Handbook of Research on User Interface Design and Evaluation for Mobile Technology

Volume II

Joanna Lumsden
National Research Council of Canada
Institute for Information Technology – e-Business, Canada

Information Science REFERENCE

INFORMATION SCIENCE REFERENCE

Hershey · New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Sara Reed
Copy Editor: Joy Langel, Katie Smalley, and Angela Thor
Typesetter: Jeff Ash
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on user interface design and evaluation for mobile technology / Joanna Lumsden, editor.

p. cm.

Summary: "This book provides students, researchers, educators, and practitioners with a compendium of research on the key issues surrounding the design and evaluation of mobile user interfaces, such as the physical environment and social context in which a device is being used and the impact of multitasking behavior typically exhibited by mobile-device users"--Provided by publisher.

Includes bibliographical references and index.

ISBN 978-1-59904-871-0 (hardcover) -- ISBN 978-1-59904-872-7 (ebook)

1. Mobile computing--Handbooks, manuals, etc. 2. Human-computer interaction--Handbooks, manuals, etc. 3. User interfaces (Computer systems)--Handbooks, manuals, etc. I. Lumsden, Joanna.

QA76.59.H36 2008

004.165--dc22

2007024493

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter LXI
Designing a Ubiquitous Audio-Based Memory Aid

Shwetak N. Patel
Georgia Institute of Technology, USA

Khai N. Truong
University of Toronto, Canada

Gillian R. Hayes
Georgia Institute of Technology, USA

Giovanni Iachello
Georgia Institute of Technology, USA

Julie A. Kientz
Georgia Institute of Technology, USA

Gregory D. Abowd
Georgia Institute of Technology, USA

ABSTRACT

The ephemeral nature of sound can be problematic when people cannot recall something they heard. Motivated by everyday conversational breakdowns, a continuous, near-term audio buffering application named the Personal Audio Loop was designed to recover audio content from the recent past using the mobile phone platform. The investigation of the potential usefulness in everyday life, the level of ubiquity and usability required of the service, and the social and legal considerations for long-term adoption is presented. The methods used include a controlled laboratory study, in-situ diary and event-contingent experience sampling studies, examination of legislation, and deployment of the technology over several weeks.

INTRODUCTION

Everyday conversations fill our lives and we are all very familiar with the kinds of breakdowns suggested by these simple scenarios:

- You are in a conversation with a friend and one of you is interrupted. When the conversation resumes, neither of you remembers what you were talking about.
- You are at a social event and you are introduced to someone new. Minutes later, you have forgotten the person's name.

Current mobile phone technologies have the potential to preserve audio to enable people to recover needed information in many similar situations. Thus, the Personal Audio Loop (PAL) is the product of a multi-year design and research project focused on developing such capabilities. PAL is an automatic, audio-based memory aid that relies on audio capture technologies and the remarkable human ability to recover information quickly with only a few bursts of the audio from that memory.

PAL represents a significant case study for the design of new mobile and ubiquitous technologies. PAL is exceptional in its design and in its use because it required emerging technologies and unique usage models. Additionally, it is a personal technology that makes use of public information: the audio in the environment within earshot of an individual. In this chapter, the innovations required for this design case are explored. A mixed technological and human-centered approach was necessary to produce a near-term audio service that could survive research in the natural environment of people's everyday lives.

The design process addressed multiple questions of technical and human significance:

- **Usefulness:** Though motivated by observations from everyday life, how often and in what situations do people actually need a near-term audio memory aid?
- **Ubiquity:** What parameters of this service would make it available every place and every time it is needed?
- **Usability:** How should the service deliver functionality to maximize its benefit and minimize its distraction?
- **Social and legal considerations:** What aspects of society may influence the uses and cultural practices surrounding an audio recording application for everyday life?

PAL and technologies like it are becoming increasingly common, but at the time of its inception, the design was arguably outside the realm of a typical person's experience. Therefore, two essential requirements of the design process were to allow potential users to interact with a working prototype to have a sense of the capabilities and to answer engineering questions, including important architectural considerations.

From a technical perspective, there are several options for designing an audio-based memory aid to provide the required capabilities. All of the designs considered during this process reflected the same basic notion of replaying a buffer of recently recorded audio, but early prototypes varied in terms of the nature of recording and playback capabilities. A fully distributed system assumes an environment equipped with microphones, speakers, and interface controls to maximize opportunities for recording and playback wherever and whenever needed. A fully localized solution provides recording and playback in an all-in-one package carried wherever needed. A hybrid solution might delegate the recording in the environment and use a handheld device that receives streamed audio from a central repository for playback.

The complete design process for the Personal Audio Loop, a solution for an audio-based near term memory aid that addresses the technical concerns of an interesting capture and access application and answers questions from the four categories described, is presented. The process involved a series of formative studies that led to the design of a self-contained service integrated into a commercial mobile phone handset, summative studies that explored real life uses for primary users, and situated imagined uses that considered secondary stakeholders.

BACKGROUND AND RELATED WORK

Similar Audio-Based Memory Applications and Technologies

Near-term capture and access applications that provide audio reminder services have been previously explored in the office as well as for telephone conversations. Xcapture, originally built to provide a "digital tape loop" of a single office, could also provide short-term auditory memory of telephone conversations (5 to 15 minutes long) (Hindus & Schmandt, 1992). Although the system was designed for use in a setting where social protocol allows recording, the authors recognized the privacy issues of subsequent use of archived recordings and suggested that social expectations change with use. In MERL's real-time audio buffering technique, captured audio persists for the duration of that phone conversation (Deitz & Yerazunis, 2001). During the course of the conversation, a user may tap the phone against the ear to move backwards in the audio and to replay any portion of the discussion. This system does not store conversations and could arguably pass legislative tests and be socially acceptable. On the other hand, a reminder application can be created to preserve the audio for long-term use. In their investigation of an audio-based personal memory aid, Vemuri, Schmandt, Bender, Tellex, & Lassey (2004) created a mobile application for the iPaq PDA that constantly records, analyzes, and indexes audio to identify the best memory triggers for events from daily life. The application stores short audio clips, which it has determined to be good memory triggers for an event based on an analysis of the audio for key features such as the speaker and the tone of the discussion. Additionally, the system performs speech recognition to allow the user to perform keyword searches over the captured content at a much later time.

