# Audio Hallway:
# a Virtual Acoustic Environment for Browsing

**Chris Schmandt**
M.I.T. Media Laboratory
geek@media.mit.edu

## ABSTRACT

This paper describes the Audio Hallway, a virtual acoustic environment for browsing collections of related audio files. The user travels up and down the Hallway by head motion, passing "rooms" alternately on the left and right sides. Emanating from each room is an auditory collage of "braided audio" which acoustically indicates the contents of the room. Each room represents a broadcast radio news story, and the contents are a collection of individual "sound bites" or actualities related to that story. Upon entering a room, the individual sounds comprising that story are arrayed spatially in front of the listener, with auditory focus controlled by head rotation. The main design challenge for the Audio Hallway is adequately controlling the auditory interface to position sounds so that spatial memory can facilitate navigation and recall in the absence of visual cues.

Keywords: digitized speech, virtual environments, spatial audio, auditory user interface.

## INTRODUCTION

The Audio Hallway is a virtual acoustic environment designed to support browsing of certain classes of digital audio recordings. It uses simultaneous, spatialized audio presentation techniques with navigation playback selection supported completely by the position and motion of the user's head. The user interaction techniques it explores are most appropriate for browsing a large number of audio files which cluster together into logically related groups.

Consistent with the clustering of the audio data, the Audio Hallway supports two levels of representation and interaction with that data. At the top levels, the Hallway is a virtual acoustic environment in which the listener travels up and down a hallway, with clustered sounds audible behind "doors" which are passed on the left and right sides. The sound emanating from the doors is an auditory collage of "braided audio", an acoustic mix of all the sounds in that cluster, processed so each is intelligible in turn. When the user enters a room, by tilting his or her head in its direction while nearby, the individual sounds inside the room are presented as a spatial array in an arc about the listener's head, with the most dominant sounds being those toward which the user faces. This allows browsing of the individual sounds comprising the cluster.

The Audio Hallway is an example of a computationally generated synthetic listening environment. We are now able to create and interact with artificial listening conditions, and research suggests that such environments may contribute to more productive audition in some situations. One task for such an auditory interface is browsing of the large quantity of digital audio material now becoming available on the Internet and from other sources such as personal digital recorders and voice mail systems. Synthetic listening environments can take advantage of various techniques including random access, time scale modification (rapid playback), simultaneous listening, spatialized audio, and digital signal processing, enhancing, and mixing.

Browsing in a visual user interface takes advantage of the fact that a variety of artifacts may be placed in the visual field and the user can rapidly scan them, as well as use peripheral vision to obtain some sense of the objects not in immediate visual focus. Presentation of sound files for browsing is made difficult by the serial and transitory nature of audio. As opposed to the graphical display, sound must be playing in order to exist; there is no auditory "still" or thumbnail which can be extracted from a sound file as from a movie or image file.

Audio Hallway is one of a series of projects exploiting computer generated simultaneous and spatialized audio presentation. Simultaneous listening, involving attending to multiple sources of audio at the same time, has been studied for some

*UIST '98. San Francisco, CA*
© 1998 ACM 0-58113-034-1/98/11... $5.00

time. Cherry [3] performed early experimental work on the “cocktail party effect”, i.e., that we appear to be able to switch our attention and focus it on one of several several simultaneously presented sounds. The *selective attention* ability is predicated on our capacity to cognitively group the arriving sounds into *streams*. Although the sounds arrive at our ears simultaneously and are mixed as they travel through the air and the components of the ear, under many conditions we perceive them in terms of their sources, e.g., a noisy fan above, a woman speaking on our right, a car passing on the street. A number of cues contribute to the streaming effect, including frequency and temporal characteristics of the sound and the direction from which it arrives at our ears [2].

Computational techniques allow for spatial presentation of sound information by processing the source audio into left ear and right ear signals through the Head Related Transfer Function (HRTF), which introduces the inter-aural phase, frequency, and amplitude differences as heard by each ear [13, 1]. Such spatialized audio is presented over headphones, and was originally developed for virtual reality applications such as in [4].

