcements and other audio events. In this paper, we present the design and evaluation of a mobile sound transcription tool for the deaf and hard-of-hearing. Our tool, Scribe4Me, is designed to improve awareness of sound-based information in any location. When a button is pushed on the tool, a transcription of the last 30 seconds of sound is given to the user in a text message. Transcriptions include dialog and descriptions of environmental sounds. We describe a 2-week field study of an exploratory prototype, which shows that our approach is feasible, highlights particular contexts in which it is useful, and provides information about what should be contained in transcriptions.

1 Introduction

"[I was] appreciative that [the tool] was available… to fill in the large gaping holes of conversation in a group I usually miss… I saw how [the translations] could create the gift of a conversation." –A study participant.

Sound plays an important role in communication and contextual awareness about interesting events and information. These sounds, and the information they convey, may not be easily available to people who are deaf or hard-of-hearing. The home environment can be controlled and augmented to support better awareness using alerting systems for things like phones, doorbells, alarms and a baby's crying (see [17] for a review of existing techniques and the sounds they support). However, other areas encountered in daily life (*e.g.,* public spaces, stores, restaurants, streets, airports, public transportation, and so on) do not always provide adequate support for communicating sound-based information to the deaf. For this reason, mobile support for better sound information awareness would be of great value to the deaf and hard-of-hearing.

We present the design and evaluation of a mobile sound transcription tool for the deaf and hard-of-hearing called Scribe4Me. When a user presses a button on her Scribe4Me PDA, the last 30 seconds of sound is uploaded, transcribed and sent back to her as a text message. Transcriptions include dialog and descriptions of environmental sounds. Scribe4Me is unique in providing support for both speech and

P. Dourish and A. Friday (Eds.): Ubicomp 2006, LNCS 4206, pp. 159 – 176, 2006.

non-speech audio in mobile environments. It is also the only speech to text system for the deaf that is robust enough to be deployed in an unconstrained setting for two weeks.

Our main goal was to explore the potential uses of and issues surrounding a sound transcription tool. We conducted a 2-week field study of an exploratory prototype, gathering data about situations in which requests were made, transcriptions were sent, and gathering qualitative feedback from users (see Fig. 1). Results showed that despite issues such as a 3-5 minute delay, the prototype tool was useful in a variety of situations and most users were excited by it, especially for understanding conversations: "I don't really get all that much of the conversation and subside on a few words here and there. There were several times I used the tool… I was privy to the complete sentences, which was very nice. I very much enjoyed having it available…" Scribe4Me was useful to people with varying levels of hearing loss, from people who are profoundly deaf to those using cochlear implants and hearing aids.

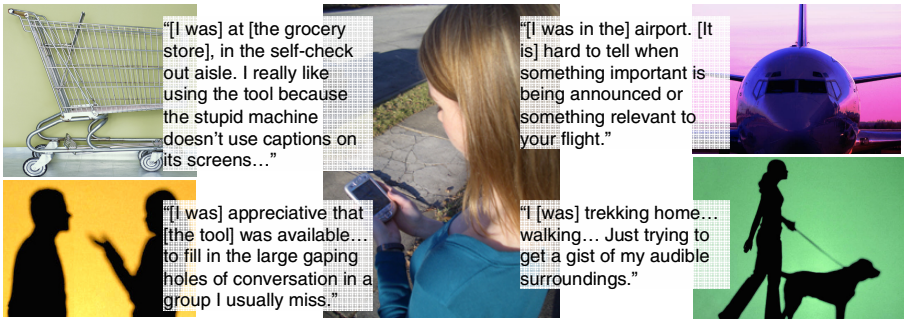

Fig. 1. Field study users found the Scribe4Me tool valuable in many situations, including at grocery stores, in group conversations, in airports, and moving about the city

2 Related Work

Research for the deaf has often focused on communication support in both *mobile* and *non-mobile* settings. However, most commercial technologies focus on *non-speech audio* in *non-mobile* environments. This section provides an overview of research and products supporting transcription and review of speech and non-speech audio. We also discuss a related area that has received attention outside the assistive technology community: mobile tools for automatic storage and review of audio.

Mobile Communication: Communication options used by the deaf have changed radically since the introduction of mobile phones and PDAs. Text messaging is common among the deaf and has increased communication between the hearing and deaf [20]. WISDOM (Wireless Information Services for Deaf People on the Move) [8] was a collaborative effort between several research organizations and companies in Europe, focused on developing video sign language transmission on mobile devices (see [1] for a description of their vision-based sign language recognition system). No products exist as a result of this project. Another commercial software

product, LipC-Cell [4], enables lip reading during mobile phone calls. The mobile phone connects to a PC, which converts the caller's voice into a 3D animated face with real-time lip movements (companies providing LipC-Cell are no longer in business and the software is not available). A number of wearable vibrotactile speech perception aids have been developed that extract voice fundamental frequency and deliver it *via* vibration. These systems improve speech perception for the deaf (see [2] for a survey). Impromptu is a mobile system for audio applications with speech interfaces, using speech recognition [21]. Though not intended for users who are deaf, Impromptu demonstrates mobile speech recognition. None of these tools can handle unconstrained audio recordings of varying quality and none include both speech and non-speech sounds.

