
Navigating Persistent Audio

Aaron Zinman

MIT Media Laboratory
20 Ames St.
Cambridge, MA 02138
azinman@media.mit.edu

Judith Donath

MIT Media Laboratory
20 Ames St.
Cambridge, MA 02138
judith@media.mit.edu

Abstract

This paper gives an overview of RadioActive, a large-scale asynchronous audio messaging platform for mobile devices. It supports persistent chat spaces that allow users to engage in discussion on demand. Our goal is to allow users to easily navigate and participate in large audio-based discussions with minimal cognitive overhead. RadioActive attempts to eliminate problems that habitually plague audio-only designs by using a novel combined visual and audio interface.

Keywords

Asynchronous audio, audio design, BBS, chat, HCI, mobile applications, persistent audio

ACM Classification Keywords

H. Information Systems, H.4 Information Systems Applications, H.4.3 Communications Applications

Introduction

Current mobile technology focuses on one-to-one conversations that tend to isolate individuals from the public environment. In this paper we explore some of the navigational techniques used in RadioActive, a platform that re-engages public interaction as a social catalyst and proxy using voice messaging. Audio is potentially a better medium than text for sociable communication on mobile phones. It is the primary medium that telephones are designed to support, and

connects individually more personally through additional metadata like intonation. Group collaboration studies have also shown that audio promotes higher level, and more social, discourse [7]. This is because humans color speech with intention and degrees of confidence, enabling better inference of meaning and social processes [7]. Furthermore, it is easier to speak than to type using small devices with even smaller keyboards. This is particularly true while walking.

RadioActive allows large-scale audio-based public conversations on mobile phones. It functions as a mobile audio forum, where users leave posts by recording voice messages. Other users can navigate through, listen to, and respond to these messages using the audio or optional visual interface. RadioActive's chat space is persistent, allowing users to drop in and out of discussion as they wish. Such conversations might appear in many forms, from augmenting Podcasts to broadcasting personal blogs within physically local communities. The combined interface is unique for persistent audio chat spaces, and provides an easy way to navigate at a global level. An automated playback mode also allows users to put the device in their pocket and listen passively (Figure 5).

RadioActive attacks several of the technological and design problems that currently make asynchronous many-to-many mobile discussions prohibitive. We hypothesize that RadioActive's combined audio and visual interface will address problems of navigability, cognitive overload, and scalability. This paper outlines our design choices and rationale behind RadioActive.

Designing Multi-Threaded Persistent Audio

Designing interfaces for persistent many-many audio conversations is difficult. Audio itself is serial, which slows navigation. We have chosen to use a visual representation of the chat space to provide instant impressions of the chat space and a more global view. Our design is both highly functional and reflects the abstractness of illustrating structured audio. The visualization centers on representing messages and thread structure through connected circles (Figure 4). Message properties are translated into visual elements, such as size for message length, to differentiate posts with gestalt (Figure 1). In RadioActive, messages have subjects and bodies, similar to email or Usenet. This is particularly important in audio, where the summary speeds navigation. Participants in informal experiments quickly understood the representation.

We support a range of interaction styles to support varied attentional requirements. The most passive version is radio mode, where an adaptive heuristic is used to play through the multi-threaded conversation space without requiring the GUI. It is designed for users who are walking or driving. This method, which uses depth-first search, may skip much of a thread to assure linear presentation, in addition to topic diversity. At points of interest, the user may choose to manually control the navigation. They can use the directional keypad to skip messages along the path, or alter playback speed. Another variation of the radio mode facilitates navigating without the visual interface by using an interactive breadth-first search to cover the chat space more thoroughly. Finally, the visual interface provides complete coverage and control of the chat space.

When users open RadioActive they view their “inbox”, which functions like an RSS reader by aggregating threads (Figure 2). The first message of threads matching subscribed or generated contexts, such as location or topic, are listed with additional information. An alternative inbox view removes additional details, showing only the topic messages (Figure 3). Hovering over messages plays their subject, while activating a node shows the corresponding thread (Figure 4). Once in the thread view, navigation works similarly, where users click and hover over messages to hear their contents. Messages pulsate during playback for visual feedback.