In a Bellcore project, non-spatial audio was used in an interesting acoustic environment for audio “window systems”, in which sounds of varying levels of priority or importance were enhanced or muffled [10]. Spatialized audio was used in non-VR situations in AudioStreamer [9], which presented three simultaneous audio streams of news stories to the left, right, and in front of the listener, and enhanced selective attention when the user leaned toward one of them. It was also used in SoundScape [7] in which multiple versions of the same sound orbited about the listener, with rotational angle being a function of time offset into the sound, and with the user able to reach up into the orbiting sound and re-position it for playback.

The Audio Hallway attempts to build a more extensive spatial listening environment, consisting of a much greater number of audio sources, and allowing the user to move about within this environment. As opposed to the conventional virtual reality applications, however, the Hallway uses no visual display. A large part of the reason for avoiding a display was the desire to exploit the fact that sound can be attended to while the hands and eyes are busy performing other tasks. This might allow auditory interfaces to be used in a variety of situations where conventional computing interfaces do not apply, such as in automobiles, while walking on the street, or exercising. Additionally, although we recognized from the beginning that a visual display would help listeners understand the hallway metaphor, we chose to focus our attention on the purely auditory aspects of the user interface, as a specific challenge.

A recurring theme in the design of each component of the Audio Hallway is the ability of the listener to build an adequate spatial model of the sounds to aid navigation and retrieval. Without any visual interface, it is difficult to successfully convey the spatial model, much less allow the listener to apply spatial memory to navigation and recall. It is highly likely that this problem is intrinsic in the auditory interface, or at least as spatial audio can currently be synthesized, and hence will be relevant to a wider range of spatially presented auditory interfaces.

The Audio Hallway is a fully operational design exploration. As it was built, it was continually if informally evaluated at several levels. Did the designers and authors find it effective, and was the design correctly implemented programmatically? Did other graduate students find the interface effective, in informal listening tests? Finally, what were the reactions of a continual stream of outside Lab sponsors who listened to the interface? These last sessions ranged from 5 minutes to half an hour, and the listeners had a wide range of technical expertise and familiarity with auditory user interface issues. Their reactions led to refinements to the interface and better valuation of what was possible.

## THE AUDIO DATA

The Audio Hallway is oriented towards browsing collections of small audio files. Each collection is a set of related audio files which can be logically grouped into a cluster, to be represented as a single entity at the top level of browsing. Once the user has selected a cluster, the individual files which make up that cluster can be scanned.

### Audio files from radio news

The audio data used in the Hallway consist of broadcast radio news actualities (“sound bites”), as used in the ABC Radio Network. The actualities are broadcast quality audio snippets ranging from 5 to 20 seconds long, and are distributed by ABC Radio to its affiliates every hour; affiliates use this material for composing their own newscasts with local anchors. An associated data feed provides text transcripts of the actualities and some summary information authored in the newsroom.