Non-Mobile Communication: Assistive technology for the deaf used in static environments has also focused on communication. Common technologies include: assistive listening devices (improving the audibility of one sound source that is likely to be lost due to distance or background sounds such as a lecturer in an auditorium or a conversation in a loud restaurant); telecommunication devices (such as text telephones (TTDs), IP relay, and video relay services [6]); and close-captioning for TV and movies [7, 16]. Communication support in classrooms includes captioned dialogue with educational transcription services, computer-assisted note-taking, and, more recently, automatic speech recognition programs [9]. Research has also explored automatic sign language recognition. Edwards [10] summarizes work on developing techniques for capturing, segmenting (delimiting), and classifying sign language gestures. A number of systems have been developed to automatically translate English into American Sign Language (see [13] for a survey). Several other systems enable speech articulation practice. Sonido offers software that visualizes speech using spectrographs and allows users to practice speech articulation with recorded speech samples [22]. Ellsman and Maki study the effectiveness of spectrographs in speech training, suggesting that they can enable students to practice alone to a limited degree [11]. Finally, many systems translate speech between languages, *e.g.*, [3, 15, 19, 25].

Non-Mobile, Non-speech Audio: Other non-mobile systems enable non-speech sound awareness. Many commercial alerting systems can be installed in buildings and homes. These typically use flashing lights, vibration, or extra-loud sounds to provide awareness of alerting sounds (phones, doorbells, emergency alarms, and babies crying). However, alerting system are expensive, difficult to install, and non-mobile. In related past work on peripheral visualizations of non-speech sounds, we interviewed people who are deaf about sound awareness needs [17] (discussed in the next section).

Mobile Audio Buffer: Outside of assistive technology, researchers have explored mobile tools for recording sound for later review as a memory aid. For example, the Personal Audio Loop (PAL) [14] helps people avoid conversational break-downs that occur when a person forgets something that was said (*e.g.,* a conversation topic or a name). PAL retains audio for about one hour, enabling users to replay conversations. Though not aimed at the deaf, PAL introduces privacy issues relevant to Scribe4Me. The unobtrusive nature of PAL makes it easy for the speech of people near the user to be recorded without their knowing or consenting. Use of Scribe4Me is hard to hide from conversation partners, can be hidden from passersby. In 38 of 50 states in the U.S., it is legal to record conversations to which you are a party without informing others [23]. PAL creators explored the ethics of this issue using an inquiry technique

in which potential PAL users asked conversation partners to complete a survey after a conversation. The survey asked about privacy and consent preferences if the conversation had been recorded by PAL. Results showed that people wanted to be informed of and asked to give consent to audio recording and its replay to others *a priori*.

In summary, past research looks at mobile and non-mobile communication (*e.g.,* video relay on a PDA or PC), non-mobile sound awareness systems (*e.g.,* phone ring flashers), and privacy in recording audio. Notably missing from past work is a nuanced understanding of what *speech* and *non-speech sound* awareness is valued in *mobile* situations. Scribe4Me enables us to explore sound awareness user needs, which are further motivated in the next section.

3 Formative Work

In past work, we interviewed people who are deaf about sound awareness needs to inform peripheral, visual displays of non-speech sounds [17]. Participants emphasized the importance of sound-based awareness in *all locations*, something not adequately supported by existing technology. They listed many sounds of interest outside (*e.g.,* vehicles, people approaching, *etc.*), in the home (*e.g.,* appliances, knocking, TV, *etc.*), and in the office (*e.g.,* activities of coworkers, printers, phones, *etc.*). Participants sometimes relied on third party descriptions of sounds. One participant reported calling her hearing husband and putting the phone near the baby monitor when she was unsure about the sounds the baby was making.

Our formative interviews motivated the work described here. Based on those results [17], we decided to explore the idea of describing recent environmental sounds to people based on recordings made on a mobile tool. After feasibility testing to check that contextual audio recorded with a cellphone-quality microphone was understandable to a human (suggesting that transcription was at least possible), we conducted two Wizard-of-Oz pilot studies with four participants each. We followed participants and transcribed sounds by writing a description of them on paper when asked. We focused on testing feasibility, identifying situations in which the tool was useful and learning about what information to include in transcriptions.

The first pilot study included four hearing graduate students. One participant listened to music using ear-buds and the other three wore earplugs, to reduce hearing accuracy. We instructed participants to walk around the campus and surrounding city streets for 20 minutes, and did not assign a specific task. Participants were instructed to raise one hand to ask "what happened?" A researcher followed each participant and wrote a description of recent sounds on paper when a hand was raised, which was shown to the participant as soon as the transcription was complete. We transcribed a number of events onto paper, including short descriptions of ambient sounds ("two girls behind you are talking"), descriptions of speech ("a woman walked by said she parked at Dicks"), and no information ("not sure what happened"). Transcriptions did not include verbatim speech transcriptions, something participants suggested we add. Feedback from this pilot indicated that most participants did not find a compelling use for the tool. Thus, we conducted another pilot study with more representative users over a longer period of time.

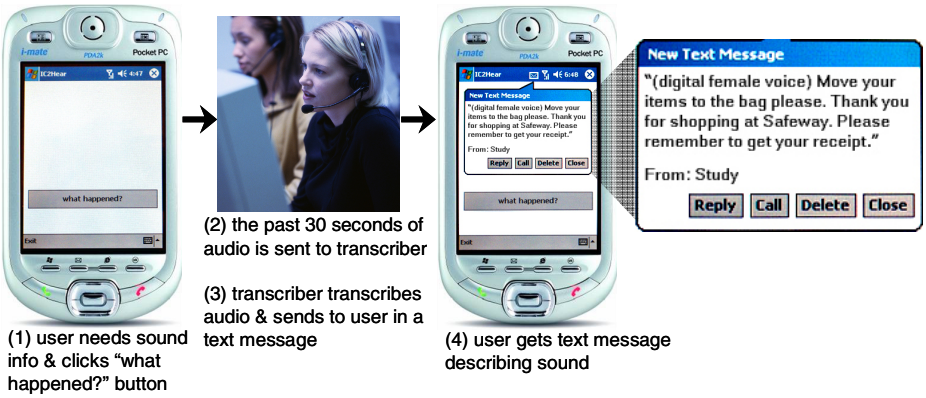

Fig. 2. System diagram showing the request process from left to right

In the second pilot study, which had similar goals to the first, we gathered feedback in more realistic situations. A researcher followed the user and transcribed audio on paper when a user pushed a portable, 1-button device (implemented using BOXES [12], a tool designed for early-stage prototyping of physically-based devices). Two people who are deaf participated while shopping at mall. Two people without hearing loss participated while grocery shopping. One also did a "staged" trip to a Greyhound bus station, in which she asked when the next bus was leaving. Study length varied from 40-90 minutes, depending on the activity.