Relevant Legal Cases and Policies

Legal cases over the past two decades have exposed the contrasting requirements and balances of privacy and utility for recording applications. The experience in the fields of surveillance in public spaces and of the privacy of private communications is drawn from. Among other sources, European Directive 95/46/EC opinions and rulings by various EU Data Protection Authorities (DPAs) (British Institute of Comparative Law, 2003; European Commission, 2004), and several U.S. Supreme Court[1] rulings were considered.

Despite the ongoing debate stressing the differences between the United States and Europe regarding privacy, legislation regulating the recording of communications by electronic means is remarkably similar. The main items are the U.S. Electronic Communications Privacy Act (ECPA) of 1986 and European Directives 2002/58/EC and 95/46/EC. ECPA regulates wiretap and surveillance and applies to any electronic recording device and conversations ("oral communication") between two persons "exhibiting an expectation that such communication is not subject to interception," even if the conversations were not transmitted through a telecommunications network. European Directive 2002/58/EC covers only personal conversations transmitted over public telecommunication networks. However, Directive 95/46/EC applies to any personally identifiable information which includes recorded voice conversations, according to multiple opinions by European national data protection authorities. Directive 95/46/EC requires a proportionality assessment between potential harm and benefits; however, the personal character of the application might exempt users from many provisions, including informed consent.

FORMATIVE EVALUATIONS OF PAL

Based on early interviews with potential users and intuition, it was determined that the platform for PAL would need to be mobile, computationally powerful, include a robust development environment, and feature easily accessible interaction techniques and an external or attachable microphone. The mobility, ubiquity, and performance of mobile phones make them an appealing platform for this application but only certain phones

support the required capabilities. The chosen Motorola iDEN i730 (Figure 1) clamshell phone features the J2ME programming environment, an external microphone, and buttons accessible while the device is closed and worn while clipped to a belt on in a pocket. The microphone is capable of recording voices in a small room with the phone open or closed in a shirt pocket or attached to a belt, with higher quality than most PDAs. Questions about the feasibility of the mobile phone as a platform as well as about the nature of situations that would require use of PAL remained. Thus, two formative studies were designed to answer these questions: a laboratory study designed to answer the questions of usability of the interface and a diary study to characterize the frequency and situations of use in everyday life, the full details of which are available elsewhere (Hayes et al., 2004).

Laboratory Study: Developing a Usable Mobile Interface

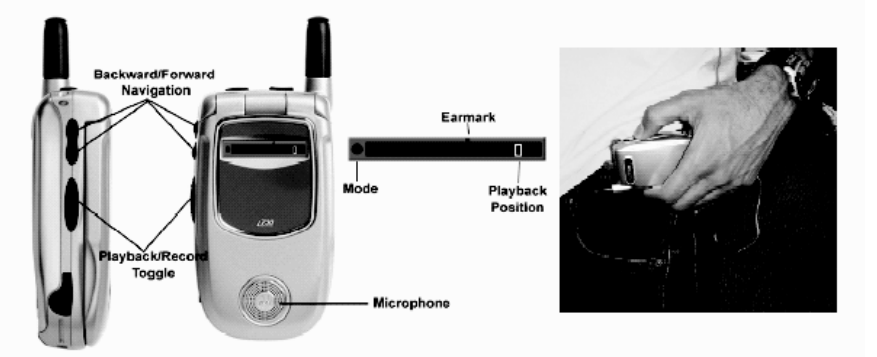
Mobile phone applications can be difficult to use. The limited display size and paucity of buttons or other interaction methods mean that creative designs are necessary to create feature-rich applications. PAL required several features including the need for quick, unobtrusive access and interaction techniques that enable a human to find relevant moments in a stream of audio. If PAL was too difficult or awkward to use or took too long to locate important audio, it would likely not be less useful for a person.

The Initial Prototype

The laboratory study necessitated a working prototype capable of testing the phone as an interface to the recent audio recorded in a person's surrounding. In its normal operating mode, the implementation of PAL continuously records audio from the user's environment. Audio older than the buffer length (in the initial prototype, 15 minutes) is deleted automatically. Recording automatically halts when the user answers or makes a call. Five buttons are available on the outside to accommodate interactions while the phone is closed (Figure 1). PAL provides simple audio navigation features (e.g., rewind), informed both by previous research on skimming (Arons, 1993) as well as by commercial video recording services like Tivo™. PAL includes a simple timeline visualization on the exterior LCD of the handset indicating application status (recording, playback, and direction of navigation) as well as the playback position in the audio buffer relative to the current time (the right edge of the timeline).

Technological limitations and engineering problems remaining to be solved meant that a fully functional prototype would take several months. In addition, developing applications for

Figure 1. Left: The Motorola i730 handset used for PAL. Three buttons control navigation and record/playback mode. A timeline on the face of the phone indicates mode and relative place in the buffer. Right: Example placement of phone when using PAL.



a mobile phone is a lengthier process compared to traditional desktops. However, to test the usability of the design before exerting such a programming effort was wanted. As a result, a "tethered" version of PAL specifically for this controlled study (see Figure 2) was built. For the tethered version, a Motorola i730 mobile handset was connected to a laptop computer through a serial connection. The laptop handled the audio recording and playback using the Java Media Framework but the visualization and interaction with the application was entirely through the i730 handset. The button events were sent to the laptop through the serial link and the visualization was updated on the handset as events were sent back from the PAL audio engine running on the laptop. This architecture provided a quick and easy prototyping platform for the design. It also provided a way to easily log the user events as the user interacted with the PAL application. Without having to worry about the technical limitation of the mobile handset, a prototype which was indistinguishable from the actual wearable version was able to be quickly created. This early prototype allowed the gathering of quantitative performance metrics as well as user reactions to the device.

Method

The laboratory study included 18 participants: students and faculty from the Georgia Institute of Technology. Participants ranged in age at the time of participation from 18 to 50 and included five females. Participants had a background in Human-Computer Interaction and usability, and were chosen explicitly with the intent of examining heuristics such as the mapping of buttons to functionality and the quality of the visualization. Participants' experience with mobile phones ranged from 7 years of consistent use to no experience at all (7 participants). The prototype was demonstrated, encouraging participants to examine the device and ask questions until they expressed comfort with its functions.