The audio is distributed digitally over a satellite network, received on our roof at a 10 meter dish, decoded from the proprietary digital format, and then presented as analog audio. This audio is fed into a Sun Sparcstation which uses the Unix *cron* command to trigger recording. Recording software watches for a cue tone, records an entire sequence of 10 to 20 actualities separated by voice cues, and then splits the recording into separate files, one per actuality. After the text transcript arrives over the satellite data feed, another program is triggered by *cron* to correlate transcripts and audio files, doing minor error checking in the process; sometimes an error is made at the transmitter, or, more likely, our decoding software improperly segments the actualities, resulting in a

```
IT’LL BE AWHILE BEFORE ANOTHER PAMELA ANDERSON SEX TAPE
IS AVAILABLE.  A FEDERAL JUDGE HAS, FOR THE TIME BEING,
AGAIN STOPPED A SEATTLE COMPANY FROM DISTRIBUTING A
HOMEMADE SEX VIDEO. ABC’S STEFFAN TUBBS HAS THE STORY
 FROM LOS ANGELES...  (ABC NEWS, LA.) :38

VERBATIM: ‘‘IT WAS RULED THE RELEASE OF THE TAPE WOULD
VIOLATE THE RIGHTS OF BOTH PAMELA ANDERSON OF BAYWATCH
FAME, AND HER THEN-BOYFRIEND, BRET MICHAELS OF THE BAND
’’POISON.’’ THE CELEBRITIES HAVE EACH FILED 90-MILLION
DOLLAR LAWSUITS AGAINST THE SEATTLE INTERNET
ENTERTAINMENT COMPANY TRYING TO RELEASE THE TAPE. THE
JUDGE RULED BOTH ANDERSON AND MICHAELS RIGHT TO PRIVACY
WOULD BE VIOLATED IF THE TAPE MADE IT TO THE INTERNET
OR ANYWHERE ELSE. STEFFAN  TUBBS, ABC NEWS, LOS ANGELES.’’


THE ALADDIN CASINO AND HOTEL IN LAS VEGAS IS NOW HISTORY
AS AN IMPLOSION BROUGHT IT TO THE GROUND MONDAY NIGHT.
FOR PRODUCTION PURPOSES HERE IS WHAT IT SOUND EDLIKE,
FIRST WITH THE EXPLOSIONS THAT BROUGHT IT DOWN FOLLOWED
BY THE FALL AND CHEERS FROM A CROWD OF ONLOOKERS.  :27

VERBATIM: ’’NATURAL SOUND’’


HISTORY OF SORTS WAS MADE MONDAY NIGHT ON THE INFORMATION
SUPERHIGHWAY. KOKO, THE FAMOUS PICTURE PAINTING GORILLA
AT THE SAN FRANCISCO ZOO THAT HAS LEARNED TO COMMUNICATE
THROUGH SIGN LANGUAGE, WENT ON-LINE, CHATTING TO FOLKS
THROUGH AMERICA ONLINE. KOKO WAS ASKED QUESTIONS IN AN
A-O-L ’’CHAT ROOM’’ AND THOSE QUESTIONS WERE SIGNED TO
HER BY HER CARETAKER, DOCTOR FRANCINE PATTERSON. :16

VERBATIM: ’’SOME PEOPLE HAVE CRITICIZED THAT THIS WAS A
PUBLICITY STUNT BUT I MEAN HERE IS A GORILLA WHO SIGNS
AND USES THE COMPUTER AND CERTAINLY A CHANCE TO BE ABLE
TO ASK A GORILLA A QUESTION AND ACTUALLY GET AN ANSWER,
SOME PEOPLE DID GET AN ANSWER, THAT IS HISTORIC SO IT
SHOULD BE VALUED FOR THAT’’
```

Figure 1: Some examples of the description and transcriptions available on the data feed.

mismatch between the number of text and audio files.

The actualities contain a wide variety of audio data. Some are story summaries by ABC reporters. Most are recordings of the people making the news, e.g. excerpts from speeches, eyewitness or interviewee commentary, and local ambient noise such as the cheering of baseball fans at the World Series or the sounds of an explosion when a famous hotel is demolished in Las Vegas. During the course of the day stories unfold and different actualities related to each story will appear at different hours; stories fade as new stories break during the morning and afternoon. We needed to sort the actualities by story, so each story would appear in its own room in the Audio Hallway.

### Clustering the audio files

The textual descriptions and transcripts are used to cluster the sets of actualities into clusters based on story topic. If the information were available in an orderly and consistent manner, we might have been able to use text-based techniques to find a good story summary to represent each story succinctly. As this was not the case, we took advantage of the transcripts merely to group actualities into stories and determine the number of stories. Summarization is provided via the braided audio technique described below.

We use the SMART text retrieval engine [11] to correlate the stories. SMART is ordinarily used to match a query text against one or more documents in a previously indexed corpus of documents. We use it in a slightly different manner to cluster the actualities, where each cluster ideally represents one story. SMART uses statistical methods based on the words in the supplied text to map that text into a vector in a multi-dimensional "term space". We map each actuality into this space, and then use minimum error metrics to cluster the stories.