All four participants found some compelling use for the tool. While grocery shopping, one participant used the tool because "I was anxious that the store was going to close soon because it's late, and I thought the announcement might have been asking us to check out." When at the Greyhound station, the same user used the tool because "I was trying to find out when the next bus was, and I thought there would be some kind of announcement, but I didn't hear anything. I was anxious that I missed it." In this particular instance, since there had been no distinguishable sounds, the tool had responded with, "Not sure what happened." The participant commented, "I liked that. It was comforting because it told me that I hadn't missed anything." A second participant was particularly interested in what others said, "…especially if they are talking about me. Don't be afraid to tell me bad things." Another participant didn't find a use for the tool in the mall, but thought the tool would be valuable in an airport or when "things didn't go according to plan." Suggested improvements centered on increasing the detail of transcriptions, especially including exactly what people nearby had said.

Overall, people in the *in situ* pilot found beneficial uses for the tool and speculated about others. In both pilot studies, the main feedback was that participants wanted the wizard to provide a complete transcript of the event instead of a summary (especially the exact dialog of others). Additionally, participants always asked for information about recent events, rather than events that had occurred some time ago. These results, along with strong user needs exposed in formative interviews, and gaps in existing technology, motivated the design, implementation, and field test of the mobile sound transcription tool discussed next.

4 Design and Implementation of Scribe4Me

Scribe4Me is a tool that provides descriptions and transcriptions of recent speech and sounds to users when they press a button. As illustrated in **Fig. 2**, Scribe4Me runs on a PDA and records 30 seconds of audio continuously. If the user wants more information about what happened in the last 30 seconds, she can *request* it by pushing a button on the screen labeled "what happened?" For each request, two 15 second audio files are sent *via* GPRS to a human transcriber, who sends back a text message describing the content of those files.

Our pilot studies helped to inform these design decisions. First, we learned that awareness of *recent* audio was useful to participants. Thus we designed Scribe4Me to buffer the last 30 seconds of audio. Second, participants wanted detailed information conveyed by both speech and non-speech sounds. To find out just how much detail is required, we designed Scribe4Me to include a human-in-the-loop to transcribe all possible sounds in detail. The field study would inform whether a human transcriber or automated sound recognition would be used in future iterations.

We implemented a functional prototype of Scribe4Me using Momento [5]. Momento is a tool designed to support rapid prototyping and situated evaluation of mobile Ubicomp applications. It has built-in support for long-term studies of `human-in-the-loop systems, of the sort we were conducting. Momento was used to implement both the mobile interface and to support the transcription by the remote transcriber.

4.1 Mobile Client Implementation

The mobile part of the prototype was an alpha version of the Momento mobile client, configured to capture and transmit audio. The only additional coding needed was to create the UI screens shown in **Fig. 2**. While sending files, the "what happened?" button was disabled and progress messages were shown: "Sending file 1 of 3 of your request;" "Sending file 2 of 3. We've begun translation;" and "Sending file 3 of 3. You should receive a text message soon." The entire process from making a request to receiving a response took 3-5 minutes. This time was composed of GPRS transmission (2-4 minutes) and transcription time (about 1 minute). We used a Windows Mobile version of the Momento mobile client, rather than a J2ME version, because it was faster and less error prone on our deployment devices. Because Scribe4Me required a sensitive and accurate enough microphone to enable detailed transcriptions, we used Pocket PC devices (five Qtek 9090 PDA2Ks and one i-Mate PDA2K).

Our goal was to minimize transmission time and maximize audio quality. Both audio quality and transmission time increase with larger file sizes. Audio is recorded by the Momento client as WAV files at 11KHz in stereo. We customized Momento to slow the audio sampling rate to 8KHz mono (file sizes dropped from 320 kb to 115 kb). Testing showed that these files were as understandable to a human transcriber as at the default sample rate, although 8KHz is the minimum for speech recognition systems. It also resulted in a 2-4 minute drop in transmission time.

Because this was a rough prototype intended for evaluation, we had to make several design decisions that would be changed in a more polished tool. First, the "what happened?" button was implemented in software, requiring the user to press it with a stylus. This software button was implemented by customizing Momento's mobile

interface to include a single button labeled "what happened?" Second, recording was paused while transmission was in progress due to limited device resources. Third, we asked users to take a photo with the PDA each time they pressed the "what happened?" button. The photo would not be part of a final tool but was important for providing the researchers and the participant with context about requests for later discussion in email journals and post-hoc interviews.

4.2 Remote Desktop Translation Implementation

The transcriber used the Momento tool to handle requests. Momento provided a generic interface for monitoring and responding to incoming multimedia data from mobile devices. When a request was made, Momento notified the transcriber with a sound. Using the Momento desktop interface, the transcriber listened to the incoming audio, typed a response, and sent it to the participant, who received it as a standard SMS (text) message. Longer transcriptions were split into multiple text messages due to software in the GSM network that imposes a 160 character size limit for SMS messages. The transcriber could begin translating the first audio file before other files finished transmitting.