Figure 2. Initial prototype of PAL tethered to a laptop used in a controlled laboratory study

PAL's intended use involves the replay of audio for which the user was present initially. The controlled study, designed to mimic this scenario, included a scripted dialog of 5 minutes. In this script, the participants asked researchers predetermined questions and researchers replied with the same answer for every participant. The script purposely involved a large amount of detail to increase the likelihood that participants could not recall the answers to all questions from memory. After completing the dialog, the researchers who had been participating in the dialog removed the script and asked the participants a series of questions about the information they just heard. Although the researchers noted whether the participants remembered the information without use of PAL, they asked every participant to find and play every answer. Participants were encouraged to "think aloud" as they used the prototype and the researchers timed how long it took an individual to find the answer, theorizing that this first time use while discussing their actions would be a worst case timing for most users. Participants answered seven questions, the first two being practice questions not used for computing timing results. An exit survey and semi-structured interview provided further evaluation of the interface by users and of their envisioned need for this kind of service.

Results

After a short demonstration, all participants were able to navigate the audio well enough to answer the questions. The exit survey had users answer

questions and rate various features on a Likert scale from 1 to 7, ranging from 1 being "Strongly Disagree" and 7 being "Strongly Agree". According to the exit survey, participants agreed that the device was easy to use with one hand and small enough to carry at all times. They could clearly understand the audio even in its highly compressed form.

With an audio buffer of 15 minutes, participants required an average of 34.8 seconds to find responses for questions that were known to be in the in the recorded audio while talking aloud about their actions. Participants reported the visualization was somewhat helpful in accomplishing the task but not overwhelmingly so. Thirteen of eighteen participants actually used PAL without the visualization, preferring an eyes-free interaction.

Although inquiring about privacy was not a goal of the controlled usability study, 10 participants raised spontaneous concerns regarding the social acceptability of a continuously recording system. The most common sentiment expressed indicated that participants were less concerned about recording their own voice than those of their conversation partners.

Diary Study: Determining the Usefulness of PAL

The laboratory study showed the feasibility and usability of PAL on a mobile phone but it did not inform about the overall usefulness in everyday life. A diary study was undertaken to explore the extent to which a near-term audio reminder service was needed, looking for frequency and characteristics of potential use. Diary studies balance the ecological validity of gathering such data in situ against interruption of everyday activity flow caused by recording personal observations, particularly in mobile settings (Czerwinski, Horvitz, & Wilhite, 2004). Specific information relating to social context including privacy concerns in the diary entries and during the follow-up interviews was asked for.

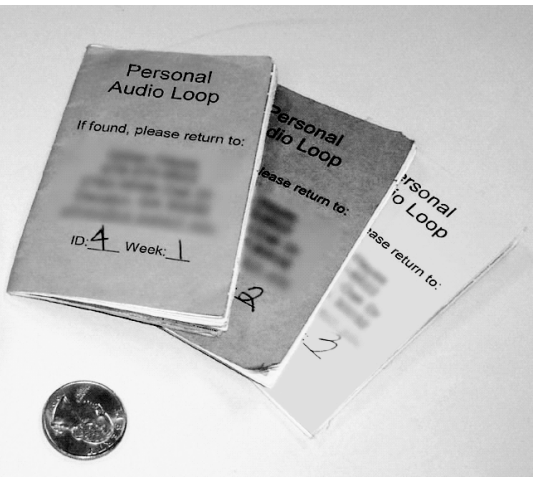
At the time of the diary study, a working version of PAL on a mobile phone was available. However, it was not ready for full deployment because of performance and reliability problems. The prototype was more than adequate for the purposes of demonstrating the system to participants. The diary study allowed the development of PAL to continue while still providing important development feedback.

Method

Twelve experienced mobile phone users (5 female, 7 male, ranging in age from 22 to 60 years) participated in the study. Participants' occupations spanned a spectrum of domains, including a psychologist, a finance manager, a realtor, a car dealer, a consultant, a professor, and a full-time homemaker. A working version of PAL was demonstrated to the participants. Then they were asked to carry small pocket-sized diaries and record an entry for each incident during the following week when they would have needed or liked to use the PAL service. Each page of the diary contained a simple form to complete for the potential instance of use, streamlined after an initial trial period. Each form in the diary included space for describing the content of the audio to retrieve, when and where the incident occurred, and whether any persons unrelated to the conversation were nearby. Participants also estimated how far in the past the salient audio content was and rated how important it was to retrieve that information. Figure 3 shows an example of an incident survey.

At the end of each week the diaries were collected from the participants and semi-structured interviews were conducted to examine a subset of entries in detail, up to six diary entries per participant per week. The detailed questions probed issues such as the kinds of information being sought, the distance of unrelated third parties from the participants, and the participants' assessments of the social appropriateness of using the device in the specific contexts. Each participant who chose to continue for another week was each then given a new empty diary to again record incidents. Participants could continue with the diary for up to 3 weeks. At the end of the study, semi-structured interviews with all participants were conducted.

Figure 3. Left: Sample diary entry; Right: Examples of pocket size diaries used in the study



The weekly and summary interviews allowed the clarification of misunderstandings in the entries as well as probing particular issues that were more difficult to garner from the diary form factor.

Results

Twelve people participated in the first week, eleven continued for the second, and eight in the third, for a total of 31 participant weeks and 109 incident reports. On average, participants reported more than three incidents per week, most of which referred to data that was less than an hour old with only 6% from over a day prior. Of the incidents reported, most occurred in public or semi-public places (defined as schools, workplaces, etc.) with less than a third in private space (predominantly car and home). In almost half of the incidents, people unrelated to the audio they wished to retrieve (e.g., other customers in a restaurant) were present. During follow up interviews, participants asserted that they would not have felt rude towards their communication partner using PAL in most of the situations. During the second and third weeks, participants were questioned about their reactions if their partners objected. Participants stated that such an objection would be "not likely" in almost every case and indicated that they would not have complied with the objection, had there been one, in three fourths of the questions. When asked how far away they would like PAL to record, two-thirds chose within a small room (10 feet) and only one individual requested a large radius, reporting that he is "just nosy." Generally, participants were willing to take the few minutes that would likely be required to access the data with willingness to spend time finding information roughly correlating to the importance of the information.