The main limitation on this technique is the short text lengths of the actualities; longer texts are more reliably correlated. As the texts are unfortunately transmitted entirely as upper case, we cannot take advantage of the occurrences of proper nouns to aid categorization. Still, SMART gets most of the stories most of the time, and its errors are not unreasonable. For example, one day actualities on a minor earthquake in eastern California, possible arson of a synagogue in Los Angeles, and difficulty finding parking at a show at the Getty Museum were all grouped into one story (disasters in California).

## BRAIDED AUDIO

The Audio Hallway supports selection at two levels of aggregation: choosing among stories and listening to individual sounds which make up a story. To enable story selection, we needed a means of representing each story as an auditory entity, much like a visual "thumbnail" might represent a still or motion image of higher resolution. Although we had text transcripts of the actualities and could possibly have used textual analysis to try to select the most representative summary, this is itself difficult and we wanted to explore more general and auditory techniques. We developed a form of auditory collage, Braided Audio, as an auditory representation.

A visual collage intermingles visible segments from multiple images; an auditory collage must incorporate the temporal nature of sound as well. Braided Audio mixes all the sounds to be represented, but sequentially amplifies each so it momentarily dominates the mix and is intelligible. We developed the “rope” metaphor after which the technique is named. A rope composed of multiple braided strands intermingles each strand, but when the rope is inspected from end to end, an individual strand is closer and most visible to the viewer at each instant.

The amplitude with which each sound is played varies sinusoidally, with overlap (mixing) over the boundaries of the adjacent sounds, as shown in figure 2. As the previous sound fades, the current sound rapidly grows louder, plays with rel-

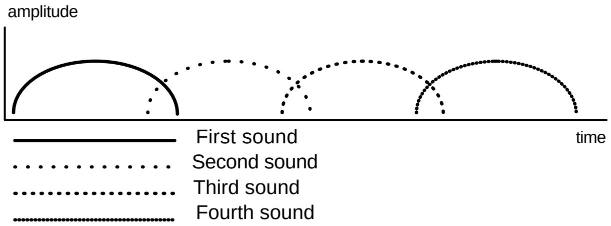

Figure 2: Braided audio, showing the amplitude and durations of segments of the sounds mixed into the braid.

atively steady volume for about three seconds, and then fades down into the next sound. Using a sinusoidal amplitude envelope instead of a step function avoids jarring discontinuities and aids in perceiving the braided audio stream as a single, internally related acoustic entity. Sudden acoustic transitions draw attention to themselves [5]. It is important that the sometimes disparate components of the audio braid be perceived as a single whole, a cluster, which can be ascribed a particular location relative to other clusters; this is an aspect of our larger theme of spatial recall. It is also confusing to listen to multiple rapidly changing sounds simultaneously, especially when all are in motion together; smooth braiding helps compensate for the changes.

Note that braiding is generated without knowledge of the acoustic content of its constituent sounds. We cannot select the most semantically salient portion of a sound, although if we could this would enable a more direct representation of its contents. Fortunately the news actualities are usually carefully edited at the ABC newsroom in New York; this makes them succinct and acoustically "dense". If they consisted of less edited speech, it might prove useful to select phrase boundaries (as was done, for example, in Stifelman's Audio Notebook [12]) for emphasis in the braiding.

The braided audio is meant to convey the general topic of a story and some of the level of its emotional content. Is it an exciting, breaking story? A major disaster? Can it be recognized as another in a series of stories unfolding over days, such as yet another sex scandal in Washington or financial news? Although performance is variable depending on the content of the actualities themselves, braiding has supplied a generally adequate representation and listeners quickly grasp the general topic of the story. That this would be true was not obvious before we attempted it. The text associated with each actuality includes a summary description and an actual transcript, and this combination facilitates text-based clustering. But the listener hears only the actual sound, and each sound is not necessarily clearly associated with a topic. Crowd noise, cheering, explosions, or eyewitnesses accounts full of adjectives but little content, as well as short direct quotes, tend to be ambiguous. The braiding works because among the intermingled sounds enough of them are distinctive, when the braid is heard for more than about 7 seconds..

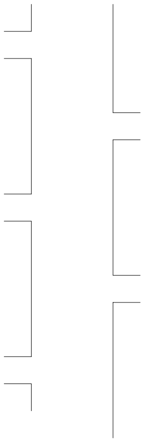

Figure 3: The Audio Hallway. The user travels down the center of the Hallway, passing open doors from which sound emanates.