The Momento desktop application coordinated communications with the PDAs. It provided a timeline visualization of all incoming and past requests. Furthermore, the application saved all communication logs in tab-separated files that could be imported into spreadsheet programs for later analysis.

5 User Study

We conducted a two-week field evaluation of a fully functional prototype with six participants who are deaf. The study helped us answer the following questions:

- Would Scribe4Me be useful for people with different levels of hearing loss?
- How would users decide when to use Scribe4Me?
- In what situations would Scribe4Me be useful?
- Would users value Scribe4Me enough to continue using it or to pay for it?
- What information is most useful to include in audio transcriptions? Are hard-to-transcribe details such as emotion valuable?
- What improvements could be made to Scribe4Me?
- How would others react to the participant using Scribe4Me?

5.1 Participants

We recruited 6 users living in two geographically distant, urban regions for the two week study: 2 profoundly deaf (hearing no sounds), 1 almost profoundly deaf (hearing only very loud sounds with little comprehension) and 3 hard-of-hearing with the help of hearing aids or cochlear implants (hearing level is highly variable; these participants all communicated verbally with others, but often missed sounds that were quiet, in noisy environments, or were not the focus of their attention). Four participants were female. Ages ranged from 25 to 51. Participants' occupations included a substitute

teacher/student, an information services worker, a clerk, the CEO of a technology company, a marketing director, and a college student. Participants were volunteers recruited *via* an email sent to several distribution lists for the deaf and hard-of-hearing.

Participants were compensated based on their participation, with $155 (USD) as the maximum for 100% participation. All participants had 100% participation, except one who completed only 8 days of the field study (due to scheduling conflicts). Table 1 describes user hearing levels and participation results.

Table 1. Participation results per user

User	Days	Total requests	Avg. requests per day	Hearing / aids
1	7	47	6.7	minimal hearing, no aids
2	10	23	2.3	profoundly deaf
3	13	19	1.5	1 aid, 1 implant (recent)
4	14	14	1.0	profoundly deaf
5	15	13	0.9	1 aid, deaf in other ear
6	8	2	0.3	2 aids

5.2 Method

During the two week study, transcribers transcribed requests 7 days a week, 9 am to 9 pm for east coast participants, and 6 am to 6 pm for west coast participants (this accomplished the goal of having 9 am to 6 pm covered for all participants). Data collected from each participant included demographic information, a great deal of qualitative feedback during four interviews, audio and photos from requests, and tool usage descriptions and feedback included in a daily email journal.

The study began with a one-hour meeting in which the user was introduced to Scribe4Me and trained to use it. During the field study, users completed daily journals (see Figure 3 for an example). We also conducted two 30-minute interviews over the phone or Instant Messenger (IM) on days 4 and 8 of the study. We asked each participant if he or she were having any problems using Scribe4Me or the PDA, when it had been most useful, if he or she had hesitated to use it in any situations, how usage of the tool had changed since the beginning of the study, what issues had affected the usefulness of the tool, and how we could have improved the transcriptions. We finished the study with a one hour exit interview, in which we asked each participant about overall impressions of the tool, situations in which it was used, ways in which it could be improved, and whether he or she would continue using it for free and for a charge.

During the field study, we sent each user a daily summary of the photo and transcription associated with each request he or she had made. We asked the user to send us an email journal each day answering a series of questions about each request, such as why it was made and how useful it was. A sample email journal from the study is shown in Fig. 3. Users responded to email journals even when they made no requests that day (answering questions about what they did, why they did not need to use the tool that day, and if they thought about making any requests but did not).

Request 2: 11:55 am *[No new picture.]*

Translation: "(male announcement) We are approaching <unclear>. <sound of a machine howling> <machine howling, digital beeps several times, muffled female and male voices in the background>"

- *Why did you ask "what happened?" in this situation? Please describe in as much detail as possible.* This one was hilarious… It's a pilot project of a crosswalk device that is beneficial to the blind pedestrian. I was very curious how it sounded. It helped me understand what blind people hear, and improved my ability to do my job on [a disability organization].
- *Where were you?* [named a landmark].
- *Was someone trying to communicate with you directly?* No.
- *How did you feel about using the tool in the environment you were in when you asked "what happened?"* Everybody is walking around with these types of devices these days, so I was one of the crowd.
- *How useful would you rate the response (1=not useful → 5=very useful)?* 5
- *If you rated the response 4 or 5: what made it useful?* Well, it helped me understand what blind people hear, and improved my ability to do my job on [a disability organization].

Looking back at the whole day, answer the following:

- *Of all the times you requested "what happened" today, which request was the most important to you? Why?* Hmmmm… Difficult question. Well, for novelty value, the howling crosswalk signal. But for information it would be talking about [a different request made that day about a conversation with a co-worker].
- *Where there any times that you thought about using the tool but decided not to or weren't able to? If so, (1) briefly describe the situation, and (2) why did you decide not to use the tool?* Oh, it's such a neat little device that I am enthused to use it anywhere and anytime.

  Please share any comments, suggestions, or stories related to your use of the tool today: It was fun!

Fig. 3. A participant's email journal with one request and one full-day summary. Text in *italics* are questions from researchers. Substitutes in brackets are for clarity or anonymization.

5.3 Data Collected

Three participants completed the full 2-week field study. One participant completed 8 days due to conflicts scheduling the training session. A second and third participant completed 10 and 7 days respectively due to technical issues with their PDAs: one had GPRS connectivity problems for 4 of 14 days and the other had battery power

problems for 7 of 14 days. On these days, users were instructed to think about and report in the email journal when they would have used Scribe4Me that day and why.

