Audio Aura: Light-Weight Audio Augmented Reality

Elizabeth D. Mynatt, Maribeth Back, Roy Want, Ron Frederick
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
[mynatt,back,want,frederick]@parc.xerox.com

ABSTRACT

The physical world can be augmented with auditory cues allowing passive interaction by the user. By combining active badges, distributed systems, and wireless headphones, the movements of users through their workplace can trigger the transmission of auditory cues. These cues can summarize information about the activity of colleagues, notify the status of email or the start of a meeting, and remind of tasks such as retrieving a book at opportune times. We are currently experimenting with a prototype audio augmented reality system, Audio Aura, at Xerox PARC. The goal of this work is to create an aura of auditory information that mimics existing background, auditory awareness cues. We are prototyping sound designs for Audio Aura in VRML 2.0.

KEYWORDS: Audio, Augmented Reality, Auditory Icons, Active Badge, VRML

INTRODUCTION

For many people in office environments much of the content of their jobs (e.g. online documents, calendars, email) is contained within computational devices. Of particular importance is timely information (e.g. the arrival of an email message or the activity of a collaborator) that is generally accessed from a desktop computer. This reliance on desktop computers forces people to focus on a small box in their environment instead of allowing them the freedom and richness of the entire physical environment.

One solution is the design of portable devices that offer some connectivity to desktop computers. However, devices like laptop computers, PDAs, and pagers require relatively heavy-weight interaction. Either users must request information such as inquiring if they have new email, or the device is difficult to ignore, such as a buzzing pager.

We are exploring a different approach to connecting people with computational devices that is based on two concepts. First, we leverage the physical environment to trigger the delivery of information. As a user moves through the office place, entering the coffee room or pausing at a colleague's office, information normally contained in one's desktop computer is summarized and sent to the receiver. The second part of our strategy is to present information via rich auditory cues [4] that build on the peripheral auditory cues people constantly process in their normal environment. By using physical-world triggers and auditory cues, we are creating a light-weight interaction that does not require active participation by the user.

AUGMENTING THE PHYSICAL WORLD

Our system, Audio Aura (see Figure 1), is based on three known technologies: active badges, distributed systems and digital audio delivered via portable wireless headphones. An active badge [6] is a small electronic tag designed to be worn by a person. It repeatedly emits a unique infrared signal and is detected by a low-cost network of IR sensors placed around a building. A location server combines all the information culled from the IR sensors, perhaps augmenting it with other information such as online calendars and email systems. Audio cues are triggered by changes in the location database and sent to the user's wireless headphones.

Using the physical world as the trigger for information delivery has been explored by many researchers [3][5]. Most systems have focused on augmenting visual information by overlaying a visual image of the environment with additional information usually presented as text. The most common configuration of these systems is a hand-held device that can be pointed at objects in the environment. The video image with overlays is displayed on the hand held device in a relatively small window. These handheld systems require the user to actively probe the environment as well as then indirectly view a representation of the environment on the video screen. Our system offers two primary distinctions. First, users do not have to actively probe the environment. Their everyday pattern of walking throughout an office environment will trigger the delivery of additional information. Second, users do not view a representation of the physical world, but continue to interact with the physical world that includes additional real-world auditory cues. This lack of indirection changes the experience from analyzing the physical world to participating in the physical world.

Providing auditory cues based on people's motion in the physical environment has also been explored by researchers and artists, and is currently used for gallery and museum tours. The systems that most closely approach ours include one described by Bedersen [2], where a linear, usually cassette-based audio tour is replaced by a non-linear, sensor-based digital audio tour allowing the visitor to choose their own path through a museum.

Several differences between our systems are apparent. First, in Bedersen's system users must carry the digital audio data with them, imposing an obvious constraint on the range and generation of audio cues that can be presented. Second, Bedersen's system is one-way. It does not send information from the user to the environment such as the identity, location, or history of the particular user.

Other investigations into light-weight interaction include Hudson [5] who demonstrated providing iconic auditory summaries of newly arrived email when a user flashes a colored card while walking by a sensor. This system still required active input from the user and only explored one use of audio in contrast to creating an additional auditory environment that does not require user input.

SCENARIOS OF USE

We are currently exploring providing a range of information with Audio Aura:

Employees at PARC can use the system to augment their physical environment with information about their colleagues. For example, they can hear a greeting at the entryway to the office of a colleague

Figure 1: Audio Aura System

who is out for the day. Audio Aura can also create an iconic summary of a person's recent activity (also at the entryway to their office), perhaps indicating that they had left just a few minutes earlier.

The audio content can also reflect the history of the person wearing the device. Following the previous example, if I again passed a colleague's office, the summary of their activity might only reflect the time since my last visit.

Since Audio Aura knows the identity of the user, the system can deliver personalized information such as a summary of newly arrived email messages while pouring coffee in the break room or hearing a reminder for a meeting that is about to start.

Various artifacts can trigger auditory cues; for example, standing near a bookshelf may trigger a message about recent acquisitions. The longer a user lingers, the more detail s/he receives.

