Nomadic Radio: A Spatialized Audio Environment for Wearable Computing

Nitin Sawhney and Chris Schmandt
Speech Interface Group, MIT Media Laboratory
20 Ames Street, Cambridge, MA 02139
{nitin, geek}@media.mit.edu

Abstract

This paper describes an on-going research project at the MIT Media Lab, exploring the use of auditory I/O as a primary interaction modality for wearable computing. Nomadic Radio is a framework developed for use on a wearable audio platform. It presents timely information and permits communication within a spatialized audio environment. The contextual state of the user indicated by time of day, physical positioning, scheduled tasks, and message content, is used to model information presented in the user's listening space. We are evaluating techniques for the design of spatial audio streams, speech I/O, and auditory cues in the interface. Issues related to asynchronous communication and peripheral awareness are also being considered.

Introduction

Many efforts in wearable computing today utilize a visual approach for information presentation, based on textual or graphical user interfaces. This requires the use of head-mounted visual displays which can be cumbersome and may cause unnatural interference with human vision. They also impose a greater cognitive load on the user and generally necessitate social and cultural adaptation. An "eyes-free" approach, using audio-based augmentation of the physical environment, can be used to express peripheral information and provide passive interaction for specific tasks and usage scenarios.

The goal of an audio environment for a wearable computing system is to convey relevant information to a *nomadic listener* based on the context of her tasks and the timely nature of her messages. Frequently accessed information sources such as voice mail, email, appointments, news, weather forecasts, and traffic reports, can be integrated within such an environment. The MIT Media Lab's *Nomadic Computing Environment* [1] enables subscribers to manage personal information via fax, pagers, and telephony access using digitized audio and synthesized speech. These services are integrated using *Phoneshell* [1], a suite of applications which provide remote telephony access to a variety of information sources.

Simple devices such as pagers provide a convenient form of alerting users to communication and remote information. Yet such devices offer extremely low-bandwidth for communication and the interface does not afford rich delivery of information content. Similarly most cellular telephones do not retain much personal information about their users or their context of use and offer a limited interface to access services easily. Mobile audio devices [2][3][4] with personalized information, localized computing and richer interaction mechanisms certainly point towards audio interfaces and applications for *wearable audio computing*. Audition can be used to enhance an environment with timely information as well as provide a sense of peripheral awareness [5] of people and background events. A combination of wearable auditory I/O and tactile interaction provides a means of unobtrusively augmenting a physical environment for use in everyday life.

Design of Nomadic Radio

In *Nomadic Radio* we provide an integrated environment for personalized messaging on a wearable device with audio as the primary interaction modality. This should allow seamless access to information and communication services to individuals situated away from their desktops such as in meetings, classes or simply while walking. Here the users do not need to actively request information from the system or require foreground interaction. Messages are periodically downloaded to the device throughout the day and selected messages are presented to the user based on her context and desired level of awareness or interruptability. The audio streams are presented simultaneously in specific spatial locations in the user's listening space. A model of the head-related transfer functions (HRTF) permits localization and externalization of sound sources, yet we must consider its effectiveness for segregation [6] of messages in an audio-only nomadic environment. The *AudioStreamer* [7] detected the gesture of head movement towards spatialized audio-based news sources to increase the relative gain of a specific source, allowing simultaneous browsing and listening of several news articles.

In a recent speech-based email application, *MailCall* [8], timely messages were filtered based on the priority and state of the information being presented as well as the context of the user's tasks. A background process, *CLUES* [8], scans various databases and automatically generates filtering rules to categorize messages. We are incorporating timely message filtering to selectively present the appropriate audio messages, coupled with position and environmental context.

In one usage scenario, the listener hears an ABC news summary at a certain time of day, and as she moves closer to a particular room, she is reminded of a meeting there via a brief preview of a related email message, while the news broadcast fades down for a few moments. Textual messages such as email, weather forecasts, and stock reports can be delivered as synthesized speech. All audio messages are localized in the user's listening space based on the time of day and level of priority of the message. Hence messages are placed at chronological positions around the user's head, where the mapping of time to space aids in audio browsing [9]. Timely messages are foregrounded, while others remain in the periphery. The user can hear a quick preview, a summary or the complete message and control the playback rate or skim through the audio. Spoken prompts provide message summaries and status information. Non-speech auditory cues indicate the length and type of message as well as an awareness of background or network events. Users can record their voice as personal memos or for asynchronous communication with other nomadic users.