For each request made during the study, we received two 15-second audio files and a photo taken by the user at the time of the request (43% of requests included photos; the rest were in similar situations or a photo was not practical). Occasionally, audio files from requests were not fully transmitted due to GPRS network connectivity issues, making it impossible to send a transcribed response. Momento did not support resending files lost in transmission because doing so would have increased response time to undesirable levels and compromised battery life. During the field study, both audio files from 16 (out of 118) requests were not transmitted, resulting in the response, "[Audio file not transmitted, possibly due to low connectivity. Please try again!]" One audio file from 36 additional requests was missing, indicated in the response by, "[1st/2nd audio file / 15 seconds missing]." Participants reported that even with only one of two audio files the responses were often useful. See Table 1 for a summary of each user's participation.

6 Results

Overall, participants were enthusiastic about Scribe4Me. All but one participant (#6 in Table 1) found at least one valuable use for the tool during the study. All participants said that they would continue to use Scribe4Me if they could install it on their mobile device. The major issue with Scribe4Me raised by all participants was the delay between requesting and receiving a transcription (which was 3-5 minutes). For 3 participants, this severely limited the scenarios in which Scribe4Me was useful, since they wanted a real-time aid for communicating with others. The other 3 participants found Scribe4Me to be valuable in a number of situations even with the delay: "It is a great idea. There are things that I'm curious about on a day to day basis. I don't mind waiting for the translation."

We next present results reflecting these user sentiments and showing the value of Scribe4Me to users with hearing loss. In particular, we describe differences in use between participants with different levels of hearing loss, situations in which Scribe4Me was useful, and users' thoughts on continued usage of Scribe4Me. Then we present results about those aspects of Scribe4Me that were successful and those that were less so in order to help improve similar tools in the future, including what information to include in transcriptions, technical limitations, user suggested improvements, and privacy issues.

6.1 When Was Scribe4Me Valuable?

Here we present the participants' experience with Scribe4Me during the 2-week field study. Feedback was largely positive and all participants wanted to continue using Scribe4Me in their daily lives. We include results about differences between users with different levels of hearing loss, especially what cues they used to decide when to use the tool. We then present situations in which the tool was valuable to users. Finally, we share participants' sentiments on continued use of Scribe4Me.

6.1.1 Equally Valuable for Users with Different Levels of Hearing Loss

Levels of hearing loss range roughly from profoundly deaf (no hearing) to moderately hard-of-hearing (hearing and comprehending most but not all sounds). Our study included participants from across this range and demonstrated the usefulness of the tool to people with diverse levels of hearing loss. There were no marked differences in how much participants with different levels of hearing loss valued using the tool or in how they used it. This was largely because all participants wanted to comprehend sounds better, which Scribe4Me helped them do. For example, one moderately hard-of-hearing participant who wore hearing aids said: "I don't think I have as good of comprehension [as the device]." Though cochlear implants and hearing aids amplify sound, they do not always improve comprehension, especially in noisy places.

The cues that prompted participants to use the tool differed between users with different levels of hearing loss. Users who are profoundly deaf often relied on a visual cue – either from the source of the sound itself or from others' reaction to it (*e.g.* "people turned their heads"). For the three hard-of-hearing participants, sounds that were audible but not comprehensible were often the cue for using the tool. Participants reported "hearing something" or missing parts of conversation and wanting a description of the sound or the missing words.

6.1.2 Valuable in Multiple Situations

Participants used the tool at home, work, restaurants, a drive-through, their own car, airports, business conferences, the grocery store, walking outside, riding public transportation, an animal shelter, church, group meetings, coffee shops, the post office, and other service-oriented businesses. All of the requests made are categorized in Table 2. Participants found the tool to be useful in a number of these situations, both for information awareness and for exploration:

Semi-public speech: Participants used Scribe4Me to learn about relevant conversations in which they were not obviously participating. "[The] meeting was over and various people were chatting. [The transcription was] useful after the fact, since this type of conversational situation is the most challenging for me. [I was] appreciative that [the tool] was available… to fill in the large gaping holes of conversation in a group I usually miss." Another participant said, "[I was in] the office. There was lots of stuff going on. Didn't know if there was a paging or just conversation. [The request was useful] because it seemed so loud." Another participant said, "It was during a break in the middle of an evening class and the professor started talking again just before the interpreter returned from her break. This is good for helping me to 'catch up' on the information prior to the interpreter's return."

Conversations in which the user was involved: Even though the delayed response made it difficult to use the tool in real-time conversation, one participant eloquently wrote about how the tool improved communication for her: "[The tool is most useful] conversation-wise. Usually I don't really get all that much of the conversation and subsist on a few words here and there. There were several times I used the tool and it may have taken a couple messages and a couple minutes but I was privy to the complete sentence, which was very nice. I think just having it as a tool… engages me into conversation which I would not normally have pursued. Even reading the responses after the fact and clarifying how the tool filled in what I did not lip read is encouraging. I very much enjoyed having it available and look forward to the day

when programs such as this will provide the opportunity to further engage the deaf and hearing impaired into interaction with others."

Announcements in public places: "[I was in the] airport. There are many announcements in airports. [It is] hard to tell when something important is being announced or something relevant to your flight."

Using machines with voices: "[I was] at [the grocery store], in the self-check out aisle. I really like using the tool because the stupid machine doesn't use captions on its screens, and it's nice knowing what automated messages it spits out." Another participant said, "[This is a] request to hear that darn fax machine. Believe it or not this thing talks and does not text its voice." Another participant was listening to the radio: "[We were] in the car [and the] game was on. Trying to see if I could pick up key moments (when crowd noise occurred, *etc*.). [The] noise showed something important was happening – [I] wanted to get the synopsis."

