Designing Auditory Interactions for PDAs

Debby Hindus
Interval Research Corporation
1801 Page Mill Road, Palo Alto, CA 94304
Phone: 415-842-6067 Email: hindus@interval.com

Barry Arons
Speech Interaction Research

Lisa Stifelman
MIT Media Laboratory

Bill Gaver
Royal College of Art

Elizabeth Mynatt
Georgia Institute of Technology

Maribeth Back
Harvard Graduate School of Design

ABSTRACT
This panel addresses issues in designing audio-based user interactions for small, personal computing devices, or PDAs. One issue is the nature of interacting with an auditory PDA and the interplay of affordances and form factors. Another issue is how both new and traditional metaphors and interaction concepts might be applied to auditory PDAs. The utility and design of nonspeech cues are discussed, as are the aesthetic issues of persona and narrative in designing sounds. Also discussed are commercially available sound and speech components and related hardware tradeoffs. Finally, the social implications of auditory interactions are explored, including privacy, fashion and novel social interactions.

KEYWORDS: Auditory interactions, speech interfaces, auditory cues, PDAs, user interactions, interface design, auditory icons, sound design, social effects of technology

INTRODUCTION
Audio is coming into its own as a medium for interacting with computing devices. This is due in part to technological progress and in part to a multimedia-heightened appreciation for the richness of auditory presentations. In addition, there has been considerable progress in designing and implementing auditory user interfaces (AUIs). However, in traditional user interface design, speech and audio for user interaction have been thought of as an adjunct to graphical user interfaces, or as a flawed substitute when GUIs are not available, as in telephone-based interactions.

This panel will elucidate the potential for AUIs by having the panelists address the design, implementation, aesthetic, and social aspects of creating an auditory user interface for a personal digital assistant (PDA). This hypothetical PDA is very small, with no graphics or minimal graphics for most uses. It is intended to be carried on one's person, so that the panelists will have to think about office, home, car, and public environments.

Each panelist will focus on a particular aspect of AUI design, using several concrete interactions as a starting point for drawing out the issues in depth. Questions to be posed include: What are the general principles to be considered? Can the design be implemented? What would be hard? What would be easy? Does it have to be transaction-based? What tools would you need? Do you have them? Who needs to be on the team? How did you go about your design/evaluation?

The goal is for the audience to come away with a clear idea of the issues, choices and tradeoffs involved in designing and building auditory interfaces, particularly AUIs for small devices.

PANELIST: LISA J. STIFELMAN
Several years ago, the Speech Research Group at the Media Lab explored the idea of a speech-driven, handheld computer with a microphone, speaker, only a few buttons, and without a keyboard or screen. My master's thesis, VoiceNotes, provided an interface for capturing and getting random access to spontaneous thoughts, ideas, or things-to-do in contexts where writing would be inconvenient. Today, a host of new PDAs have entered the market; while the devices are mobile, they are awkward to interact with when the user is mobile (e.g., driving, walking) or engaging in other activities. In addition, they are generally too large to keep in a shirt, pants pocket, or wallet. When devices are small enough to carry with us at all times and use in a variety of contexts, speech and sound will become more prevalent components of interaction. Most of today's PDAs don't even come with a microphone (e.g., Newton) let alone a coherent audio interface. A variety of VoiceNote-like devices have also appeared in the last several years (e.g., Voice Organizer) but they are uninformed by basic principles of speech user interface design.

What is needed to make usable and commercially viable audio-based PDAs? First, we must begin with basic

UIST 95 Pittsburgh PA USA
© 1995 ACM 0-89791-709-x/95/11..$3.50

principles for designing with speech and sound to avoid the poor design now prevalent in interactive voice response systems. This poor design has led to terms like "voice mail jail," and the annoying bleeps that emanate out of our desktop computers. Audio interfaces cannot be designed the same as visual ones, and a visual interface should never be directly translated into an audio one.

