
RadioActive: Enabling Persistent Mobile Communications for Groups

Aaron Zinman

MIT Media Lab
20 Ames St
Cambridge, MA 02139
azinman@media.mit.edu

Judith Donath

MIT Media Lab
20 Ames St
Cambridge, MA 02139
judith@media.mit.edu

Abstract

Current mobile technology works well to connect individuals together at any time or place. However, general focus on one-to-one conversations has overlooked the potential of always-on group and community links. We hypothesize that asynchronous persistent audio is a superior medium to support scalable always-on group communication for mobile devices. To evaluate this claim, one must first have an adequate interaction design before its possible to investigate the qualities and usage patterns over the long-term. This design does not exist for mobile devices, and is non-trivial given the weaknesses of audio and the practical limitations of mobile devices. Our work takes the first step in this direction by creating and evaluating an initial design called RadioActive. RadioActive is a technological and interaction design for persistent mobile audio chat spaces. Our work focuses on strategies to navigate and structure large asynchronous audio discussions. In this paper we examine related work, describe our approach, highlight a conceptual framework for navigation, discuss our evaluations, and provide suggestions for future research.

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

Social uses of audio, such as the telephone, are typically synchronous. Research has made strides in scaling synchronous group communication [27], but the medium's usefulness is constrained by the cognitive limits of concurrent presentation. Asynchronous communication is time-shifted, which instead allows all participants access to the cataloged information independently. This enables more of "the right people" can be active in the conversation by not constricting participants by who is available during a specific point in time. We already can see the success of asynchronous communication in email, Usenet, and even voicemail, yet there has been a general lack of designs that support group and community-level asynchronous mobile communication. While IM and SMS are useful and ubiquitous, their designs are fundamentally limited by the minimal physical constraints of mobiles. The costs of text display and entry are too high to scale towards *n-to-n* communication sessions. Instead of a text-based approach, we hypothesize that asynchronous persistent audio is a superior medium to support scalable always-on group communication.

To evaluate this claim, one must first have an adequate interaction design before it is possible to investigate the qualities and usage patterns over the long-term. This

design does not exist for mobile devices. Our work takes the first step in this direction by creating and evaluating an initial design called RadioActive. RadioActive is a technological and interaction design for persistent mobile audio chat spaces, where individuals gather to exchange messages asynchronously.

Audio is well suited for sociable communication on mobile phones because 1) it is the primary medium that telephones are designed to support, 2) it is better able to convey emotion and personal identity than text, and 3) humans color speech with intention and degrees of confidence, enabling better inference of meaning and social processes [7]. Audio is more effective than text in group collaboration because of its ability to promote higher level and more social discourse [7].

In RadioActive, it is possible to browse in a persistent archive any message created since the inception of the chat space. The archive permits out-of-order exchanges, where one can reply to any message across time in a multi-threaded fashion. This allows old conversations to re-emerge, with or without the involvement of the original participants.

Browsing text archives is a familiar problem, but designing for audio archives involves new challenges. RadioActive focuses on this key issue of ways to navigate an audio chat space. In order to know what to listen to next, users need an impression of the chat space to guide them. Audio is slow and serial, making it difficult to traverse large volumes of messages. It is also much more difficult for machines to parse and interpret. We approach this problem in RadioActive through visualization, where many of the many qualities of the chat space are presented visually rather

than orally. A message and its subsequent replies, known as a thread, are represented by visuals that vary in size, shape, and color, to represent the underlying metadata. Users interact with these visuals using a touch screen to play and respond to messages.

In this paper we examine related work, describe our approach, highlight a conceptual framework for navigation, discuss our evaluations, and provide suggestions for future research.

Related Work

While our exact goals have not been achieved in existing designs, shared elements of related work help us understand what to avoid and what to follow.

Nishimoto AVM and ScanMail

The most similar system to RadioActive in both goal and medium is Nishimoto's **A**synchronous **V**oice **M**eeting System [25]. AVM is a desktop-based asynchronous voice forum that uses automatic speech recognition to transcribe messages. It uses a newsgroup-like interface where users can quickly browse the transcriptions. In case of machine error, users may choose to playback the original audio of user selected portions, much like ScanMail [37]. AVM additionally features interruptions, or "barge-ins", as an audio analog of in-lined quotations. They theorized it would reduce the total amount of communication required to discuss the same topic.

In the user study, the barge-ins were not heavily used, and AVM with voice was perceived less similar to real-life than a text-based equivalent. This is because the users read the transcripts instead of using the audio. Users preferred the usability of the textual BBS than

AVM, which likely has more to do with the respective interface designs than the quality of the mediums.

ScanMail is a visual desktop-based interface to voicemail. Like AVM, ScanMail uses ASR to transcribe text for reading speed and to support search. It is useful for users with high volumes of voicemail to treat it like email in desktop environments. However, ScanMail does not provide any mechanisms for reply, a critical function of a complete communications environment. Both solutions mask audio's problems through transcription, which is inappropriate for large volumes of text on mobiles.