Table 2. All situations in which Scribe4Me was used, categorized by type and whether or not the request raises privacy concerns. Out of 118 requests, 4 were put in multiple categories and 11 were not categorized because the user did not describe the request in an email journal.

Use Category		#Requests
privacy concern		46
	public speech (others talking)	13
		captured unintentionally 8
		captured for curiosity, info gathering, or unknown intent 5
	semi-public speech	33
		class (teacher and student questions) 17
		general workplace chatter 13
		church 2
		non-work meeting 1
no / limited privacy concern		65
	conversation in which user was involved (hard to hide use of tool)	20
		with friends / other people 10
		with coworker or boss 5
		with family, at home 3
		strangers in stores / restaurants 2
	announcements	14
		airport announcements 9
		BART train / bus digital announcements 3
		announcements at restaurants / coffee shops 2
	machine / electronic / TV / radio voice	16
		TV 8
		grocery store self-checkout machine 3
		radio 3
		calling card with digital voice prompts 1
		fax with voice prompts 1
	environmental sounds (no voices)	15
		misc. environmental sounds 8
		traffic / city street environmental sounds 4
		environmental sounds at home 2
		electronics / computers 1

Environmental sounds: "I must have been trekking home, which would be walking along [the street]. Just trying to get a gist of my audible surroundings." Another user said, "I was wondering if you could hear the [train] sounds (if it makes them)… The train was just arriving and I was curious if they had made an announcement about it (speaker) and if you could hear it. Wanting to know if [the train] makes noise isn't exactly useful to me (in that specific situation). It just satisfies my curiosity."

These examples show that the tool has great potential for improving situational awareness and communication abilities for the deaf.

6.1.3 Participants Want to Continue Using Scribe4Me

All users said they valued the tool enough to continue using it and to pay for it as part of a mobile service plan. Two users thought the tool would be useful on a frequent, ongoing basis. Four users said they would likely use the tool infrequently, but that it was useful to have. One of these participants said: "If it had the 1 min response time and background sounds were no longer an issue, I'd probably use it a couple of times a week. Maybe even daily if the results continued to come back accurate."

6.2 What Did We Learn About Scribe4Me's Design?

Participants valued Scribe4Me and wanted to continue using it, indicating that a wider deployment is worthwhile. The results presented next provide design implications for future iterations of the tool. In the next five subsections, we present results about what worked and what didn't to inform the design of similar tools in the future.

6.2.1 Participants Want Translations of All Types of Information

Participants wanted descriptions of environmental sounds in addition to speech. Of 102 total transcriptions, 75% included speech and other sounds, and 25% included only environmental sounds. One participant said he appreciated the non-speech transcriptions, "…and that's why a voice-recognition system would be more limited. It was good to know that a sound came from the TV, *etc.*" Another participant wanted a description of the crosswalk for the blind (see Fig. 3). She said, "It is interesting to me to reveal not only conversation but also certain sounds, like the howling crosswalk noise. I remember pressing that button before and noticing that people around me noticably raised their eyebrows, and I guess I would too if I had heard it howl!"

Also, participants wanted descriptions voices. For example, they wanted to know the gender of speakers: "Knowing the gender of people who were talking made it easier to understand." They also appreciated knowing when speakers were making an announcement or coming from a TV or radio (all of which human transcribers could usually determine). Participants also wanted to know the *mood* of speakers. A participant asked that moods be added to transcriptions mid-study: "[It would be nice] if you could add the emotion of the people talking. Not for all of them, but if the person were mad or teasing, it is good to have that information." Transcribers began including notable speaker emotions, resulting in positive user feedback.

6.2.2 Technical Issues Limited Communication

Two main problems reduced the effectiveness of the tool for all participants in face-to-face communication situations: the delay of 3-5 minutes between when a user made a request and received a transcription; and transcription errors. These problems were caused largely by limitations of current technology.

The delay made the tool less useful to participants, especially in situations where the participant was in a conversation or listening to a speaker because "language happens quickly." One participant summed up the issue well: "Three to five minutes [delay] leaves one as an observer rather than an active participant." All participants thought the value of the tool would be much higher if the delay was reduced.

The second major issue was with inaccurate transcriptions caused by less-than-perfect audio quality and/or lack of context. Though most transcriptions were accurate and detailed, there were times when background noise overwhelmed speakers (*e.g.,* at restaurants), speakers were too far away for all words to be audible (*e.g.,* when overhearing conversations at work), and words were mistaken for others that sounded similar (*e.g.,* the word "terminations" was mistaken for "determination"). Lack of context made it hard to determine which sounds to transcribe and which to leave out. Due to these issues, participants felt the tool would not be useful in some critical situations, *e.g.,* "if there was a large group conversation."

6.2.3 Suggested Improvements: Less Delay, More Control

Based on issues encountered in the study, users suggested many improvements to the tool. Foremost was reducing the response delay. Participants also wanted the ability to choose the length of audio files and to send contextual information to transcribers (*e.g.,* which speaker to focus on among many). Participants thought long transcriptions sent in multiple messages were confusing, requesting that entire transcriptions be sent in one message. Finally, all participants wanted more accurate transcriptions and many suggested that the tool have an external microphone to better record audio. All of these suggestions are feasible to address in future iterations of the tool.

6.2.4 Privacy and Effects on Social Interactions

Because the Scribe4Me tool records the voices of people nearby, it introduces difficult privacy issues. Both privacy issues and the discomfort of using a device while interacting with someone, affected participants' use of the tool in social situations.