## AUDIO HALLWAY

The braided audio is a single acoustic representation of a cluster of related audio files, in this case, related to a single story. These story representations are used at the top level of the user interaction to provide a means of browsing and selecting between stories, based on the metaphor of a physical hallway. The hallway is a virtual acoustic environment; spatialized audio is presented over headphones. The user moves up and down the hallway by tilting his/her head forwards or backwards; velocity is proportional to the degree of head motion.

Situated along the hallway are rooms, alternating on the left and right sides (figure 3). Each room is a distinct news story, and is represented by the braided audio from that story. The hallway wraps around, such that going forward beyond the last story restarts with the first story.

The auditory sensation of motion through the hallway is that of stories sweeping by alternating on the left and right sides. As a story approaches in front, its volume grows until it is quite loud when it is situated right next to the appropriate ear. If the user continues down the hall, that story fades to the rear and a new story appears on the other side. Up to three stories can be heard simultaneously; the closest story dominates one ear and the next two stories, guaranteed to be on the other side of the hallway, are heard in front of and behind the user on the other side of the hall.

### Perceptual effectiveness

Once users understood the hallway metaphor, they could generally navigate effectively. In the initial version of the system, head tilt was mapped to acceleration; this made it very difficult to stop moving and resulted in significant overshoot when the user decided a door just being passed was interesting. Subsequently, tilt was mapped to velocity, which proved more effective. When traveling at high speeds, however, overshoot remained a problem, and users sometimes found it difficult to locate a room recently passed. This became a problem only at speeds resulting in passing a room every two or three seconds or less. Otherwise, users could generally construct a mental map of the various stories and find one after browsing up and down the hallway.

At comfortable travel speeds, the braided audio technique usually provided enough of a glimpse into the story's contents to be useful in selecting which rooms to enter and for hearing the whole story. This was definitely enhanced by hearing a room approach in the distance for some time while moving towards it. If a room was passed quite quickly, however, only one or two of the braids might be heard, which often failed to adequately convey its content.

When attending to the braided audio from a nearby doorway the more distant sounds tended to be distracting and interfere with auditory attention. Although the interfering rooms were considerably more distant linearly, they may not have been adequately separated in terms of angular azimuth about the head. Rather than attempt to change the hallway geometry, we compensated by artificially increasing the amplitude of a room when it is very close. Acoustically, this changes our model from one in which the sound is generated right at the jam of the virtual door to one in which the sound comes from deeper inside the room and hence radiates in a more directional manner.

The main difficulty noted with listeners' experience was their inability to adequately resolve the left/right locations of rooms in the hallway. Sound localization by HRTF, at least with generic head models, is not as accurate as we would like. Several factors contribute to this problem. Foremost, our audio data was sampled at 8kHz, giving an effective audio bandwidth of less than 4kHz, resulting in loss of some acoustic information useful for localization. Using higher bandwidth audio cues resulted in slightly more accurate user perception.

A common complaint was poor lateral localization as a doorway was passed; the sound was sometimes described as passing from left to right (or vice versa) through the listener's head. If we widened the hallway, we could gain more effective lateral localization, as the audio sources would be rendered further apart. But this would render all the rooms further away and hence of similar amplitudes, and so the nearby room would require an even additional gain boost to dominate the spatial mix. Instead we distorted the hallway, spreading the walls and making it wider in the distance, as shown in figure 4. This means that as a doorway is passed, it sounds close to the head in the left-right direction, but rapidly moves further to the side once passed. This technique helped prevent confusion as to on which side of the hallway a door was, and thereby reinforced the hallway metaphor.

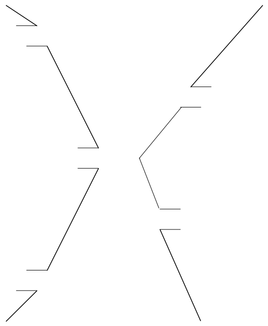

Figure 4: The modified hallway. With increasing distance from the head, doors are positioned further to the sides, to help maintain lateral acoustic discrimination.