We purposefully did not include in our design process the availability of automatic speech recognition. More than limiting scope, we believe it is more important to first maximize design-wise known accurate meta-data before misinterpreting speech.

Mobile audio chat approaches

TattleTrail [18] and Symphony [21] allow users to share a conversation space using mobile devices. Their goal is to enable small work groups to converse both synchronously and asynchronously in a linear sequential structure. A linear structure is a practical constraint on speech interfaces, as evidenced by the simplicity of the "catchup" function, which plays unheard messages chronologically. Sequential messaging structures also avoid the need for random-access browsing. However, the projects are relevant in their ability to move between asynchronous and synchronous. Such fluidity might be appropriate in future work.

Hyperspeech [2] and Hypervoice [26] were systems designed to navigate highly structured hyperlinked audio data on telephones. The works focused on manipulation of structured databases without a real computer. Given the power of current mobiles, we can free ourselves from many of the constraints under their systems.

Browsing audio

Audio navigational techniques often focus on expanding a sense of place in an audio-space. Espaces2 [29] is an attempt to provide an audio-only environment using an "acoustic bubble" metaphor. Users browse hyperlinked audio data using 3D spatialized audio and audio icons. Spatialization and similar techniques such as earcons [6], Sonic Browsing [13], Ambient Soundscape [17], and Audio Hallway [30], and as well as parallel browsing techniques [1, 13] attempt to multiplex content in the audio domain. Such techniques incur large cognitive loads, which dangerously starves our attention for the outside world. Furthermore, it would be difficult to physically adapt these techniques for mobiles, given they often require precise speaker placement and low background noise.

Although structural data has been conveyed using sonification techniques [17, 24], it is restricted in practice to one dimension of expression. Because the mappings are unnatural, adding any more dimensions loses meaning in perception. It is also difficult to separate or estimate data points given a large simultaneous spread in any number of dimensions. In addition to perception problems, the artificial mappings are difficult to learn.

Speech interfaces rely on the user to maintain state in working memory, which inevitably will be forgotten. Refreshing state can be as costly as the initial input. Even with the best simultaneous presentation algorithms, speech-based navigation is limited in the complexity and size of an audio space it can successfully convey. This can only be solved with a visual representation.

Audio Skimming

Time-compression techniques, such as Speech Skimmer [3], shorten listening time by applying DSP to a sound clip. The most common algorithms speed playback whilst maintaining pitch using techniques such as Speech Skimmer's SOLA. New techniques have surfaced [32], which are likely to be appropriate when lightly applied.

Design Theory

To guide us in the process of designing "search" for our chat spaces, it is useful to think within a small set of useful domains. In "Running Out of Space: Models of Information Navigation" [11], Dourish suggests there exists at least three such navigational domains to manage large bodies of information: social, semantic, and spatial.

Social

Dourish defines social navigation as occurring when "movement from one item to another is provoked as an artifact of the activity of another or a group of others". Examples include exchange by word of mouth and collaborative filtering. In RadioActive, we can think about the networks of the people who author and view, in addition to the physical contexts of the user.

Semantic

General semantic navigation is accomplished with user-provided or computed metadata, and through hyper-linking the results of feature extraction. In our application, semantic features might come from the user (i.e. reply to this message), the machine (i.e. NLP shows these messages are on the same topic), or another inherited context as interpreted by machine or user (i.e. chat space for a museum piece).

Spatial

Dourish differentiates between information that is "inherently spatial" and mapping a semantic relationship onto a spatial arrangement. The latter is "problematic when semantic and social navigation are seen only to take place as a result of spatial organisation." Thus our chat space might arrange messages relative to where people live on a map if we wish to label our axes. But what about conversations where that's not appropriate? As discussed in the following section, we use space to project the message hierarchy (our semantic relationship) onto visual elements (non-intersecting graph edges) rather than along a coordinate system. Further, an energy-minimizing graph layout assures edges do not cross by loosening the restriction that distance is meaningful.

Design

RadioActive exists to support on-demand social conversation amongst groups. Our principal strategy to support the medium is visualization. The visual channel enables parallel presentation, a significant advantage over audio. However, creating an effective visual interface for mobile devices is difficult. Visually representing tens or even hundreds of messages is not

trivial on small screens, and mapping the underlying metadata to visual cues has few real-world metaphors to draw from. Our goal is to use a clear compact representation of the audio chat space with a zoomable interface.

The visual abstraction represents the structure of the persistent chat space. Keeping navigational state in the visual domain has many strategic advantages over a speech interface approach. Speech interfaces require users to navigate blindly, projecting state transitorily and sequentially through the audio channel. Mentally retaining state is at odds with participating in the actual conversation. Instead, the visual representation retains the navigational state for the user.

Effective use of the visual domain also exploits additional computational abilities of users, as "spatial arrangements simply choice, spatial arrangements simplify perception, and spatial dynamics simplify internal computation" [20]. We have chosen not only to exploit spatial arrangements, but also other visual features like color that move useful affordance cues to the visual domain to simplify choice [19].