SUMMATIVE EVALUATIONS OF PAL

Deployment Study: In Situ Evaluation of Real Life PAL Use

The formative studies of PAL indicated how people might use PAL but it was difficult to know if these predictions were true without testing the service in the "real world." Thus, a robust version of PAL was developed on a mobile phone, which was informed by the results of the laboratory and diary studies. Having a fully functional system on the common form factor of a mobile phone allowed the exploring of how people would use such a system were it to be widely available.

The Final Prototype

The PAL application runs continuously on the i730 handset in one of three modes. The default mode is recording but that mode can be switched to either pause or playback. Due to the phone's performance limitations, there is some latency in the overall interaction including an 800-1000 ms delay for playback to actually begin after seeking. The laboratory study indicated this delay was still within a usable range. The phone that was used has enough memory to support up to 3 hours of audio; considering silence periods, the phone could provide recall functions for considerably longer. For the initial deployment, an arbitrary limit of 15 minutes for the recorded audio buffer was chosen although the application allows users to set their own buffer length. The i730 features a high quality built-in microphone and an external speaker. The external microphone is sensitive enough to pick up voices in a meeting room with the phone closed and in a shirt pocket. This range was sufficient based on the situations found in the diary study. The PAL application runs over 12 hours on a fully charged 1400 mAh lithium-ion battery with it continually recording and an average of 15 seconds of playback every hour.

The PAL application runs with the phone shell "flip" closed and the user interacts with the three buttons on the left side of the handset. PAL can be operated with the handset closed for complete heads-up operation, especially when worn on belt clip with a hands free ear bud and microphone; in this configuration, the external display provides status information. When the user opens the phone flip either to answer a call or dial a call, the recording pauses. In this mode it is also possible to browse the recorded audio and the internal display provides information to facilitate skimming. The users can skim backward, forward, and toggle between record, pause, and playback modes. In playback mode, quick and simple navigation through the audio stream is important. PAL is controlled using the available side buttons of the handset and a graphical display on the external panel for quick visual feedback. When a user needs to re-listen to part of a conversation, the small buttons are used to skip backward or forward a fixed amount of time in the recorded audio stream. This simple forward and backward jumping allows a user to skim a conversation to find a portion that reminds them of some salient feature of a previous conversation. The prototype defaults to a skip backward of 30 seconds and a skip forward of 10 seconds based on the controlled study. Significantly different forward and backward jump times were purposely chosen because similar forward and backward times would cause too little net change in skim mode when alternately using forward and backward.

It was found that different people have a different conception of which of the two navigation buttons is forward and which is back. As the orientation of the phone changes, these mappings also change. To alleviate this problem the notion of smart buttons, where the user is never wrong, was introduced. Each time the user initiates skim or playback of the recording, the only logical direction is backwards. The application makes whichever button the user presses first, the back button. The opposite button becomes forward. After the user goes back to record mode, the mapping resets. The next time the user initiates playback mode, again, the first button selected becomes the back button.

Operating on the metaphor of a running timeline, the display shows the full 15 minute audio buffer as a horizontal bar. There are two moving parts. The first is the buffer (dark region of the horizontal bar) that moves during recording to show that audio that has been recorded. After 15 minutes or the specific buffer length, the entire bar is filled in, which will be the case most of the time. The second movable part is a bead that shows the playback position relative to the recording position. By looking at the display, the user can position the playback point relative to the current recording point. In addition, users can use the visualization to roughly locate recordings.

To aid human's inaccurate recollections (e.g., when something occurred 5 minutes ago compared to 10 minutes ago), additional context information on the visualization is provided. One feature shows the silent time resulting from

pauses during recording on the visualization. The recording stops when the user goes into playback mode, when a call is placed or received, or when the user explicitly pauses the recording. Since the timeline is a temporal representation, a dead space placeholder is provided (indicated by white space) to indicate where the mentioned situations occur relative to the current audio buffer. The white space is skipped when skimming. However, there is an important tradeoff to note with this scheme. Consider someone who talks for over 15 minutes on the phone or re-listens for a significant amount of time; most of the buffer would just be filled with white space. Therefore, an option for PAL to record either the last 15 absolute minutes or the last 15 actual audio minutes is provided. Finally, a marking feature where users tag important points in the audio loop is provided. This could be the result of something heard during record or playback mode. For example, if the user receives directions, they may place a bookmark at the time of recording because they know they will need to go back and access it later. To bookmark, the user simply taps the large button to indicate a tag point and a vertical tick mark appears on the visualization. The actual mark is placed 5 seconds before a button press event. In playback mode, holding the navigation button jumps between markers.

Method

A working version of the application was deployed to four of the diary study participants for 7 weeks. Furthermore, four members of the research team also used PAL for over 2 months, one as long as 4 months. During the first four days of the deployment, participants were asked to carry a diary to note their uses of the device. Throughout the deployment of the technology, their uses were also logged. By deploying the devices to even a small number of users, both expected and emergent uses could be observed and greater understanding about the dependency users might develop on the service could be gained.

Results

On average, participants used the device more than twice a week. Although this average is lower than what was indicated by the diary study, participants also reported an average of 1.5 incidents that they thought about using the device and chose not to use it. In one case, the user's conversation partner recovered the information before the user was able to try with PAL. In all other cases, the reason not to use PAL was reported as forgetting it was available. Interviews with the users since this initial probe indicate that ordinary use subsequently remained fairly consistent with the rate observed in the first 4 days and that the frequency of use for exploring the application or showing it to others has decreased substantially. Overall, satisfaction as reported through qualitative interviews has been high. All four users requested to continue using the devices after the first 4 days and reported that they believed they would use them more, over time. Each user changed the buffer length (ranging from 10 minutes to 60), the initial jump backward (ranging from 15 seconds to 60), or both. Users expressed that configuring the application was important and one user even indicated that he changes the buffer length depending on the situation he is about to encounter.

In the initial 4 day probe, the most frequently reported situation for use was to remember forgotten details (60%). Other unexpected situations have also been reported in the following weeks. Specifically, users have been employing PAL as an instructional aid. For example, one user recorded conversations with customers and then replayed them for an employee in training. Another recorded a negotiation with an employee to replay it for another, again as part of job training. One user regularly used the PAL service to augment her medical journal. Her physician had asked her to record data about symptoms. Her job required a great deal of driving, however, which made it difficult for her to record these symptoms by hand. Instead, she chose to speak them aloud when she could not write them down at that mo-

ment. She would then pause the recording long enough to replay it, when she could write in her health journal.