When users better understood the hallway metaphor, they were more easily able to experience the desired spatial audio configuration. We found that a short description of the experience, using terms including "hallway" and "doors", and explaining the navigational model and its link to head motion resulted in significantly more satisfying listening, with less confusion as to the direction in which sounds were moving. Experience to the hallway was at best mixed, however, with some listeners not becoming comfortable during the course of approximately five minutes of listening and interacting with the system. Listeners with substantially more experience seemed to be able to navigate effectively and claimed to perceive the hallway as intended, but they were all in some ways associated with the research and hence do not demonstrate convincing usability.

## ROOMS

When the user tilts his or her head in the direction of a doorway while passing near it, entry is gained into the room. At this point the braided audio from the hallway goes silent, and the user is presented with all the audio, typically six to twenty individual files, arrayed in front of his or her head. Up to four of these files play simultaneously, based on the angle of rotation of the head in the ground plane. Sounds play simultaneously to facilitate browsing, as the listener can simply change attention between them without any need for additional head motion, and the fading in and out of neighboring sounds with head rotation again helps convey the spatial model of an ordered array of sounds about the user's head.

With the number of sounds to be presented, if they are spaced equally about the user's head, they end up being too close together to attend to separately. The greater the angular distance between audio sources the easier it is to separate them (and the longer it takes to switch attention between them). In order to spread out the playing files so they can

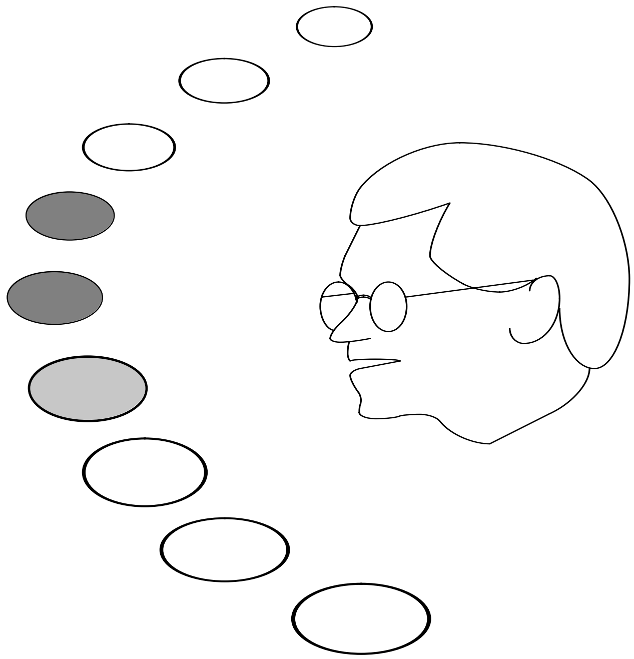

Figure 5: Inside a room, individual sound files are situated around and equidistant from the head, in a plane parallel to the ground.

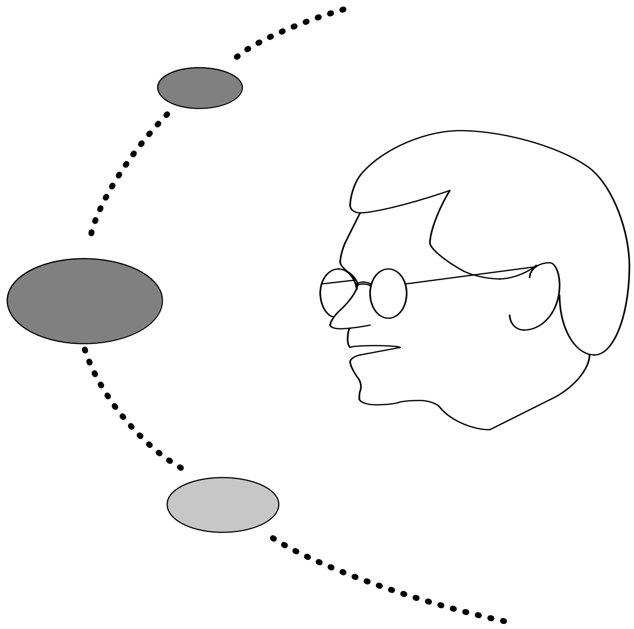

Figure 6: The apparent position of the three sounds highlighted in the previous figure, as rendered through the "lens". In this figure only, size of the sounds represents their playback amplitude.