The visual representation conveys many dimensions of structure and metadata simultaneously to informatively guide the user through the chat space. Because we can simultaneously express more qualities of the space than possible in audio, users are freed to interactively participate in non-linear multi-threaded conversations.

Interface overview

We start with an asynchronous chat space structurally modeled after Usenet, a text-based large-scale group chat system. We can't use a similar interface due to

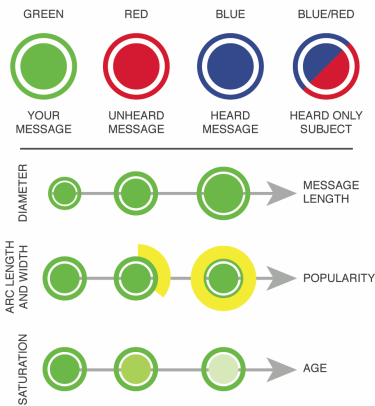


Figure 1. Circles visualize meta-data associated with a given post.



Figure 2. Three chat spaces with the same structure are visualized. Color effectively segments chat spaces by utility. Chat spaces with few unheard messages (left) are not as attention grabbing as a space with mostly unheard messages (middle). When only specific branches have been heard (right), color and layout-based clustering logically segments the space. Concentric rings help distinguish the first message, and function as navigational aid.

the physical and contextual constraints of mobiles: screen size, user's attention, ambient noise, etc. However, Usenet and email's interaction strategies to navigate messages are commonly understood. In order to shrink the design, we seek to preserve existing navigational strategies while limiting what we present.

At the heart of the visualization is the unique representation of a group of related messages. Each message is represented as a colorful circle. Variations in size, shape, and color represent the underlying properties of each message (Figure 1). These variations meaningfully differentiate posts through gestalt (Figure 2). The choice of dimensions of meta-data to present came from choosing those that best balanced germane information and visual expressibility. Figure 3 shows a first-pass list of potentially useful dimensions from which we picked, starting from the most significant in email and newsgroups.

When a user replies to an existing message, the new representative circle is placed near the parent post and connected with a gray line. This results in a graph-like representation for a group of related messages to depict multi-threaded hierarchy. Users navigate by some-what blindly following conversation paths, listening to each message by touch (we are designing for a touch screen—a growing trend in mobiles). Sliding one's finger across the messages constitutes one method of skimming audio, but to reduce listening time to understand gist we have to add an additional structure for long posts: subjects. Upon creation of a long message (longer than seven seconds), we ask the human authors to summarize their posts after creation. Thus gliding our finger causes only the subject to play. Shorter posts only possess a body and are already

quick to skim. Touching only one post will play its message. During playback, a button appears to allow the user to respond to the post currently playing. This takes them to a new screen to facilitate message creation, confirmation, and eventually insertion.

After playback completes, a dynamic playlist is optionally constructed to automatically choose the next child using depth-first-like search. This lets the user put the device in their pocket and passively listen to the chat, hence the radio in RadioActive. The heuristic is informed by the user's listening history and the properties of the children. The resulting playlist is projected into the visual structure, available for manual manipulation (Figure 4).

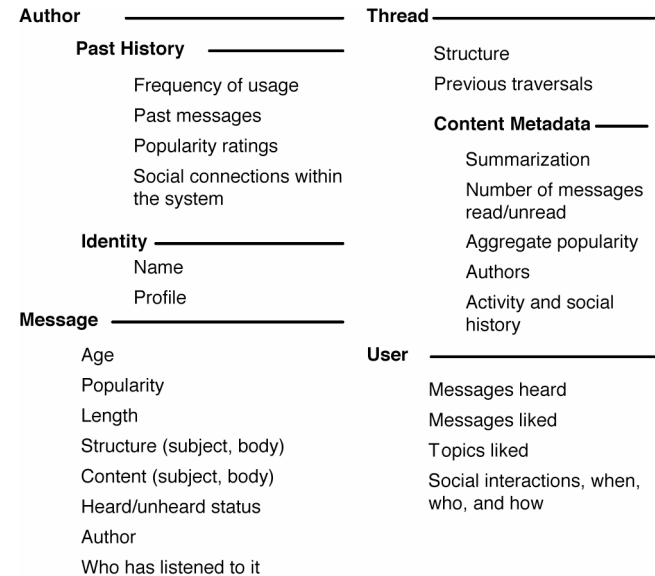


Figure 3. A first-pass approximation of the important dimensions from Usenet and email

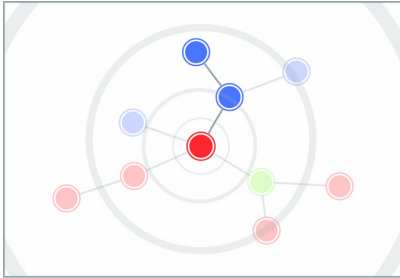


Figure 4. Because rings around messages and the lines are fully saturated normally, transparency works well to show and manipulate the current playlist.