All of the users expected the service to be available, reporting that they would choose not to write information down when it is already spoken aloud. Ultimately, this expectation of availability became a significant hurdle to long-term adoption. No user continued to use PAL for more than 6 months. The primary reasons for discontinuing use were: (1) difficulties encountered when the operation of the device as a phone and the device as PAL interfered with one another and (2) annoyance with carrying two phones in the case of those users who were still carrying their personal mobile phones as well as the PAL-enabled phone.

Social contract issues recurred more often in the case of the real deployment than the results of the diary study had revealed. Users expressed that conversation partners aware of the device sometimes responded negatively initially, but relaxed after the application and its buffering and discarding functions were explained. Interestingly, all four users reported informing new conversation partners about PAL less frequently as time went by. After several weeks, users had almost stopped alerting conversation partners altogether. They would use their own judgment of the situation to decide if it was acceptable to record. They did, in fact, sometimes choose not to record despite reporting wanting to at those times. In both situations in which they ultimately chose to record and those in which they turned PAL off, participants reported deciding not to alert people to the possibility of the recording. As frequently as users reported negative social repercussions from PAL, they also reported positive cooperative uses of the device. For example, one user's wife consistently used PAL on his device by walking near him and speaking when she needed to remember something. Despite the in-depth usage information that was able to be obtained from this deployment of the service, the choice on the part of the users to avoid discussing PAL when they thought it might be confrontational left no in-depth understanding of the experience of the conversation partners and others nearby.

Contextualized Survey Study: The Experience of the Conversation Partners

Motivated by the results of the diary, laboratory, and deployment studies as well as available legislation and instincts about societal norms, it was chosen that the problem of PAL be explored further, from the standpoint of the conversation partners. As uncovered in the first three studies, primary stakeholders sometimes may have legitimate interests in using PAL, for example, due to a memory dysfunction or simply because of cognitive load. This interest may be opposed to that of secondary stakeholders or third parties (who might not want to be recorded, even if only temporarily, or who might want further explanation of the service when the user may find taking time for this explanation to be inconvenient or even impossible in the particular situation). Thus, there is a significant question of whose interests should prevail. A second question considered use of PAL relating to the proportion of individuals opposed to the application. If only a small minority of secondary stakeholders and third parties oppose PAL and the vast majority does not care, should the contrary minority be yielded to and a large market potential be curtailed?

These issues echo the classical ethical debate between utilitarianism and normativism in the context of privacy and technology (Terrell & Jacobs, 2002). It was chosen to explore these issues using the concept of proportionality (Iachello, 2006) borrowed from the legal community. To make determinations about the risk and benefit tradeoffs, the following needed to be understood:

1. To what degree, and in which situations, secondary stakeholders are most likely to object to the use of a device that can potentially cause the recording of their conversation (i.e., are objections unqualified or do they depend on the location, the topic, the identity of the conversation partner, or on the perceived confidentiality of the conversation?); and
2. What application parameters (e.g., retention time) can be adjusted to meet a compromise

between the interest of the primary users and conversation partners.

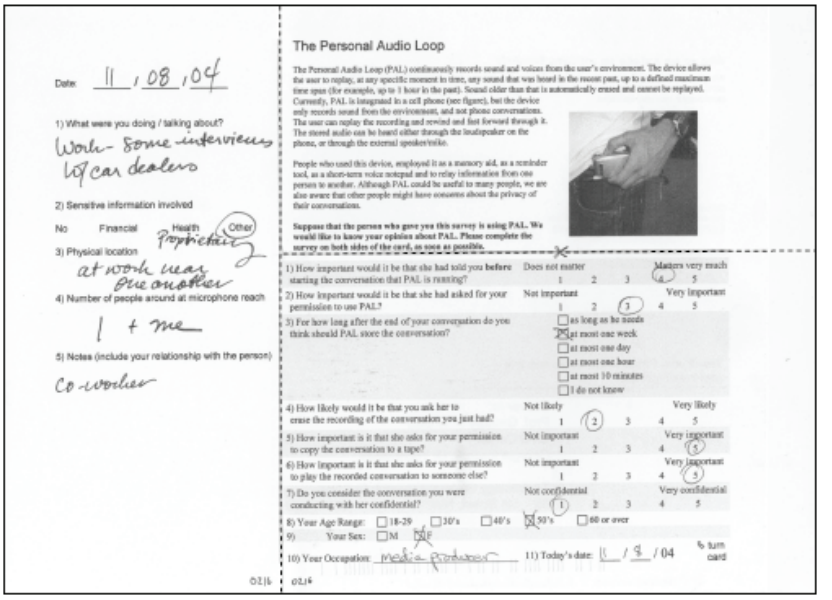
As stated in the section on the deployment of the technology, users might have been frustrated with the limitations of the prototype and thus, did not participate in a long-term study of the service. Rather, an event-contingent experience sampling procedure targeted at the conversation partners of PAL users was designed (Iachello, Truong, Abowd, Hayes, & Stevens, 2006). Experience sampling is a technique that has been in use since the 1970's to measure user responses in everyday settings to various stimuli (Scollon, Kim-Prieto, & Diener, 2003). Wheeler and Reis define event-contingent sampling procedures as those initiated by the occurrence of a specific event, in this case a conversation (Wheeler & Reis, 1991).

Method

A group of individuals whom are called "proxies" acted as substitutes or probes for the researchers in conducting the study. As the proxies went about their daily activities, they asked people older than 18, acquaintances as well as unknowns, at the end of their normal conversations, to read a description of PAL and to fill out an anonymous survey about the application. The proxies were instructed to hand out the surveys only if the conversation exceeded 3–4 sentences—weeding out very short interactions such as "thank you" and "excuse me." Also, the survey was only submitted once to the same individual to avoid respondent bias. The survey was administered in various locations in the United States.

The goal of the survey was to link the questioning to real life experiences as closely as possible. Thus, participants were asked to imagine that the person with whom they had just been speaking was using PAL. The survey included two parts, linked by a unique number (Figure 4). The investigator completed the first part (on the left) with information about place, participants, and the activity being achieved with the conversation, as salient elements of the social setting (Goffman, 1966). The investigator then detached the portion on the right and gave it to the participant along

Figure 4. Example of "proxy" survey. The survey form is divided in three parts, here shown after being reassembled. The left side is filled out by the researcher. The right side is given to the participant. It contains a description of the application and the survey.



with a consent notice to participate in the study. The proxy asked the participant to fill out the survey immediately if possible, to increase recall accuracy. Otherwise, the survey portion of the card was return addressed on the backside and could be mailed back at the participant's convenience (a postage stamp was affixed for this purpose). The questionnaire included six questions on a 5-point scale, one multiple-choice question and, on the backside, a blank space for optional comments in addition to the lab's address and space for postage. This structure minimized completion time and, in fact, most participants were able to complete the survey immediately.