Implementation

Nomadic Radio consists of client and remote server components that communicate via Wireless LAN and CDPD modems. The *Nomadic Clients*, developed in Java, operate on Toshiba *Libretto* PCs (1.85 lb. Pentium-based mini-notebooks worn on the waist) running the MS Windows 95/NT platform. The *Soundbeam,* an audio I/O device developed by Nortel, is worn around the neck. It consists of two *multipole* speakers for directional audio, mounted on the user's shoulders, and a microphone placed on the chest. Tactile input is provided by a three-button wireless mouse. Spatialized audio is delivered via a Java interface to Intel's *RSX 3D* audio libraries. Speech synthesis and recognition is based on AT&T's *Watson* SDK. On the wearable device, IR and GPS-based receivers will provide positioning data indoors and outdoors. The current architecture relies on server processes (written in C and Perl) running on Sun SPARCstations that utilize the telephony infrastructure in the Media Lab's Speech Interface group. The servers extract information from live sources including voice-mail, email, hourly updates of ABC News, weather forecasts and traffic reports. The clients, when notified, download the text/audio files stored on a web server. Text files are converted to synthesized speech and rendered as spatialized audio in the listening environment.

Future Work

Nomadic Radio can be considered an active information agent that adaptively manages the user's listening space, based on her location, context of activity, prior listening patterns and desired level of interruption. The nomadic client will be designed to capture a continuous model of the user's tasks and listening preferences for more effective and personalized delivery of audio streams. Future work also includes integration of situational awareness on the wearable system via a classification of sounds in the environment [10]. Hence environmental sounds would act as contextual cues for delivery of timely information. A train schedule is played if the system hears the sound of a train approaching or audio sources faded down if the listener were engaged in a conversation. The design of effective audio interfaces sensitive to nomadic listeners and their environments presents a challenge for wearable computing.

References

[1] Schmandt, Chris. "Multimedia Nomadic Services on Today's Hardware". IEEE Network, September/October 1994, pp.12-21.

[2] Stifelman, Lisa J., Arons, B., Schmandt, C., Hulteen, E. A. "VoiceNotes: A Speech Interface for Hand Held Voice Notetaker". *Proceedings of INTERCHI'93*, April 1993.

[3] Roy, Deb K. and Chris Schmandt. "NewsComm: A Hand-Held Interface for Interactive Access to Structured Audio". *Proceedings of CHI '96*, April 1996, pp. 173-180.

[4] Wilcox, Lynn D., Schilit, B. N., Sawhney, N. "Dynomite: A Dynamically Organized Ink and Audio Notebook". *Proceedings of CHI '97*, March 1997, pp. 186-193.

[5] Mynatt, E.D., Back, M., Want, R. and Frederick, R. "Audio Aura: Light-Weight Audio Augmented Reality". *Proceedings of UIST '97 User Interface Software and Technology Symposium*, Banff CA., Oct 15-17. 1997.

[6] Arons, Barry. "A Review of the Cocktail Party Effect". *Journal of the American Voice I/O Society*, Vol 12, Jul 92.

[7] Schmandt, Chris and Atty Mullins. "AudioStreamer: Exploiting Simultaneity for Listening". *Proceedings of CHI '95*, May 1995, pp. 218-219.

[8] Marx, Matthew and Chris Schmandt. "CLUES: Dynamic Personalized Message Filtering". *Proceedings of CSCW '96*, November 1996, pp. 113-121.

[9] Kobayashi, Minoru. "Dynamic Soundscape: Mapping Time to Space for Audio Browsing". *Proceedings of CHI '97*, March 1997, pp. 194-201.

[10] Sawhney, Nitin. "Situational Awareness from Environmental Sounds", MIT Media Lab Tech. Report, June '97. *http://www.media.mit.edu/~nitin/NomadicRadio*