2D Space

We can only fit so much on the screen at once, so not all of the conversation is shown in our infinite 2D plane. Starting a drag in empty space results allows translation of the conversation space, much like Google Maps. Arrow keys similarly support translation. We also implemented zooming [5], supported by buttons on the side of the device as well as in the interface (Figure 5).

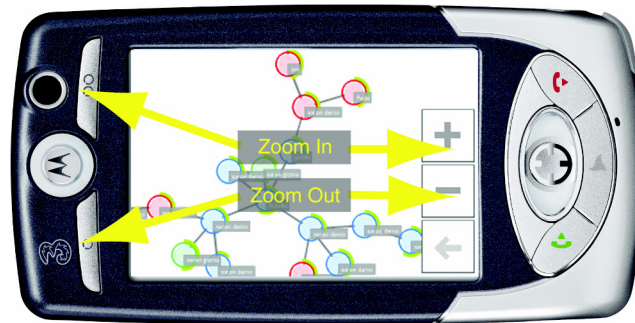


Figure 5. Zooming is supported by buttons in software and in hardware on our target platform, the Motorola A1000.

Originally we thought to keep all threads inside an infinite 2D plane, allowing the human operators to place threads next to other related threads. Aside from being confusing at first, separating persistent messages in space by topic is actually a poor strategy when the set of topics is large. Humans lack the metaphors to usefully project a multi-dimensional ontology onto a 2D plane. The progression of ideas in space is fundamentally limited by the non-linear and multi-dimensional relationships of the real world. Clusters eventually occlude their neighbors, impeding navigation. Furthermore, going left versus right only

has meaning along the related arbitrary dimension for a specific group of messages.

On a small screen, it is costly to traverse across threads with extraneous detail in between. We can remove these problems by dedicating a separate 2D plane for each thread, called the message-level view. A separate screen called the Inbox is used to navigate across threads (Figure 6).

Layout

We layout our messages according to an energy-minimizing algorithm to minimize the distance between each message in all directions to better fill the canvas. This algorithm has several advantages for mobile social data: 1) it emphasizes clustering, 2) it minimizes the display size constraint, and 3) the resulting layout has a more organic aesthetic. We further optimize the display by biasing the layout to fit within a specific *width:height* ratio. Thus when the user completely zooms out, the complete graph will exactly fit within the bounds of the screen.

Energy-minimizing layouts also help navigation. Spatial distance from the root post gives an easy approximation of reply-level depth, something that is difficult to estimate with Venolia's method of thread visualization [35]. Further, by spreading possible routes away from each other we simplify choice: unrelated messages are less likely to cluster.

Popularity

The most helpful information to choose the next post to listen to is to know about the content, from which we can assess message value. However, because everything is in audio, it is difficult to present such

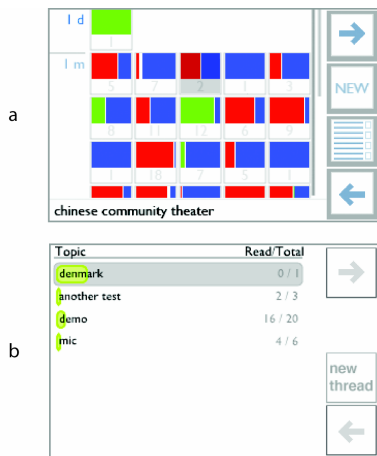


Figure 6. The original inbox (a) simultaneously shows many threads, providing a quick graphic feedback on the user's history within the space by showing proportions of read/unread messages. User studies revealed it was more opaque in usability than the redesigned interface (b), which only showed the name of the topic with its popularity behind it in green.

information. Another way of assessing message value given enough usage is popularity. Unlike text, by streaming the sequential message we know how many people have listened to what percent of a message of n minutes. The more people listen to longer messages, the more worthwhile it is. These parameters are used to derive popularity without requiring any extra effort from the user, unlike Slashdot's moderation system. We calculate a value range where most messages fall between 0-100, but they can go higher. Popularity is calculated with linear weighting according to the following formula:

$$popularity = \frac{numListensForThisPost}{20 \cdot globalAvgListensPerPost} + (70 \cdot avgHeardPercentForThisPost)$$

We visualize popularity with a partial ring around a post, increasing its thickness where popularity exceeds 100.

Evaluation

Without an extremely large user base and a long-term study, we cannot evaluate the true potential of group audio messaging on mobiles. However, we have performed several evaluations and iterations of our designs using small groups for feedback on our interface.

Initial Evaluation

The first version of RadioActive used the entire desktop as an initial testing ground for the design. It was spatially grounded, requiring users to start new threads at arbitrary points in the space. Threads were surrounded by boxes to help visually group them (Figure 7). The initial test comprised of eight males and two females. They were given a short explanation of how the chat space worked and were then asked to

join in the current discussion. They could use it as long as they liked, averaging approximately 3 minutes. This was the amount of time it took to browse the small chat space and leave a reply. Afterwards they participated in an open-ended interview.