SOUND DESIGN

We combine sound effects, music, and voice into a rich, multi-layered environment which allows behavioral complexity. For example, because speech tends to carry foreground information, it may not be heard unless the user lingers in a location for more than a few seconds. For a user who is simply walking through an area, the sounds remain at a peripheral level, both in volume and in semantic content.

Because we intend this system for lightweight interaction, the design of the auditory cues must avoid the "alarm" paradigm so frequently found in computational environments. One method we are exploring is imbedding cues into a running, low-level soundtrack, so that the user is not startled by the sudden impingement of clearly artificial sound. The running track itself carries information about global levels of activity within the building, within a workgroup, or on the network. This "group pulse" sound forms a bed within which other auditory information can lie. We have tried several different types of sounds, varying the harmonic content, the pitch, the attack and decay, and the rhythms caused by simultaneously looping sounds of different lengths. For example, by looping three long, low-pitched sounds without much high harmonic content and with long, gentle attacks and decays, we create a sonic background in which we leave room for other sounds to be effectively heard.

PROTOTYPING

Mapping Audio Aura's complex matrix of system behaviors to a multi-layered sound design has been aided by prototyping it in VRML 2.0, which allows interaction with 3D graphics and audio in Web browsers [1]. By moving through a 3D graphical representation of the CSL Lab and triggering audio cues either through proximity or touch, a user can get a sense of the functionality of the Audio Aura system.

PERSONALIZATION

Mapping of sounds to system functionality in Audio Aura can be done either by a system designer or by the end user, who can personalize their sound environment via a Web-based graphical front end. The user can select sounds from a database or create their own, load them, and then test them via the VRML prototype. These VRML prototypes help users decide what pattern of sounds work best for them.

DEMONSTRATION

The VRML 2.0 prototypes of Audio Aura's current and proposed states form one part of the demonstration of the system. The system itself is tied to the physical infrastructure of the lab where the IR sensors are installed (PARC's Computer Science Lab).
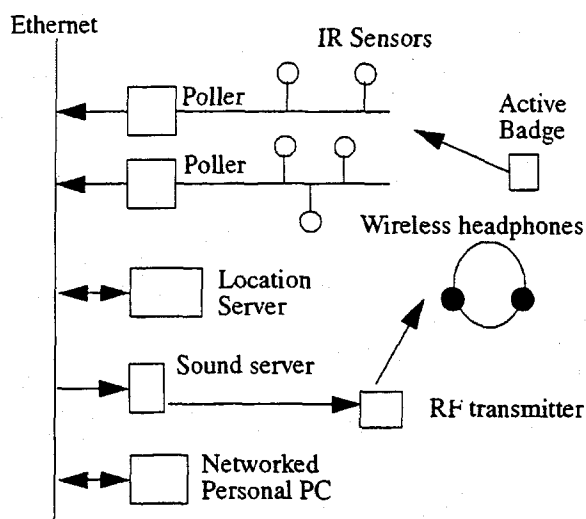
CURRENT AND FUTURE EFFORTS

We have developed several prototypes of Audio Aura, both real systems built on top of the current Active Badge system and VRML representations of Audio Aura's current and proposed behaviors. Identity and location of the user are primary system inputs. Not surprisingly, timing and responsiveness of the system are crucial elements. We are exploring improving the performance of the system as well as optimizing the physical placement of the sensors. Our experiences with Audio Aura indicate that some applications may require marking the presence of the auditory nodes of information (auditory hotspots) through the use of light, sound, or some other dynamic effect. For example, a small light outside a conference room door may indicate the availability of auditory information about that room's schedule.

Aspects of this system that remain to be explored include refining both system granularity and sonic capabilities. Currently we measure users in transit between rooms. Measuring motion within a room and measuring interaction with an object in the room are both logical extensions of the system's capabilities. We are exploring the effectiveness of different types of auditory cues, including speech and nonspeech; as well as the relationship between background and foreground sound in creating lightweight interaction.

REFERENCES

[1] Ames, A., Nadeau, D., Moreland, J. *The VRML 2.0 Sourcebook.* Wiley, 1996. See also the VRML Repository at http:// sdsc.edu/vrml.

[2] Bederson, B.B and Druin, (In press) A. Computer Augmented Environments: New Places to Learn, Work and Play, ed. Jakob Nielsen, in *Advances in Human Computer Interaction*, Vol. 5, Ablex Press.

[3] *Communications of the ACM*, (1993) Special Issue on Augmented Environments, 36 (7).

[4] Gaver, W.W. (1994). Using and Creating Auditory Icons. In Kramer G. (ed), Auditory Display: The Proceedings of ICAD '92. SFI Studies in the Sciences of Complexity Proc. Vol. XVIII, Addison-Wesley.

[5] Hudson, Scott E. and Smith, Ian, (1996) Electronic Mail Previews Using Non-Speech Audio, *CHI '96 Conference Companion*, ACM, pp. 237-238.

[6] Want, R., Hopper, A., Falcao, V. and Gibbons, J., (1992) The Active Badge Location System, *ACM Transactions on Information Systems*. Vol. 10 (1), pp. 91-102.