Beyond the basic design issues associated with any interface employing speech and sound, what issues are specific to an interface for a PDA? Perhaps the interface with the greatest similarity is the telephone. There are distinctions, however, between a telephone-based audio interface and one for a PDA. Telephone interfaces are transaction-based—they have very definite beginnings and endings; a device that is always with you requires a more continuous style of interaction, without a definite start or finish. An audio interface for a PDA will also be strongly influenced by the device's physical form. Form reveals functionality, and provides affordances for use. Should the device be handheld or worn like clothing? Where should the microphone(s) and speaker(s) be placed or worn? Can the user easily alternate between talking, listening, and looking? Some people envision a PDA like a Dick Tracy wrist watch, but imagine bringing the device close to your mouth for private talking and then having to pull it further away to look at a tiny display. The form of the device also has a strong influence on the modalities employed—if a device is handheld it may have physical buttons or a display, while if it is worn like a pin (as on Star Trek), the interaction may employ audio alone.

Since the device will be with us at all times, unlike a desktop computer, it will be used in a variety of social contexts and acoustic environments. How can the style of interaction be adapted to the user and the context of their current activity? This is critical as audio will not always be appropriate—e.g., when others can overhear and the information is private, when talking aloud would be considered rude. Imagine a PDA used in a hospital, where the doctor dictated in front of the patient "condition is worsening, may have to amputate!" The form of the device also impacts its social acceptance. Should an audio-based PDA be shaped like a cellular telephone so one can appear to be talking to another person rather than to a machine or to oneself? The acoustic context is also very important. Although speech is useful when the user's hands and eyes are busy as when driving, a car environment is noisy, causing speech recognition errors and increased difficulty in understanding synthesized speech output.

Perhaps most importantly, we need to develop metaphors for audio user interfaces. The Macintosh user interface laid the ground work for direct manipulation, providing a consistent style of interaction that could be applied across applications and machines. Visual metaphors (files, folders, desktops) pervade today's graphical interfaces. What will be the new metaphors for tomorrow's audio interfaces?

Biography Lisa Stifelman is a doctoral candidate at the MIT Media Laboratory. She has been a user interface designer and developer at AT&T, in the Human Interface Group at Apple Computer, Inc., and is currently an Interval Research Corporation student fellow.

PANELIST: ELIZABETH D. MYNATT

For the past four years, I have been investigating the design of a system that transforms graphical interfaces into auditory interfaces. This system, called Mercator, provides access to graphical user interfaces for blind computer users. This project has given me the interesting challenge of representing typical graphical user interfaces concepts (such as windows and menus) with auditory cues.

In many ways, the design of an audio-based PDA will hold the same challenges. Much of the untapped potential market for PDAs is made up of users who already have experience with desktop graphical interfaces. One way to support these users is to provide an interface that leverages their knowledge of graphical interfaces. A central question in this effort is determining what characteristics of these interfaces are central to our understanding of graphical interfaces. Most existing commercial PDAs have tried to move graphical interface concepts into PDAs and have failed because they have moved too much of the space-consuming implementation of graphical interfaces into much smaller visual displays. One challenge to the designer of an audio-based PDA will be to support interaction concepts such as direct manipulation that are often supported well in graphical interfaces.

Another interesting design challenge with Mercator that is also critical in the design of an audio-based PDA is finding the balance between using speech and nonspeech auditory cues. For example, most of the Mercator's representations of graphical interfaces, excepting textual information, is conveyed with nonspeech auditory cues. This design is based on the assumption that analogous speech cues would be more distracting and would take longer for the user to process mentally. Many speech interface designers would argue against this assumption, adding that speech output is also less ambiguous.

When discussing the design of auditory interfaces, the concern that nonspeech auditory cues are not appropriate for business-oriented applications is generally raised. It is interesting to note, however, that both the Newton and General Magic's Magic Cap make use of simplistic nonspeech auditory cues. These cues aid the user's understanding of pen input, for example, conveying when a pen-based gesture is recognized by the system. It follows that nonspeech cues could serve the same purpose in a voice recognition interface. Perhaps nonspeech cues could fulfill the same role as body language, which is critical in human to human voice interaction.

Despite concerns about synthetic auditory cues, people utilize real-world auditory cues in many situations. If actions such as turning pages, writing notes or opening a mailbox didn't make sounds, we would be quite disoriented.

Given that most PDAs will not support tactile feedback, the incorporation of nonspeech auditory cues is even more important. Auditory feedback can also be utilized by those in the periphery of the PDA. For example, when an office assistant is typing a form based on my spoken input, it helps to hear the keystrokes and line-feed. These auditory cues give us an awareness of what someone is doing, assisting our interaction with them.