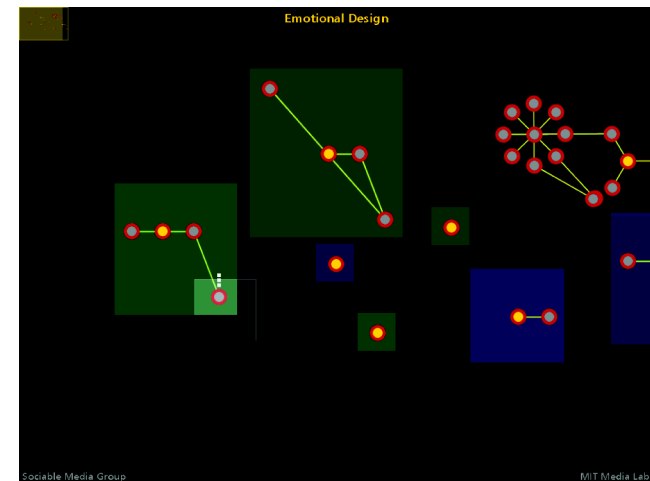


Figure 7. The first RadioActive iteration used a spatially-grounded concept, allowing multiple threads to be presented simultaneously on an infinite 2D plane.

All of the subjects found the concept of asynchronous audio intriguing, imagining scenarios ranging from collaborative work to blog activities. One said it was "the next obvious step from message boards". Nine of the subjects found the visual representation and its associated navigation intuitive, with comments ranging from "I can't see how else you'd do it" to "once you get used to it, it makes sense". Three subjects were concerned with how the design might translate to mobile usage where users might need to navigate

without looking at the screen, such as in an urban context. However, the most contentious subject was browsing. The results were split between users who wanted one unique summary per message and those who preferred a single summary for the entire thread. Eight of the subjects thought text-based summaries or tags would help in understanding the chat space. Requiring the user to record a short subject regardless of message length proved to be confusing for the users, as many left short messages and did not feel the need for a separate recording. They often repeated the same message over again, assuming it was an error. For the longer messages, the recorded subjects were often of poor descriptive quality, or summarized a viewpoint as "I disagree".

It is possible that the users who preferred a separate subject for the thread did so because of the spatial organization of the chat space. As previously stated, this study illustrated how it is confusing to place a thread seemingly arbitrarily in 2D space. In future iterations where threads were not spatially related, one would have a different starting place before launching a thread. Such a view is different from seeing a collection of messages linked with no indication of direction. However, it was felt that both views had merit and a balance could be struck for future iterations. Future designs still retained a unique subject for each message, but only when the message itself was of short duration (under ten seconds). Threads themselves were given a summary beyond the initial message in the form of text tags in a separate space.

However, the problem of acquiring usable summaries remained. This perhaps could be solved with better

interface prompts, but more likely, it comes from the user's lack of experience and the lack of netiquette for the medium. We hypothesize the more RadioActive is used, usage patterns will self-regulate the quality of a subject because the value and cost of good and bad subjects, respectively, is quite high. Unfortunately, such a claim requires a far longer-term user study.

Second Evaluation

With the results of the previous study, we iterated the design further which eventually became integrated into the eLens project. The eLens is a system that creates a unique chat space for a physical object by combining unique tag recognition [9] with a social network. Using the 3G data network, the client would identify the result from the scanned tag to the server, which then sends the chat space details and streams the appropriate audio. Low-latency properties of 3G (<300ms round-trip) make this interaction-style possible. The tags themselves are printable from a web page and can be affixed to a building using blue tac.

In order to test the functionality of the eLens, which includes RadioActive, we collaborated with the Catalonian government to obtain a test subject set of 9 male and 6 female volunteers aged 15-19 from a local technical high school in Manresa, Spain. Each student was loaned a Motorola A1000 with unlimited data and voice for a week. They were encouraged to use it on their own, in addition to the two-hour workshops with our team that took place during the test period. The workshops were constructed to have the test subjects play with the system under supervision to ensure they understood how the interface worked, observe their navigational strategies, answer any questions, and keep the system in working order.

The controlled class assignment was to act as a proxy for the local tourist office, placing tags on historic buildings and annotating them using RadioActive. It was hoped that their interaction would encourage alternative uses of tool, and discussion amongst each other. Instead they followed the directions too literally. The students simply placed the tag, scanned it in, named it, created a thread, and left a root message that was read verbatim from the official tourist pamphlet. They played back the message to make sure it was recorded, and moved to the next location on their list.

Even though the usage was unimaginative, we were still able to ethnographically observe how they used the design, resulting in useful critique of the design.