Results

Of 45 distributed surveys, 41 usable responses were received. Only one person refused to accept the survey at all. Most surveys were completed immediately, and nine were mailed back afterwards. Twenty-four respondents were in IT or research occupations (students, research scientists, university professors, etc.). The remaining respondents ranged across professions, including: teachers, designers, hairdressers, managers, attorneys, and business owners. Respondents spanned all age groups between 18 and over 60, with the distribution biased towards the younger age groups, reflecting the age group of the proxies. Seventeen respondents were female. There was no strong correlation observed between the opinions expressed by participants and type of occupations, age, nor gender.

Both the proxy and the participant were asked to provide a measure of the sensitivity or confidentiality of the conversation they had just conducted. The participant was asked to rate subjective "*confidentiality*" on a five option scale. Proxies indicated whether the conversation was *sensitive* following precise guidelines given by data protection legislation (i.e., financial, health, religious, and some work related topics are sensitive).[2] The difference between confidentiality and sensitivity will be stressed: significant correlation between sensitivity and confidentiality as indicated by the proxies and the participants was not observed.

Participants wanted to be informed that the recording was happening. They wanted to be informed regardless of the sensitivity of the topic of the conversation as classified by the researcher and the place where it happened (public or not). Participants indicated that it was important that the PAL user (the proxy) ask for permission before using PAL in the conversation that had just occurred. Both these variables correlated highly to perceived confidentiality as well. Participants stated it was important that the person using PAL ask permission before he or she copies or replays the audio to others. There were relatively weak correlations of these two variables with confidentiality, suggesting that there is a concern with what happens with the recording, regardless of its perceived confidentiality. The desire to be informed does not necessarily imply that the participants would have likely asked to erase the recording after the fact. In most situations, people would not have asked to erase the conversation. There appears to be a "confidentiality threshold" in this respect: participants would have asked the user to erase the recording primarily in cases of elevated confidentiality. In most cases of conversations of medium or low confidentiality, they would not have likely asked to delete the recording. Finally, participants indicated that a long retention time would not be an issue. The original estimate of appropriate retention time was in the range of 10 minutes–1 hour, but these participants were comfortable with a much longer retention time. Apparently, participants were more concerned with the misuse of the recording (e.g., by replaying to others) than with its mere storage.

CRITICAL FEATURES FOR USE

Informed by findings from the laboratory and diary studies, the multi-month deployment, the exploration of privacy regulations, and the contextualized survey of conversation partners, the critical features of PAL were uncovered. Many of these features fed into the iterative design process of developing the final prototype of PAL on a mobile phone.

Making PAL Useful

Given the rates of 2.5 and 3 incidents per week as reported by the deployment and the diary study, the need for PAL is justified. Analysis of the stated purpose for recovering the audio provided additional information (see Table 1 for a synthesis).

Information minimization requires collecting the minimum amount of personal information needed by the application. Given that 58% of the diary incidents referred to content within 1 hour, a buffer of up to 60 minutes should suffice, with a 15 minute default. EU and U.S. law diverge in this regard, as ECPA does not make any distinction based on stored information retention time. A more conservative way of looking at this issue would be that of understanding the duration of the "social contract," implicit among parties engaged in a conversation, to determine how long a recording can be maintained after the end of such conversation. This measure relates to the relation between distance and place (Citro, Iglen, & Marrett, 2003): how long does it take to move between places with incompatible social contracts? Because PAL could be abused when crossing place boundaries, the recording should be limited to minimize such risks. While valid from a phenomenological standpoint, it was decided to postpone this assessment, given the unsolved issue of gathering reliable contextual data.

Making PAL Ubiquitous

A mobile/wearable solution for PAL was targeted. The intuition was that the mobile phone would likely be with an individual much of the time (at least during working hours, perhaps also at home). Of the participants in the laboratory study who owned a mobile phone, all but one was carrying it upon arrival for the study. Furthermore, in 79% of the diary entries queried, the participant's mobile phone was on her, or within reach. A study was conducted later and it was learned that people's mobile phones are within an arm's proximity to them only 50% of the time (Patel, Kientz, Hayes, Bhat, & Abowd, 2006), however, the mobile phone still represents the only truly personal and ubiquitous device available today.

The results of all of the studies associated with PAL demonstrated the need for and appropriateness of this service to be wearable, as opposed to environmental. The argument can be made that an audio buffering service in the environment might be preferable for a variety of reasons, including power concerns, better audio quality, and the convenience of users not needing to wear a device. Every participant reported, however, that there are times when it would not be possible for the service to be environmental. Every participant who recorded any entries recorded at least one at a public place or outdoors, where environmental solutions would be difficult. Participants also ex-

Table 1. Purpose for recovering audio reported from initial diary study (total 109 entries in diary study)

Purpose category	**Occurrences**
Forgotten previous details (e.g., making a list, retrieving details)	36 (33%)
Replaying for conversation partner (replaying for person who either spoke the audio originally or was present to hear it)	20 (18%)
Interrupted (external activity took focus away from important audio)	18 (17%)
Explicit tape recorder behavior (participant was aware prior to the incident that she wanted to record it)	13 (12%)
Distracted (another concurrent activity took attention)	13 (12%)
Relaying information from one partner to another (replaying for person not present when original audio was recorded)	9 (8%)

pressed control concerns about an environmental version of PAL versus a wearable solution. One participant noted, "[I would] rather have the control of it being on my person."

While advocating a wearable solution, however, participants were not interested in a completely separate device but instead as a "value added feature" to the mobile phone already owned and carried. Although this may seem obvious in retrospect, it implies the fairly strict requirements that PAL must run unattended on the mobile handset, without recharging for at least a day, and it must not interfere with the call functions of the phone. These requirements are met by the currently deployed prototype, resulting in an arguably ubiquitous service.