Nonspeech cues can enrich a PDA interface in many other ways. Similar to the use of graphical icons, auditory icons, which leverage the user's understanding of the physical world, can help convey many underlying metaphors that might be used in a PDA interface. Likewise, nonspeech auditory cues can be used to extend the user's perception of physical spaces. Just as baby monitors allow us to listen in on the nursery while in other parts of the house, audio PDAs could help us monitor events in home and office environments. Given sufficient infrastructure, we could hear symbolic cues when a colleague returns to the office, when the kids are home from school, or when an important package has arrived.

Certainly the question is not whether we should use nonspeech auditory cues in a PDA, but how to effectively utilize these cues for a wide variety of purposes.

Biography Elizabeth Mynatt is a doctoral student in the Graphics, Visualization and Usability Center at the Georgia Institute of Technology. She directs the Multimedia Computing Group which focuses on research in auditory and collaborative environments. In addition to her work with blind computer users, Mynatt is interested in the use of audio to support serendipitous collaboration and interaction in virtual spaces.

PANELIST: BARRY ARONS

The time is right for integrating speech and audio into user interfaces and applications. This is particularly true for mobile devices such as PDAs and personal organizers where the display is typically one of the most expensive components, and often dictates the overall size of the device. An auditory interface could lower the cost of these types of devices, make them truly palm-sized, and allow them to be used while mobile (i.e., while driving or gardening without being forced to look at a screen).

The processing and storage devices needed for building auditory interfaces are now becoming available in low-cost, low-power, packages. A speech PDA should at least contain a speech compressor to maximize the capacity of user recordings and prompts in a limited data store. However, designing a more advanced auditory-based PDA requires a synergy of speech and audio processing technologies. Depending upon the target usage of the PDA, a more capable device would include some form of speech recognition for issuing commands, time compression to enable efficient interactions, text-to-speech synthesis for presenting free-form text messages, and a sound synthesis module for generating nonspeech auditory feedback.

There are many tradeoffs that can be made in terms of hardware, software, and user interface design in an auditory PDA. Is it best to use general purpose processor, a specialized digital signal processing chip, or a RISC chip? Is flash RAM, battery backed up DRAM, or a tiny hard disk the best way to store audio data. Should the speech recognizer be speaker-dependent, speaker-independent, or a hybrid? Should the device have a display or be audio-only? Is building a stand-alone device the best way to provide the desired PDA features, or would these features be better integrated into a service accessible through a cellular phone?

Unfortunately, there is no single right answer to these questions. In designing an audio PDA, there are difficult tradeoffs in areas of development time, power consumption, and cost. In terms of user interface design there are additional tradeoffs in adding new features versus simplicity, uniformity, and ease of use. An audio PDA should not be thought of as a collection of technologies, but as a synthesis of both hardware and software components that provide users with tools for getting work done in a very portable and easy to use package.

Biography Barry Arons has been researching speech-based user interfaces and developing software tools for supporting audio in workstations since 1982. He received his Ph.D. from the MIT Media Laboratory in 1994, and is a consultant in the area of speech interfaces for mobile devices.

PANELIST: BILL GAVER

Technologies are fast developing that make it possible to produce audio-only PDAs. But do we want to? There's a certain thrill about the idea of carrying around a small device that you can talk to, and that will use speech and sound to communicate back. But what of the social effects? Would such a device join slide rules and calculators as nerd-badges, signaling that the user is a slavish follower of new technologies without regard to social consequences? Would it convey the same sort of public display of self-importance that portable telephones sometimes do?

The problem here is that audio is an inherently public medium. You can be relatively discrete using a PDA with a graphical interface. But imagine sitting in the midst of a crowd of people all talking to their PDAs, and, worse, hearing their PDAs talk, whistle, beep and clatter back at them. Imagine your PDA reminding you of a meeting in the middle of a concert hall. Imagine trying to sleep on a plane while your neighbor works on her to-do list. Clearly, annoyance will be a crucial issue for such gadgets.

Moreover, an audio PDA is like a portable telephone in pulling its user to engage in an electronic world at the same time that it disrupts the local environment. As you become increasingly involved with your PDA, you are likely to become less engaged with your physical surroundings. And as you do so, the broadcast nature of audio blurs the boundaries between local and electronic environments, making it dangerous to forget that you are simultaneously in both.. While you are concentrating on telling the PDA

about your latest development project, who's listening from behind? Will we have to develop subtle codes for communicating with our computers so that eavesdroppers can't understand (e.g., don't forget to take the c-a-t to the v-e-t)? Privacy is another key issue for the development of these devices.