What worked

As a whole, the students were able to quickly learn, interpret, and use the interface. Just as was observed and reported in the previous informal interviews, users took at most 3-4 full cycles of scanning a tag, naming it, creating a thread, recording a message, and listening to it, to feel comfortable with the interface. Afterward, they effortlessly traversed chat spaces, carefully manipulating the 2D plane and scanning the messages. The students that were originally adverse to the new design liked it by the end of the week. Some enjoyed the integration of physicality and chat, in particular, "I can really see this as a possible future of communication. I think it's powerful that I can leave a message on a location. We could see being late and my friend knowing where I went." Another remarked on persistence, noting it functions "like an Internet forum, I like the idea of the persistent conversation, that messages would stay for one year or longer. I think it's

a great way to talk to lots of people and to meet new people who share my interests." In free-form interviews, each student said the interface was intuitive, and almost all liked using voice (13/15). One responded that voice carried too much identity, and the other preferred the qualities of SMS. They also expressed interest in adding videos or pictures to the chat space. The experience of voice versus SMS was described as "more personal" and "easier", although there were concerns over the speed and cost of the phone to function equivalently. Most of the students said they imagined it being used as a tourist guide--a product of the experiment itself--but several mentioned it would be useful to leave messages for friends in social situations, such as at cafes or bars. Several also thought of it functioning like Citysearch, where users could rate and talk about a business outside of it. The only complaint about the interface was the lack of author names directly on the post. The student wanted the ability to find messages from specific people. This request was fulfilled in the follow-up design.

Immediate Lessons Learned

RadioActive's visual design was found to have had two main problems: the thread-screen and subjects versus messages. The problems with the thread screen were discussed above, but ultimately came down to three classes: separation between object and action, occlusion of topic in favor of message count visualization, and a lack of understanding relevance in the message count visualization. The object/action problem was corrected in the next iteration by assigning a default action to the double click of each object. It is also the case that the thread visualization may not have correctly served the user's purpose because it gave visual priority to a user's history within

a thread. However, this preference can be explained by the goal of finding specific threads by summary rather than casually browsing, which was the expected goal. By moving away from this view completely, we achieve both goals in the new thread-level view by specifying summary, message count, and popularity together.

The users also did not understand the purpose and interaction of messages that were split into subject and body. When prompted for a summary, they often repeated the same message assuming the original did not go through. It is likely this behavior was due to a language problem, but it also may be explained by an artifact of the usage patterns. Since most messages were five to seven seconds, the extra prompt for the rare longer message was foreign and unexpected. Users then assumed there was a problem because when they touched the circle it only played the subject. This has been corrected by better visually presenting subject or body status during playback.

The original design required users to tap a message to hear its subject, and tap again during subject playback to hear the body. This mechanism was motivated by the preference of browsing subjects over content since the bodies were the most expensive bit. However, with no visual feedback or other indication the mechanism became opaque. The next iteration corrected this by incorporating the presence of a subject and body in the circle representation (Figure 1), text-based feedback to tell which part is playing, and changing the touch-based strategies to prefer bodies. It is now the case that only pressing an individual message will play the body, while dragging a finger across messages or holding a single message for 300ms will preview the corresponding

subject. Thus we use subjects as a power-user feature. Further investigations are necessary to understand the best strategy.

An unexpected observation was the use of the stylus. The interface was carefully designed to permit finger-based interaction on the touch screen, yet each time the students chose the stylus even after being told it was not necessary. The behavior may be rooted in the need for a stylus throughout the rest of the phone's software. Other explanations include the dedication of the task in terms of time allocation, low cognitive load requirements, and embracing of the touch screen for its novelty.

Another unforeseen consequence, but retrospectively obvious, was the problem of street noise when using speakerphones in urban situations. While RadioActive was originally envisioned using Bluetooth headsets, speakerphones provide a more accessible and cheaper design. Future work might consider DSP treatments to help remove excessive urban noise, if headsets are not available.

Future Work

Long-term studies are needed to evaluate the true potential of asynchronous persistent audio as a medium. Regardless, we have had many ideas in the course of our research to improve navigation within our current design. We'll start with some specific examples, and then discuss the more general recommendations that break our existing assumptions and supporting new conversation structures.

Improving our design

The most contentious feature is the separation of subject and message. Further investigation into both the usefulness and the presentation of subjects is necessary. We have already redesigned the interaction to default to playing messages when taping a message, and visual confirmation is the next logical step.

If well done, the user is likely to benefit from the ability to move around individual or clusters of messages. Spatial dynamics help people perform computation on objects more quickly than in their head for set manipulation tasks [20]. Users should be able to interactively drag, delete, sort, collapse, and expand [4] threads, messages, and topic clusters. The resulting arrangement of messages might be meaningful if shared, as only users can inject appropriate meaning into spatial arrangement; our energy-minimizing algorithms are ignorant of any context.

Relevant hyperlinks to information filters or alternative views can aid navigation. A useful starting set presented during message playback might consist of 1) a list of other message by the author, 2) other messages of similar topic, and 3) an interface that reflects upon other users who also listened to the message. Such views combined with back/forward buttons help the user non-linearly explore the conversation space.

Lastly, while we did not explore an interruption-based model similar to Nishimoto, it is worthwhile to consider improving such a system for automatic device-in-your-pocket playback.