Making PAL Usable

The final PAL prototype provides asymmetric backward/forward skip features over the recording, with default values of 10 and 5 seconds, respectively. Although most participants of the laboratory study liked these defaults, the values can be adjusted, and anecdotal experience shows that individuals do optimize them. Effective use of fast forward or rewind skimming features were not observed during the laboratory study. Considering the limited capabilities of the handset, it was opted to support earmarks instead. The user can set earmarks and then use the backward/forward skip buttons to traverse these earmarks or simply navigate without using them. One issue identified in the laboratory study related to the mapping of the pair of navigation buttons is: there is no "natural association" between the buttons and backward and forward navigation. The variety of ways the handset can be mounted on a belt or carried in the pocket or purse exacerbates this issue. Thus, the "never-wrong" mapping described in the final prototype was opted for.

Making PAL Socially and Legally Acceptable

The common opinion that people must adapt to technological evolution by changing their social expectations is not necessarily endorsed. However, a case could be made that PAL does not impinge on constitutional rights and that, in the long term, practice could show the harmlessness of this application, granted specific guarantees, namely, small recording radius, short buffer length, and some form of notification to the conversation partners. It is stressed that it is not in the scope of this chapter to provide conclusive legal opinions—a task best left to courts and DPAs. The purpose is to provide a balanced, if necessarily concise, overview of PAL's social and legal impact.

A number of different stakeholders can be identified with regards to PAL. Three are considered: the user, conversation partners, and unrelated third parties. Considering the third category, diary results indicate that 69% of the entries related to recordings in public or semi-public spaces and 44% stated that other, unrelated people were present. These figures support the concern with third-party privacy which contrasts with the fact that the vast majority of participants neither were preoccupied with a third party's privacy nor with that of the conversation partner. These observations are particularly interesting because they diverge from legislation in force. ECPA does prohibit capturing a third party's conversation when the owner of the device is not part of that conversation and the conversation takes place with reasonable expectation that it is not being intercepted (e.g., non-public space). On the other hand, it must be noted that the perceptual properties of sound might not grant constitutional basis (in the U.S.) for an expectation of privacy in public space, as suggested among others by numerous cases adopting the "plain view" rule. This could allow adapting surveillance legislation to permit limited memory aid devices such as PAL.

Interface affordances and information retention policies greatly impact social acceptability. Altering the coverage of the microphone is an essential factor of a proportionality determination, as suggested by analogous DPA opinions involving personal uses of video surveillance (namely, outdoor camera units at home entrances) (Ackerman, Cranor, & Reagle, 1999). Likewise, DPAs have used retention time and deletion policies to

evaluate the social impact of surveillance applications. Completely eliminating the risk of recording third parties' conversations is extremely difficult, given the characteristics of sound transmission. PAL arguably presents lower risks than traditional audio recorders; the retention properties of this application do support the claim that PAL does not serve archival purposes, nor does it vastly facilitate surveillance.

In the relationship with conversation partners, informed consent is one fundamental tool of social action, embodied in privacy law. Its implementation presents formidable technical and usability challenges. In this case, anecdotal evidence collected during the deployment suggests that the participants have, over time, renounced to preventively explain or ask permission to use the service. At times, participants turned off the device due to social pressure. Both observations support the previous findings from the diary study. This could hint at a gradual adaptation to the technology and the adoption of appropriate social behavior, similarly to what is currently happening with camera phones.

Although strong feedback was not received from the participants requesting that PAL provide a notification cue while recording, in view of the considerations, it was decided that such function be incorporated in the deployed handsets. When recording, the outer LED integrated in the round ornament on the phone shell (see Figure 1) lights up red. During playback, the light turns green. Although recording is usually associated with a red indicator, there is awareness that people might not understand its meaning and those users could obviously conceal the LED as well as the recording device: the user remains ultimately responsible for abiding to the social contract and mores.

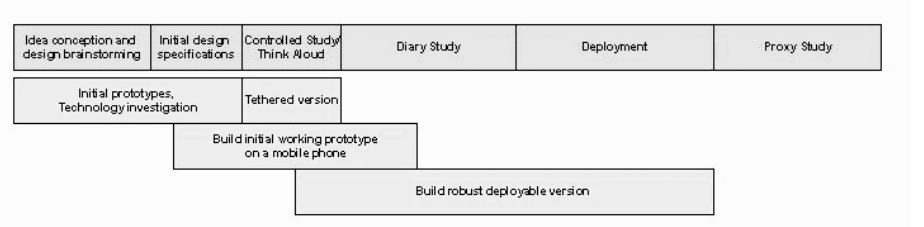
FUTURE TRENDS IN DESIGNING FOR MOBILE APPLICATIONS

As computing moves more off the desktop and mobile devices become more computationally powerful, we as a community must begin to examine the ways in which we design, develop, and evaluate mobile applications. Many lessons can be learned from the 4 year process of designing, developing, and evaluating the Personal Audio Loop. Several important features may be valuable for other researchers and designers to consider. First, the team had a fundamental advantage in conducting this type of design work based on the diversity of skills and perspectives represented. Expert mobile developers used their technical skills to prototype working applications on the mobile phone very quickly. Additionally, they were able to rapidly iterate on the software as the project progressed. Designer experts in conducting formative and qualitative studies as well as those who were knowledgeable in legal and regulatory issues, determined how the design complied with laws and social considerations. This team of varied expertise afforded a mixed-method evaluation of the designs, including the lab studies, diary studies, deployment studies, and the proxy study.

Concurrent development and evaluation of prototype systems throughout the project also contributed significantly to its success. The timeline outlines the general schedule for the phases of the project (see Figure 5). Notably, the development of the working prototype occurred concurrently with the laboratory study and the diary studies. The development of the tethered prototype allowed these studies to be conducted while the development occurred.

Finally, this project included the use of not only tried and true methods but also those that were invented or adjusted for the particular needs of mobile applications, most significantly, the proxy study. Many mobile applications will be used in public settings, and thus exploration of direct situated public reaction was essential. Few well tested techniques would have provided this information. Using proxies in a field study, however, provided the data to allow for a determination of bystander reaction. This study revealed some surprising results (such as bystanders not caring how long data is kept) and was essential to the project. These types of studies, relatively lightweight, could be deployed by several people over a short amount of time with minimal effort.