In designing the interface, we have sought consistent design principals that support mobile constraints. Our visual interface is high-contrast and clustered by functionality. Three menu functions keep the big changes simple and contextual. All objects support direct manipulation, and use standard GUI controls where appropriate. Extraneous information is hidden until the proper context, and uses little real estate when shown. Finally, we space items so users can use their fingertips instead of a stylus on touch-sensitive screens. These features help to produce an interface that is simple to use in constrained mobile contexts.

Related Work

Outside of Nishimoto’s Asynchronous Virtual Meeting System [17], little research has been performed on combining visual and audio interfaces for large-scale group chat. In audio-only work, speed is most often addressed with a combination of time-compression [3] and parallel presentation [1, 2]. Previous audio-based navigation themes include hyperlinks [19, 23], 3D spatialized audio [8, 2], earcons [6, 15, 8], and linear

threading models [13, 14]. Spatialization techniques increase cognitive load and work best with multiple sound channels, excluding most mobile devices. As Arons and Fernström have shown [1, 8], it may be possible to combine the visual interface with modified simultaneous audio streams to allow parallel browsing of subjects. However, simultaneous presentation increases cognitive overhead.

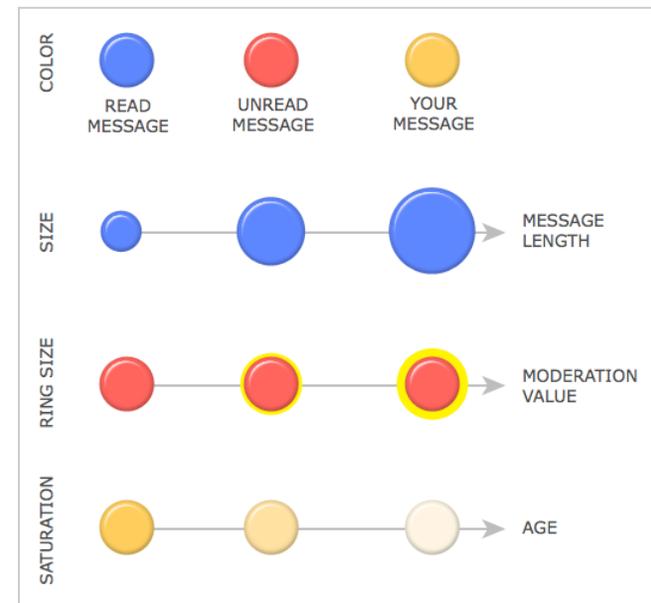


Figure 1. Dimensions of visual encoding for message representation: the different colors represent read/unread status, while size represents message length. Aging is shown by desaturation, and moderation value by increasing the size of a surrounding yellow ring.

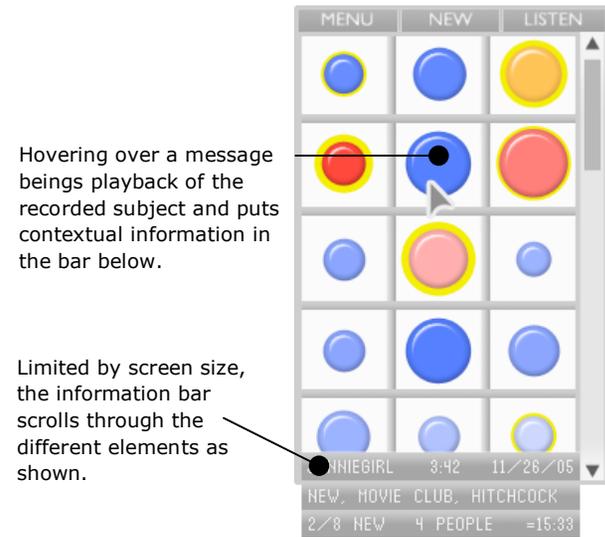


Basic information, such as when the post was created and by whom, is presented adjacent to the visually encoded circle.

Optional text tags help speed navigation. Their aggregation can yield a folksonomie, providing alternative techniques to filter topics.

Figure 2. The "inbox" is shown on the Motorola A1000. Users browse message subjects of initial thread postings by clicking on corresponding menu items or using arrow keys.