Visualizing conversation structure

When we originally conceived of RadioActive, we used Usenet as a model structure. There would be many separate chat spaces, found after traversing through a topic hierarchy. In the topic there would be multiple threads, and each thread would be represented by our visual model. This model works for Usenet, but it doesn't necessarily match all contexts. By context, we are not referring to context-aware, but in-situ context. In-situ context is important when considering the application of mobile chat to physical location, as a user will enter the space with an existing mental model for the code of conduct based on the path to get there. One would not expect a tag on the outside of a bar to carry questions about math, nor would one expect a tag located on a McDonalds inside a Mall to carry architectural information.

Thus we must be flexible in our design to suit the in-situ context of a chat space. This was especially evident recently at a showing of RadioActive. While not enough was done to warrant its own evaluation section, the results are relevant for future thought. The MIT Museum asked us to give a live demo for a program featuring current graduate research. The museum curator took it upon himself to seed the chat spaces with questions and comments according to the exhibit it would augment. He acted as a moderator, creating the initial structure of the asynchronous conversation. To make sure that his comments did not get lost amongst a possible sea (and for practicality), we treated all anonymous museum-goers as one logged in user separate from the curator's account. Consequently, our existing color model differentiated between official and unofficial posts (Figure 8). In watching the users interact with the system, they clearly started with the

curator's messages and worked their way out, eventually adding their own on-topic post. This was a clear departure from usage in previous user studies, where the conversations were forced and the contexts artificial. We theorize the successful interaction was supported by a legible UI, content appropriate for the inherited social conduct of the museum, and an existing conversation structure that was easily discernible.

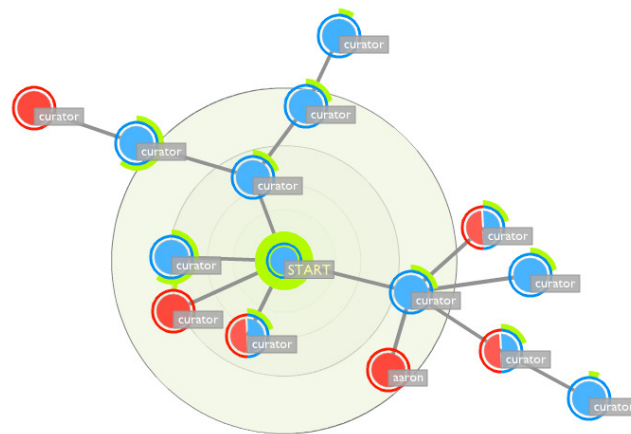


Figure 8. We reappropriated the design such that it extended the entire screen for MuseumActive, relabeled and reinforced the beginning message, and labeled the curator's messages.

Of course seeded content always helps, but more importantly the curator setup a conversation structure that matched in his mind the surrounding context. On a limited screen, we must offload to the visual domain as much audio-based information as we can. These structures can be put in the content of general communications mediums such as e-mail, but the cost of audio is too high to not make explicit as much as we can to speed navigation. This is difficult because each

separate context has its own appropriate structure. Wikipedia has one such structure appropriate for an Encyclopedia, where the shared "fact" is separate from surrounding meta-communication. Expert text-based web forums often augment the standard question-answer structure by highlighting the designated correct answer. When considering the navigational advantages of making explicit the right structure, it is clear that much research is needed in developing mobile-appropriate structures that work with what the contexts require. Of course the more flexible the structuring the better, as people need to be able to re-appropriate the chat spaces for unforeseeable needs.

Conclusion

Asynchronous persistent audio, an unexplored medium, might be the proper way to flexibly and loosely connect groups on mobile devices. Mobiles uniquely afford fuzzier and more distributed social connections by better integrating with the physical environment. However, the technology to create even a context-less persistent audio chat space did not previously exist. Our work has focused on the key problem of navigating asynchronous persistent audio under the practical constraints of a mobile device. We have discussed the possibilities of a novel visualized audio chat space, persistent group chat for mobiles, automated playback of multi-threaded chat spaces, using listening time to passively generate moderation values, and augmenting a physical location with a persistent chat. By creating, iterating, and validating a basic design, potential for the medium has been demonstrated but not proven. Long-term usage is needed to evaluate the full potential and characteristics of the medium.