As discussed previously, the PAL memory aid system introduced similar privacy issues, recommending informed consent for recording peoples' voices and replaying them to third parties (*e.g.,* a transcriber). Arguably, a person who is deaf is not getting any information from Scribe4Me that a hearing person could not and so using the tool is socially justifiable. However, the transcriber certainly introduces privacy issues. We did not instruct participants to inform others of the tool, though most were very sensitive to the issue, informing conversation partners who were recorded or not making requests when they were unable to inform others. One participant said, "Everybody that I talked to, I showed it to them before I had to use it with them. I compare it to when you might ask someone if you might take their picture… In one sense, it makes others be more drawn into communicate, though there was still the hesitation of someone else hearing in on the conversation." Another said, "[When others saw me

using the tool I felt the need to tell them about it] because otherwise it would feel like spying – even though I was getting information that would be readily available to a hearing person." Use of the Scribe4Me tool was somewhat discrete, but hard to hide in face-to-face conversation since users had to push a button and read the screen when a transcription arrived. This visible use may have encouraged participants to inform others, socially enforcing informed consent in conversational situations.

Table 2 categorizes all requests by whether or not they introduce privacy concerns. Consent is naturally enforced in face-to-face conversations and announcements are public, but other instances of speech are a privacy issue. To protect privacy, we need ways to prevent capture or enforce consent. By adding technology that detects speech client-side [21], we can put users in control of privacy decisions. For the deaf, we cannot indicate if the speech is an intended conversation partner or a nearby conversation. To be legally sound, the user would not make a request if the tool informs them speech is present and others are nearby. For other applications (*e.g.,* language translation) the tool could replay the audio so the user can tell if unintended speech was captured. Another possible solution is to change policy, enabling users who are deaf to use transcriptions tools and certifying transcribers.

Users also hesitated to use the device in some situations to avoid being rude. The most commonly avoided situation was interacting with others who did not know about the tool: "it is rude to check a phone when in the middle of a conversation." One hard-of-hearing participant felt uncomfortable using the tool in church, even though she thought it would be useful for catching parts of the sermon she missed: "In Church, we've been harped on about not using cell phones, people looked at me kind of sideways and I wanted to say, 'this is not a cell phone! I'm not talking to anybody!'"

Despite these awkward situations, people felt comfortable using the tool around people they could explain it to and in less participatory situations: "Everybody is walking around with these types of devices these days, so I was one of the crowd." One participant thought that the value added was worth some awkwardness: "I feel kind of awkward using the device in a public setting… With that said, if the device is truly helpful, it doesn't matter what the people think. So, if I were in a position to need your device, I would use it, regardless of what people around me think!"

6.2.5 Applying Lessons Learned to Other Ubicomp Applications

The Scribe4Me field study, though limited in size, may provide some guidance for similar applications that involve mobile sound recording and transcribing. In particular, the field study informs the sound information needed, situations of use, social implications, and the need for feasible human-in-the-loop systems.

People found the following information valuable in transcriptions: speech, descriptions of speaker voices (*e.g.,* gender, emotion), and environmental sounds. Other applications that transcribe sound will want to consider these pieces of information. An example application similar to Scribe4Me is a tool for travelers that translates speech into the user's language. Even though a user is primarily interested in speech, descriptions of speaker's voices and environmental sounds would be useful for placing the voices in context. Some situations in which users found Scribe4Me valuable may be useful for travelers: in conversations, using machines with voices,

and announcements. Travelers may have addition needs, such as reading signs or notices that could be supported by translating text in photographs taken with the tool.

We learned about limitations of Scribe4Me that would likely be issues for similar applications. It is difficult to use in face-to-face conversations due to social expectations of eye contact and attention. Also, the most difficult situations for accurate transcriptions are loud environments (*e.g.,* restaurants) and situations with multiple conversations (*e.g.,* in a classroom). Difficulties were caused by background sounds drowning out more interesting sounds, not knowing which sounds were of interest to the user, and parsing sentences and conversations when multiple occurred at once.

Finally, results from our study provide motivation for exploring lower cost solutions to including humans in the system. While human are often included in telecommunication and video relay services for the deaf [6], these services are costly. Some projects have explored distributing work among many people on the Web [18, 24]. Scribe4Me or similar applications, could be powered by a huge pool of Internet users who share the transcribing load.

6.3 Summary

Results of the 2-week field study with 6 users showed that Scribe4Me had valued uses for both the deaf and hard-of-hearing. Though two weeks is not long enough to eliminate novelty effects, data gathered about when Scribe4Me is useful and how we can improve it are valuable. Participants acknowledged the affects of novelty on their usage but still considered the tool useful: "I could see a tool like this being a useful ongoing thing in my life… but I wouldn't use it that often as I did in the last couple of weeks in my daily life." Participants found it useful to "hear" people speaking, machines with voice prompts, announcements in public places, untranslated parts of lectures, the radio, and environmental sounds. The tool was often not useful in face-to-face communication due to the response delay and limited audio quality. Participants suggested several other improvements: that the length of captured audio could be chosen, an external microphone for better audio quality, a way to indicate which thing or person to focus on in transcriptions, and long transcription text in one message.

Recording others' voices introduced privacy issues to which participants were sensitive. Despite social issues, participants used the tool throughout the study and all predicted they would continue using it and would pay for it as part of a service plan.

7 Future Work and Conclusions

We presented the design and evaluation of Scribe4Me, a mobile sound transcription tool for the deaf and hard-of-hearing to improve awareness of sound information in any location. Results of a 2-week field study with 6 users showed many valuable uses. Technical hurdles for future iterations are to reduce the response delay and to improve the recorded audio quality. With these issues solved, users expected that it would be an even more powerful tool for communication and environmental awareness.