Structural data has been conveyed using sonification techniques [12, 16]. However, sonification best conveys single dimensions of data, has high learning curves and artificial mappings, and doesn't effectively separate data points.



Hovering over a message beings playback of the recorded subject and puts contextual information in the bar below.

Limited by screen size, the information bar scrolls through the different elements as shown.

Figure 3. An alternative view to the "inbox". Messages are stripped of extraneous information and tiled for quicker browsing.

Summary

We outline the design of RadioActive, a mobile application that facilitates persistent audio conversations on physically constrained devices. We believe a combined audio-visual interface best makes use of a mobile device's features, and complements audio as a medium. We have developed an intelligent radio mode navigates non-linear threads passively, requiring less interaction. We hope this paper can inform designers of future audio chat spaces of potential interaction strategies.

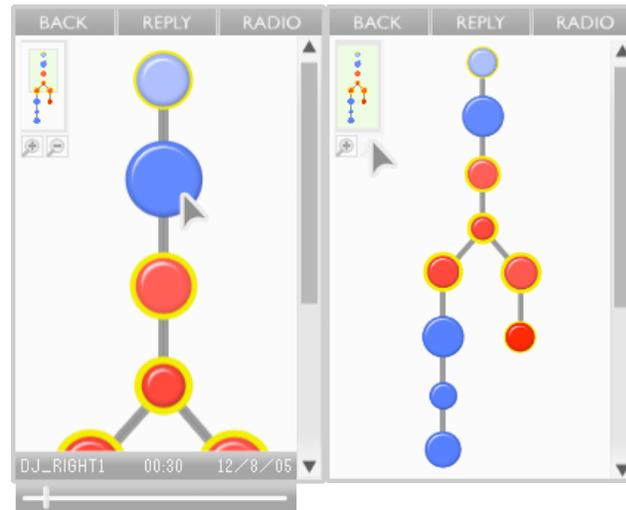


Figure 4. Once a message is selected from the “inbox”, the entire thread is displayed. A tree-view maximizes the vertical space. The first message starts at the top, and is attached to subsequent replies. Hovering over a circle plays its subject. Clicking a circle plays its body. During playback, the information bar alternates between contextual information and a control for scrubbing¹. A representative map in the corner shows the entire thread and features zooming buttons. Users navigate threads larger than the screen by zooming to the desired ratio of detail to size. At that level they may translate across the view using either the scroll bars or by moving the rectangle inside the map.

¹ Scrubbing refers to fast forwarding and rewinding the audio during message playback.

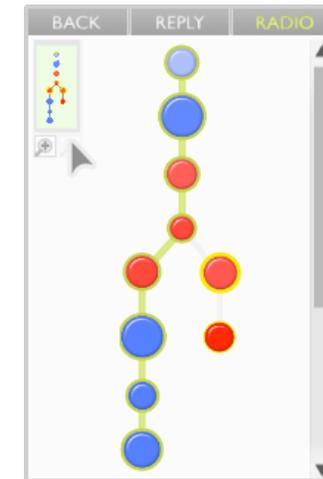


Figure 5. In radio mode, messages along the generated path are encircled in green. Connecting lines are shown in a darker green. All other paths are desaturated.

References

- [1] Arons, B. A Review of the Cocktail Party Effect. *Journal of the American Voice I/O Society*. 12 July, (1992)
- [2] Arons, B. Hyperspeech: Navigating in speech-only hypermedia. *Proc. Hypertext*. ACM Press (1991), 133-146.
- [3] Arons, B. SpeechSkimmer: A system for interactively skimming recorded speech. *Trans on CHI* (March 1997), 4, 1, 3-38.
- [4] Baudisch, P., Xie, X., Wang, C., and Ma, WI. Collapse-to-zoom: viewing web pages on small screen devices by interactively removing irrelevant content. *Proc. UIST '04*, ACM Press (2004), 91-94.
- [5] Bederson B., and Holland J. Pad++: a zooming graphical interface for exploring alternate interface physics. *Proc. UIST '94*, ACM Press (1994), 17-26.