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, W.I. Collapse-to-zoom: viewing web pages on small screen devices by interactively removing irrelevant content. *Proc. UIST 2004*, ACM Press (2004), 91-94.
- [5] Bederson B., and Holland J. Pad++: a zooming graphical interface for exploring alternate interface physics. *Proc. UIST 1994*, 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, and Kraut, R.E. Expressive richness: a comparison of speech and text as media for revision. *Proc CHI 1991*. ACM Press (1991), 21-26.
- [8] Connell, J. Mendelsohn, G., and Robins, R. "Effects of Communication Medium on Interpersonal Perceptions: Don't Hand Up on the Telephone Yet!", *Proceedings of GROUP '01*, Boulder, CO, (2001), pp. 117-124.
- [9] Costanza, E., Shelley, S. B., and Robinson, J., d-touch: a Consumer-Grade Tangible Interface Module and Musical Applications. *Proc. Designing for Society HCI2003*, Bath, UK, 8-12 (September 2003)
- [10] Donath, J., Karahalios, K., and Viegas, F. Visualizing conversation. *Proc. HIICS-32*, System Sciences, (1999)
- [11] Dourish, P. and Chalmers, M. Running out of space: models of information navigation. Short paper presented at HCI'94, Glasgow. (1994)
- [12] Ellson, J., Gansner, E.R., Koutsofios, L., North, S.C, and Woodhull, G. Graphviz and dynagraph - static and dynamic graph drawing tools. *Graph Drawing Software*. (2003)
- [13] Fernström, M., and Brazil, E. Sonic Browsing: An Auditory Tool for Multimedia Asset Management. *Proc ICAD '01*, Espoo, Finland, July 2001.
- [14] Fritz, G., Seifert, C., Luley, P., Paletta, L., and Almer, A. Mobile Vision for Ambient Learning in Urban Environments. *Proc. International Conference on Mobile Learning, MLEARN (2004)*
- [15] Hindus, D., Schmandt, C., and Horner, C. Capturing, structuring and representing ubiquitous audio. *ACM Trans. on Information Systems*, 11, 44 (Oct. 1993), 376-400.
- [16] Kerr, B.J. Thread arcs: An email thread visualization. *Proc Infovis '03 (2003)*.
- [17] Kilander, F., and Loennqvist, P. A Whisper in the Woods-An Ambient Soundscape for Peripheral Awareness of Remote Processes. *Proc ICAD '02*.
- [18] Kim J.S. TattleTrail: An Archiving Voice Chat System for Mobile Users Over Internet Protocol. MIT M.Eng Thesis, May 2002.
- [19] Kirsh, D. Metacognition, Distributed Cognition and Visual Design. *Cognition, Education and Communication Technology*. Lawrence Erlbaum, 2004
- [20] Kirsh, D. The intelligent use of space. *Journal of Artificial Intelligence*, (1995) 73(1-2), pp. 3168.
- [21] Lakshminpathy, V., and Schmandt, C. SimPhony, A Voice Communication Tool for Distributed Workgroups. *BT Technology Journal*, 22, 4 (Oct. 2004), 52 - 61
- [22] Lam D., Rohall, S.L., Schmandt, C., and Stern, M.K. Exploiting E-mail Structure to Improve Summarization. *Proc. CSCW 2002*, ACM Press (2002)

- [23] McGookin, D., and Brewster, S. An Investigation into the Identification of Concurrently Presented Earcons. Proc. ICAD, Boston, MA, July 6 - 2003, 42-46.
- [24] Mynatt, E., Back, M., Want, R., and Federick R., Audio Aura: Light-Weight Audio Augmented Reality. Proc UbiComp '97. ACM Press (1997).
- [25] Nishimoto T., Yuki, H., Kawahara, T., Araki, T., and Niimi, Y. Design and evaluation of the asynchronous voice meeting system AVM. Systems and computers in Japan, Wiley Periodicals (2002), 33, 11, 61-69.
- [26] Resnick P. HyperVoice: A Phone-Based CSCW Platform. Proc. CSCW 1992, ACM Press (1992) 218-225.
- [27] Rodenstein, R., and Donath, J. Talking in Circles: Designing a Spatially-Grounded Audio Conferencing Environment. Proc CHI 2000. ACM Press (2000) 81-88.
- [28] Rohall S., Gruen D., Moody P., and Kellerman S. Email Visualizations to Aid Communications. IEEE Symposium on Information Visualization, 2001.
- [29] Sawhney, N., and Murphy A. ESPACE 2: An Experimental HyperAudio Environment. Proc CHI 1996. ACM Pres (1996), p. 105-106.
- [30] Schmandt, C. Audio Hallway: A Virtual Acoustic Environment for Browsing. Proc UIST 1998. ACM Press (1998) 163-170.
- [31] Schmandt, C., Kim, J., Lee, K., Vallejo, G., and Ackerman, M. Mediated Voice Communication via Mobile IP. Proc. UIST 2001. ACM Press (2001), 141-150.
- [32] Tucker, S., and Whittaker, S. Time is of the Essence: An Evaluation of Temporal Compression Algorithms. In Proceedings of CHI06 Conference on Human Factors in Computing Systems, New York: ACM Press (2006).
- [33] Tufte, E. Envisioning Information. Graphics Press (May 1990)
- [34] Venolia, G. D., Dabbish, L., Cadiz J., and Gupta, A. Supporting email workflow. Technical Report MSR-TR-2001-88, Microsoft Research, Collaboration & Multimedia Group, One Microsoft Way, Redmond, WA 98052 USA, December 2001.
- [35] Venolia, G., and Neustaedter., C. Understanding sequence and reply relationships within email conversations: A mixed-model visualization. MSR-TR-2002-102. (2002)
- [36] Whittaker, S., and Amento, B. Semantic Speech Editing. Proceedings of Conference on Computer Human Interaction, 527-534, New York, ACM Press. (2004).
- [37] 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 2002, ACM Press (2002), 275-282.