be distinguished, they are distorted horizontally in a manner motivated by Furnas' work on Fisheye Lenses [6]. As the user's head rotates, a virtual lens moves across the audio sources, so that a small movement of the head results in a greater, distorted movement of the sources. This allows the user to bring a single sound into focus and also scan the sounds fairly rapidly. Figure 6 shows the three sounds highlighted in figure 5 as they would appear during playback, due to the distortion of the lens effect.

In order to maintain a sense of continuity across the distorting angular motion, it was important to obtain a smooth, sensitive, and continuous motion of the audio sources around the user's head as they moved through the audio field from side to side. Just as a lens magnifies what is immediately underneath it, we made the audio at the center of the lens proportionally louder in order to aid the selection process. The initial implementation of this interface imparted excessive latency between the head motion and the movement of the sounds; it proved critical to have these very tightly coupled to convey adequate control to the user.

Users generally had much less trouble navigating inside a room than in the hallway. In the room, sound position was more directly coupled with head position (rather than having velocity or acceleration mapped to head position as in the hallway). This makes it is easier to return to a known location to listen to the sound there. One limitation is that once a sound stops playing, because the head is turned away from it, it will restart from the beginning when it starts playing again, making scanning sometimes tedious, as the beginnings of sounds must be heard repeatedly. Nonetheless, this playback style was chosen to assist with the mental model of always having the same sound be in the same spatial location (taking account lens distortion of course). If playback resumes from the point where it previously stopped, it is harder to identify an audio stream as belonging to the audio file which was at that location earlier.

This is a case where the demands of building a mental spatial model actually interfered with efficient operation of the interface; a possible refinement would have been to provide a gesture to jump ahead in the sound, or for each sound to start with a signature (e.g. the first few seconds of the sound) and then segue to the last listening position of the sound. But the latter is difficult because at any moment up to four sounds are playing simultaneously and we don't know which was being attended to when any individual sound ceased playing.

## SYSTEM ARCHITECTURE

The Audio Hallway uses a rather sprawling distributed architecture, dictated by the abilities of various components as they became available. The main hallway application runs as a Unix program on a Sparcstation. Hallway audio is localized via a Crystal River Beachtron card on a PC, and room audio is generated on a separate PC running Intel's RSX audio spatialization software.

An audio server configuration had been previously developed for the AudioStreamer project. This server uses two stereo audio cards on the Sparcstation to generate four channels of mono audio output; the audio samples from separate sound files are merged into a stereo audio byte stream for playback. These four channels feed into a Crystal River Beachtron card on a 386 computer running DOS. The Beachtron is configured as a server and controlled over a serial port from the Sparcstation. The Beachtron produces stereo audio output, with the position and amplitudes of the four sound sources under control of the main program. Head position is gained from a Polhemus sensor mounted on the headphones; the Polhemus controller is plugged into the Beachtron PC and its data made available through the Beachtron server API.

The Unix audio server software had memory management difficulties when rapidly starting and stopping a large number of audio files, and rather than attempt to debug this legacy code the software author wrote a separate Java based server on an additional Intel box running Intel's RSX spatial audio package. This is active when the user is inside a room. The stereo audio outputs of the RSX audio and the Beachtron audio output are mixed by an analog audio mixer, and presented over headphones.

This architecture is of course sub-optimal and suitable only for a prototype experimental system, due to the large number of processes which need to be created and the time required for a cold restart.

## DISCUSSION AND FUTURE DIRECTIONS

The sound data used for the Audio Hallway had some unique characteristics, and we would certainly not expect to have transcripts generally available for grouping audio information. But techniques such as these might prove useful for audio corpora which can be characterized as clusters of somehow related sounds. One example could be an archive of voice mail messages or interactions over a telephone based customer service line, with messages or recordings grouped by the identity of the caller. Another example is an historical audio sound archive, where the clusters represent famous individuals or events, such as significant speeches and audio clips from the early days of the space program or the civil rights movement in the United States.