While it is possible to use humans for transcription in a full system, automated translation of non-speech audio may be an option in the future. We plan to address

response time issues by adding support for a live (phone) connection in interactive settings, and by leveraging faster 4G networks or open WiFi networks when possible. Finally, future versions will encourage informed consent when audio recording other people, perhaps with a prompt to remind users.

Lastly, we are interested in using what we learned to explore new applications with remote transcription. For example, a similar tool could be used by travelers to translate different languages in settings where existing automated translation tools fail.

References

1. Akyol, S. and Alvarado, P.: Finding Relevant Image Content for Mobile Sign Language Recognition. In: Proc. of SPPRA. (2001) 48-52.
2. Auer, E.T., Bernstein, L.E. and Coulter, D.C.: Temporal and Spatio-Temporal Vibrotactile Displays for Voice Fundamental Frequency: An Initial Evaluation of a New Vibrotactile Speech Perception Aid with Normal-Hearing and Hearing-Impaired Individuals. J. of the Acoustical Society of America, 104, 4, (1998) 2477-2489.
3. Black, A.W., Brown, R.D., Frederking, R., K. Lenzo, J., Moody, A., Rudnicky, Singh, R. and Steinbrecher, E.: Rapid Development of Speech-to-Speech Translation Systems. In: Proc. of ICSLP. (2002) 1709-1712.
4. California Foundation for Independent Living Centers (2003), NorthView Wins Exclusive Rights to Market SpeechView's Solutions for Deaf. Assistive Technology Journal, 65, www.atnet.org/news/2003/jan03/011504.htm.
5. Carter, S. and Mankoff, J.: When Participants Do the Capturing: The Role of Media in Diary Studies. In: Proc. of CHI. (2005) 899 - 908.
6. Federal Communications Commission (2006), What You Need To Know About TRS. www.fcc.gov/cgb/dro/trs.html.
7. Cook, A.M. and Hussey, S.M.: Assistive Technologies: Principles and Practice. Mosby, Inc., St. Louis, (2002).
8. Deaf Studies Trust (2003), WISDOM: Wireless Information Services for Deaf people On the Move. www.deafstudiestrust.org.uk/research_projects/wisdom.htm.
9. Doyle, M. and Dye, L. Mainstreaming the Student Who is Deaf of Hard-of-Hearing, Alexander Graham Bell Association for the Deaf and Hard-of-Hearing, 2002.
10. Edwards, A.D.N.: Progress in Sign Language Recognition. In: Gesture and Sign-Language in Human-Computer Interaction. (1997) 13-21.
11. Elssmann, S.F. and Maki, J.E.: Speech Spectrographic Display: Use of Visual Feedback by Hearing-Impaired Adults During Independent Articulation Practice. American Annals of the Deaf, 132, 4, (1987) 276–279.
12. Hudson, S. and Mankoff, J.: Rapid Construction of Functioning Physical Interfaces from Cardboard, Thumbtacks and Masking Tape. In: Proc. of UIST. (2006) To Appear.
13. Huenerfauth, M. Survey and Critique of ASL Natural Language Generation and Machine Translation Systems Technical Report MS-CIS-03-32, University of Pennsylvania, 2003.
14. Iachello, G., Truong, K., Abowd, G., Hayes, G. and Stevens, M.: Event-Contingent Experience Sampling To Evaluate Ubicomp Technology In The Real World. In: Proc. of CHI. (2006) 1009-1018.
15. Liu, F.H., Gu, L., Gao, Y. and Picheny, M.: Use of Statistical N-Gram Models in Natural Language Generation for Machine Translation. In: Proc. of ICASSP. (2003) 636-639.
16. Mann, W.C. and Lane, J.P.: Assistive Technology for Persons with Disabilities. The American Occupational Therapy Association, Inc., Bethesda, MD, (1995).

17. Matthews, T., Fong, J., Ho-Ching, F.W. and Mankoff, J.: Evaluating Non-Speech Sound Visualizations for the Deaf. Behaviour & Information Technology, (2006) In Press.
18. Mycroft (2006). harbinger.sims.berkeley.edu/dmc/public/.
19. Pastor, M., Sanchis, A., Casacuberta, F. and Vidal, E.: EuTrans: a Speech-to-Speech Translator Prototype. In: Proc. of Eurospeech. (2001) 2385-2389.
20. Power, M.R. and Power, D.: Everyone here speaks TXT: Deaf People Using SMS in Australia and the Rest of the World. J. of Deaf Studies & Deaf Education, 9, 3, (2004) 350-360.
21. Schmandt, C., Lee, K., Kim, J. and Ackerman, M.: Impromptu: Managing Networked Audio Applications for Mobile Users. In: Proc. of MobiSys. (2004) 59 - 69.
22. Sonido Incorporated (2003), Auditory Visual Articulation Therapy Software. www.sonidoinc.com.
23. The Reporters Committee for Freedom of the Press (2003), The First Amendment Handbook. www.rcfp.org/handbook/c03p01.html.
24. von Ahn, L., Liu, R. and Blum, M.: Peekaboom: A Game for Locating Objects in Images. In: Proc. of CHI. (2006) 55-64.
25. Woszczyna, M., Coccaro, N., Eisele, A., Lavie, A., McNair, A., Polzin, T., Rogina, I., Rose, C.P., Sloboda, T., Tomita, M., Tsutsumi, J., Aoki-Waibel, N., Waibel, A. and Ward, W.: Recent Advances in JANUS: A Speech Translation System. In: Proc. of Eurospeech. (1993) 1295-1298.