- [6] Blattner M., Sumikawa D., and Greenberg R. Earcons and icons: Their structure and common design principles. *Human Computer Interaction* 4, 1 (1989), pp. 11-14.
- [7] Chalfonte B.L, Fish, R.S, Kraut, R.E. Expressive richness: a comparison of speech and text as media for revision. *Proc CHI '91*. ACM Press (1991), 21-26.
- [8] Fernström, M., and Brazil, E. Sonic Browsing: An Auditory Tool for Multimedia Asset Management. *Proc ICAD '01*, Espoo, Finland, July 2001.
- [9] Goodman, B., Geier, M., Haverty, L., Linton, F., and McCreedy, R. A Framework for Asynchronous Collaborative Learning and Problem Solving. *Proc AI in Education '01*, IOS Press (2001)
- [10] Hewitt, J. Beyond threaded discourse. *Proc WebNet '97*. AACE (1997).
- [11] Hindus, D., Schmandt, C., and Horner, C. Capturing, structuring and representing ubiquitous audio. *ACM Trans. on Information Systems*, 11, 44 (Oct. 1993), 376-400.
- [12] Kilander, F., and Loennqvist, P. A Whisper in the Woods-An Ambient Soundscape for Peripheral Awareness of Remote Processes. *Proc ICAD '02*.
- [13] Kim J.S. TattleTrail: An Archiving Voice Chat System for Mobile Users Over Internet Protocol. MIT M.Eng Thesis, May 2002.
- [14] Lakshmiathy, V., and Schmandt, C. SimPhony, A Voice Communication Tool for Distributed Workgroups. *BT Technology Journal*, 22, 4 (Oct. 2004), 52 - 61
- [15] McGookin, D., and Brewster, S. An Investigation into the Identification of Concurrently Presented Earcons. *Proc ICAD '03*, Boston, MA, July 6 - 2003, 42-46.
- [16] Mynatt, E., Back, M., Want, R., and Federick R., Audio Aura: Light-Weight Audio Augmented Reality. *Proc UbiComp '97*. ACM Press (1997).
- [17] Nishimoto T., Yuki, H., Kawahara, T., Araki, T., Niimi, Y. Design and evaluation of the asynchronous voice meeting system AVM. *Systems and computers in Japan*, Wiley Periodicals (2002), 33, 11, 61-69.
- [18] Ohlund, B., Yu, C.H., and Jannasch-Pennell, A. Impact of asynchronous and synchronous internet-based communication on collaboration and performance among k-12 teachers. *Journal of Educational Computing Research* (2002), 23, 405-420.
- [19] Resnick P. HyperVoice: A Phone-Based CSCW Platform. *Proc CSCW '92*, ACM Press (1992) 218-225.
- [20] Rodenstein, R., and Donath, J. Talking in Circles: Designing a Spatially-Grounded Audio Conferencing Environment. *Proc CHI '00*. ACM Press (2000) 81-88.
- [21] Rohall S., Gruen D., Moody P., and Kellerman S. Email Visualizations to Aid Communications. *Proc InfoViz '01*.
- [22] Ross, K.W. Asynchronous voice: a personal account. *IEEE Multimedia*, 10, 2, April-June 2003
- [23] Sawhney, N., and Murphy A. ESPACE 2: An Experimental HyperAudio Environment. *Proc CHI '96*. ACM Pres (1996), p. 105-106.
- [24] Schmandt, C. Audio Hallway: A Virtual Acoustic Environment for Browsing. *Proc UIST '98*. ACM Press (1998) 163-170.
- [25] Schmandt, C., Kim, J., Lee, K., Vallejo, G., and Ackerman, M. Mediated Voice Communication via Mobile IP. *Proc. UIST '01*. ACM Press (2001), 141-150.
- [26] Smith, M., and Fiore, A. Visualization Components for Persistent Conversations. Microsoft Technical Report MSR-TR-2000-98.
- [27] Viegas, F., and Donath, J. Chat Circles. *Proc CHI 1999*. ACM Press (1999), 9-16.
- [28] Whittaker, S., Hirschberg, J., Amento, B., Stark, L., Bacchiani, M., Isenhour, P., Stead, L., Zamchick G., and Rosenberg, A. SCANMail: a voicemail interface that makes speech browsable, readable and searchable. *Proc. CHI '02*, ACM Press (2002), 275-282.