The mixing effect of the audio braiding seems to be effective and reasonable to attend to; it could certainly be improved by any means of making the transitions between sounds in the braid being correlated to their contents, even if only with pauses and phrase boundaries. The lens effect in individual rooms worked well once it was made responsive enough to shift rapidly when the user's head turned; initially the audio playback lagged and this spoiled the listening experience. These aspects of the Audio Hallway are worth continuing.

We were disappointed at how difficult it was to entice listeners into having the desired auditory experience while traveling through the hallway. A visual display would significantly improve the spatial mapping of the room sounds, but we do not necessarily claim that such an acoustically oriented technique would be the most effective if alternative display media were available. The evidence from this project suggests that it will be difficult to build virtual acoustic-only environments which are convincing with the combination of multiple audio sources and user motion within the auditory space. This echoes themes of our earlier work with spatial auditory presentation, although was more evident in the Audio Hallway because this project makes much more extensive use of navigation as position in a virtual acoustic world, position centered on and controlled by the listener.

Several aspects of auditory streaming indicate that at least some of this difficulty is inherent in our perceptual systems. The first is that visual cues are known to be strong components of our ability to separate audio into streams [2, 8]. Another is that the history of the sound is an additional streaming cue. Our braided audio sounds violate the history cue, because they shift every few seconds between sounds which, although semantically related, are acoustically different, coming from different talkers in different acoustic ambiance and background noise. Factors such as these must be factored into the design of future auditory virtual environments.

## ACKNOWLEDGMENTS

Brenden Maher wrote the Audio Hallway software as his MS thesis project. Bruce McHenry wrote the audio parser to separate the ABC actualities into separate audio files, Jonathan Sheena wrote the ABC data feed server, and Steven Lee wrote the story classification via SMART. Jordan Slott wrote the stereo mixing audio server. Atty Mullins and Minoru Kobayashi built some of the underlying software architectures during their thesis projects. Pat Turner assisted with production of this paper. As is so often the case, the anonymous reviewers contributed insightful comments which have hopefully been successfully incorporated into this paper.

We are deeply grateful to ABC News Radio for supplying us with the audio and data feeds and satellite decoding units for the sound recordings used in this project. This work was done under the support of the News in the Future research Consortium.

## REFERENCES

[1] D. Begault. *3D Sound for Virtual Reality and Multimedia*. Academic Press, 1994.

[2] A. Bregman. *Auditory Scene Analysis: the Perceptual*

*Organization of Sound*. MIT Press, Cambridge, MA, 1990.

[3] E. Cherry. Some experiments in the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25:975–979, 1953.

[4] S. Fisher E. Wenzel C. Coler and M. McGeevy. Virtual interface environment for workstations. *Proceedings of the Human Factors Society*, 1988.

[5] R. Want E. Mynatt, M. Back. Designing Audio Aura. In *Proceedings of CHI99*, pages 566–572, New York, 1998. ACM.

[6] G. Furnas. The FishEye view: a new look at structured files. In *Proceedings of CHI86*, New York, 1986. ACM.

[7] M. Kobayashi and C. Schmandt. Dynamic soundscape: mapping time to space for audio browsing. In *Proceedings of CHI97*, New York, 1997. ACM.

[8] B. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, 1989.

[9] A. Mullins and C. Schmandt. Audio Streamer: Exploiting simultaneity for listening. In *Proceedings of CHI96*, New York, 1996. ACM.

[10] L. Ludwig N. Pincever and M. Cohen. Extending the notion of a window system to audio. *IEEE Computer*, 23(8):185–188, 1987.

[11] G. Stalton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York, 1983.

[12] L. Stifelman. A paper-based audio notebook. In *Proceedings of CHI97*, New York, 1997. ACM.

[13] E. Wenzel. Localization using nonindividualized head-related transfer fuctions. *Journal of the Acoustical Society of America*, July 1993.