Figure 5. Timeline of the entire development cycle. The blue cells indicate the studies conducted and the yellow cells show the development phases in parallel



Despite its many advantages, the approach also left room for improvement upon which must also be reflected on. First, there still exists a need for a longer, richer, more in depth deployment study with more users. Several people using PAL as part of this project carried an additional mobile device, and thus it would be difficult for them to adjust to having a new device primarily for this service for the long term. Additionally, it was not deployed for long enough to determine how long it would take for the novelty effect to wear off. Adding more users to the deployment study would also have provided better results. Though choosing a wide range of people was tried, it was difficult to make categories of user types with such a small sample and with such infrequent use.

When considering applications for mobile phones and other mobile devices, designers and developers must recognize the varied approaches and skills that must be applied to a particular project. The Personal Audio Loop was created based on formative and summative evaluations, legal and social considerations, and technical advances and expertise. Without this multi-angled approach, the project would not have been possible. In the future, it is expected that more project teams use similar processes and team compilations to push further the boundaries of usability, usefulness, novelty, and adoption in the mobile applications space.

CONCLUSION

Based on controlled and field studies of use of a mobile audio-based memory aid, it is concluded that not only is the service desirable for users, but also that its implementation on a mobile phone is possible and usable. Users find the information needed in less time than they reported being willing to spend. They need this service at least once a week and they are willing to carry a mobile phone most of the time to have access to it. The analysis shows that this application falls within a legal "gray area" and that its legality cannot be definitively asserted or denied. The interface and retention characteristics of the application, along with observation of initial deployment, suggest that the application might be socially acceptable.

Participants stated that awareness about PAL was important to allow "boundary-setting" to occur. They were not concerned as much by retention time as with potential misuse of the recordings. They also stated that they would have rarely asked to delete a recording after the fact. These observations have broad consequences because they suggest that traditional privacy guidelines and quantitative privacy policies may not be appropriate or sufficient for the development of this kind of personal ubicomp applications and that designers should focus on the purpose of use of information and interpersonal dynamics instead.

REFERENCES

Ackerman, M., Cranor, L., & Reagle, J. (1999). Privacy in e-commerce: Examining user scenarios and privacy preferences. In *Proceedings of the ACM Conference on Electronic Commerce 1999* (pp. 1–8). New York: ACM Press.

Arons, B. (1993). SpeechSkimmer: Interactively skimming recorded speech. In *Proceedings of UIST'93: The 6th Annual Symposium on User Interface Software and Technology* (pp. 187–196). New York: ACM Press.

Citro, C. F., Iglen, D. R., & Marrett, C. B. (Eds.) (2003). *Protecting participants and facilitating social and behavioral sciences research*. Washington, DC: National Academies Press.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the CHI 2004: ACM Conference on Human Factors in Computing Systems* (pp. 175-182). New York: ACM Press.

Deitz, P., & Yerazunis, W. (2001). Real-time audio buffering for telephone applications. In *Proceedings of the UIST 2001: The 14th Annual ACM Symposium on User Interface Software and Technology* (pp. 193–194). New York: ACM Press.

Goffman, E. (1966). *Behavior in public places*. New York: Free Press.

Hayes, G. R., Patel, S. N., Truong, K. N., Iachello, G., Kientz, J. A., & Farmer, R. (2004). The personal audio loop: Designing a ubiquitous audio-based memory aid. In *Proceedings of the Mobile HCI 2004: The 6th International Symposium on Mobile Human-Computer Interaction* (pp. 168–179). Berlin/Heidelberg: Springer.

Hindus, D., & Schmandt, C. (1992). Ubiquitous audio: Capturing spontaneous collaboration. In *Proceedings of the CSCW'92: The ACM Conference on Computer Supported Cooperative Work* (pp. 210–217). New York: ACM Press.

Iachello, G. (2006). *Privacy and proportionality*. Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA.

Iachello, G., Truong, K. N., Abowd, G. D., Hayes, G. R., & Stevens, M. M. (2006). Event-contingent experience sampling to evaluate Ubicomp privacy in the real world. In *Proceedings of the CHI 2006: ACM Conference on Human Factors in Computing Systems* (pp. 1009–1018). New York: ACM Press.

Patel, S. N., Kientz, J. A., Hayes, G. R., Bhat, S., & Abowd, G. D. (2006). Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. In *Proceeedings of the Ubicomp 2006*: *The 8th International Conference on Ubiquitous Computing* (pp. 123-140). Berlin/Heidelberg: Springer.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4*, 5-34.

Terrell, T., & Jacobs, A. (2002). Privacy, technology, and terrorism: Bartnicki, Kyllo, and the normative struggle behind competing claims to solitude and security. *Emory Law Journal, 51*(4), 1469–1511.

Vemuri, S., Schmandt, C., Bender, W., Tellex, S., & Lassey, B. (2004). An audio-based personal memory aid. In *Proceedings of the Ubicomp 2004: The 6th International Conference on Ubiquitous Computing* (pp. 400-417). Berlin/Heidelberg: Springer.

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59*(3), 339-355.

KEY TERMS

Audio Buffer: Some quantity of constantly recording audio that can be accessed for a set period of time before being deleted automatically.

Deployment Study: A study that asks people to use a new piece of technology as part of their daily lives and activities for some extended period of time and typically includes interviews, surveys, and other means of measuring user response to the technology.

Diary Study: A study that asks people to keep a diary or journal of their interactions with a computer system, any significant events or problems during their use of a system, or other aspects of their working life.

Event-Contingent Sampling: A variation of the experience sampling method in which users are probed upon the occurrence of a particular event.

Experience Sampling: A set of techniques to capture people's behaviors, thoughts, or feelings as they occur in real-time, also known as a pager study.

Memory Aid: Tool or other support designed to augment human memory or correct for human memory error.

Personal Audio Loop: Mobile memory aid application buffering audio on mobile phone.

Usability: The effectiveness, efficiency, and satisfaction with which specified users of a particular piece of technology can achieve specified goals. Often considered a basic requirement of all new interactive applications and sometimes regulated using ISO9241 11.

Usefulness: As applied to mobile applications, usefulness has to do with the specific utility of that application in the user's life. Often, this level of usefulness directly impacts the adoption of the mobile application or device and may be measured by deploying the technology to such users.

ENDNOTES

1 The United States do not have DPAs specifically appointed to examine privacy issues.

2 Proxies were not asked to rate confidentiality because it was wanted that they collect as objective information as possible.