But finally, the inherently social nature of audio provides unique opportunities as well as problems. Sounds can be annoying, but they can also provide subtle, unobtrusive information about relevant events. Withdrawing into cyberspace can be antisocial and disruptive, but technology can be used as a medium for community as well. Imagine having a PDA that would play sounds indicating ongoing and planned events, both within your local environment and from other places around the world. Imagine using a PDA as a surrogate for your presence at a remote meeting. Imagine being able to whisper comments about a talk (or panel discussion) to a friend across the auditorium, or using a PDA just to chat with other users. From this perspective, such devices could provoke and provide access to new sorts of social groupings. Your children will love them.

The challenge is to design audio PDAs to be more than noisy versions of graphical ones, by balancing the tensions between annoyance and information, privacy and community. Careful sound design can help reduce annoyance, and a combination of sound and product design can help protect privacy. At the same time, designing to embrace the inherent sociality of sound might allow audio PDAs to become powerful catalysts for community.

Biography Bill Gaver is a tutor and researcher with the Computer Related Design course at the Royal College of Art, and a consultant for several companies. His work focuses on nonspeech audio interfaces, collaboration, and ecological approaches to design.

PANELIST: MARIBETH BACK

Aesthetic design for an audio PDA must overcome the constraints of a severely limited playback device with impatient users who need interactivity. Because our reactions to audio tend to be personal, allowing PDAs to take on individuality will be important. Sonic personas may be incorporated into the devices: my PDA will sound like a friendly kitten, while yours may sound like Larry, Curly, and Moe. Our PDAs may sound businesslike during the day, and more individually expressive in the evening.

The sonic aesthetics of the PDA may be enhanced by use of some standard sound design practices from other disciplines. The design of auditory icons or other nonspeech audio cues involves careful understanding of what people expect to hear. As a theatrical sound designer, I found that my task was not to replicate a particular sound faithfully, but rather to construct a version of it that matched a sort of aural template in the audience's minds. These aural templates are mental patterns developed from the instances of a particular sound that a person has heard. People's memory patterns of a sound are often strongly influenced by mass media versions of it.

Narrative is another useful sound design device. The hearer's aural template is especially forgiving if the narrative context that cues it is well designed. Maintaining a consistency within the PDA's aural iconography allows the user a sense of relative place, and of moving logically from one place to the next. Such shifting between related subsets of audio icons may be useful in a navigational context. For example, if the current set of sounds happens to be primarily mechanistic, it is inconsiderate of the designer to insert a single animal sound; this forces the user to switch narrative context.

Narrative is also important in the construction of a single audio icon. In order for a sound to "read," that is, make sense to the hearer, it must contain a logical sequence. When a glass shatters, it cracks explosively, then the pieces collapse to the ground, then a few final shards tinkle before coming to rest. These sounds in a different order do not describe a glass shattering: the tinkling cannot come before the explosion. The sequence tells a miniature story; thus, the sound contains a simple narrative structure. Elements of the narrative can be designed separately to be most effective.

If the sound designer understands the essential elements of a particular sound's narrative, each element can be greatly condensed: a jet can fly overhead in seven seconds, a toilet can flush in four. In the constrained environment of the audio PDA, recognizability, rather than realism, is the designer's aim. Thus, the narrative of a designed sound can be even further compressed

Finally, understanding the impact of technical decisions upon design quality at every stage of the sound's signal path is vital in developing an aural aesthetic. Frequently, even a well-produced sound loses its meaning when it loses its bandwidth or when it is played on a small and inaccurate speaker. For the designer, technical limitations imply an iterative design process that includes listening to the sounds on the final delivery platform. Adopting this approach throughout the design process will produce a more coherent, more elegant, and more usable product.

Biography Maribeth Back is a doctoral candidate at Harvard's Graduate School of Design. Her professional work includes sound design for theater, video and CD-ROM, virtual environments, art and science museum exhibits, and performance installation pieces. For four years she was the resident sound designer for the American Repertory Theater in Cambridge, MA.