(63)

Navigating a Spatialized Speech Environment Through Simultaneous Listening within a Hallway Metaphor

by

Brenden Courtney Maher

B.A. Clark University 1990

SUBMITTED TO THE PROGRAM IN MEDIA ARTS AND SCIENCES, SCHOOL OF ARCHITECTURE AND PLANNING, IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MEDIA ARTS AND SCIENCES

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 1998

Signature of Author _____
Program in Media Arts and Sciences
January 7, 1998

Certified by _____
Christopher Schmandt
Principal Research Scientist
MIT Media Laboratory
Thesis Supervisor

Accepted by _____
Departmen
Program in Media Arts and Sciences

Navigating a Spatialized Speech Environment Through Simultaneous Listening within a Hallway Metaphor

by

Brenden Courtney Maher

SUBMITTED TO THE PROGRAM IN MEDIA ARTS AND SCIENCES, SCHOOL OF ARCHITECTURE AND PLANNING, ON JANUARY 7, 1998 IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN MEDIA ARTS AND SCIENCES

Abstract

This thesis presents a system for browsing and navigating a database of recorded speech. It uses a spatialized audio environment, taking advantage of the human ability of simultaneous listening. A browsing environment is provided in the form of a virtual acoustic hallway. The user hears several audio recordings simultaneously; each stream of audio is of a specific topic coming from several "virtual" doors along the hallway. The user's head position in real space controls the rate of movement past the doors and the subsequent playing of the audio recordings. Behind each door is heard a "Braided Audio" collage of related news stories. Listening to an individual news recording for a given topic is accomplished in "Audio Rooms" through a new method of selection and representation based on the metaphor of a "FishEye" lens.

Thesis Supervisor: Christopher Schmandt
Title: Principal Research Scientist

Thesis Committee

Thesis Advisor_____

Christopher Schmandt
Principal Research Scientist
Media Laboratory

Thesis Reader_____

Hiroshi Ishii
Associate Professor of Media Arts and Sciences
Media Laboratory

Thesis Reader_____

Michael Hawley
Assistant Professor of Media Arts and Sciences
MIT Media Laboratory

Acknowledgments

I would like to thank the following people:

Christopher Schmandt, my advisor, for his insight for his insight into what makes strong research.
Michael Hawley for his perspective and support.
Hiroshi Ishii for his inspiration and criticism.

Delphine Lui for bringing great joy and love into my life- to which this works owes much.

Lisa Stifelman for her support within the Speech Group and earlier research.
Nick Sawhney for his insight and perspective.
Janet Cahn for being Janet.

David Levitt for his inspiration on spatialized audio.
Debbie Hindus for her support.

Bill Butera for his consistent insight into "what really matters".
Steve Waldman for his friendship and expertise.
Flavia Sparacino for her support and perspective.
Nuria Oliver also for her support.

Celia Pearce for being a great friend and inspiration.
David Zeltzer for support, insight and perspective.
Rita Addison for her support and belief.

I would also like to thank my mom(who would be very proud of me) for teaching me that there is nothing I can't do if I set my mind to it.

ABC New for their sponsorship and perspective in the development of this work.

Minoru Kobayashi and Atty Mullins for their earlier research on which this thisis is based.

The Boston VR Group who have inspired me and taught me many things needed to make this work possible.

Contents

Abstract 2
Thesis Committee 3
Acknowledgments 4
Contents 5

Chapter 1 Introduction 7
1.1 Problems 7
1.2 Simultaneous Listening and Spatial Mapping of Audio Content 8
1.3 Navigating/Browsing Spatialized Speech 9
1.4 Representation and Selection of Specific Speech Recordings 10

Chapter 2 Overview and Related Work 12
2.1 Approach
The Basic Idea of The Audio Hallway Browsing System 12
2.1.1 The Audio Hallway 13
2.1.2 The Audio Rooms 14
2.2 Related Work 16
2.2.1 Simultaneous Listening Research 16
2.2.2 Spatialized Audio Research and Applications 16
2.2.3 FishEye View 17
2.3 Related Work at The MIT Media Lab 17
2.4 Overview of This Thesis 20

Chapter 3 Browsing By Topic The Audio Hallway 22
3.1 Simultaneous Listening A Context for Braided Audio 22
3.1.1 Audio Braiding A Context for Browsing 23
3.1.2 The Audio Braiding Process 23
3.1.3 Audio Braiding A Context for The Audio Hallway 24
3.1.4 Audio Braiding Design Decisions 25
3.1.5 Audio Braiding Design Decisions in The Hallway 25
3.2 The Audio Hallway Spatial Interaction 26
3.2.1 Spacing of Doors 27
3.2.2 Initial Navigation Design Keyboard 27
3.2.3 Initial Navigation Design Head Tracking 28
3.2.3.1 Navigation In The Hallway 28
3.2.3.2 Navigation Into a Room 30
3.3 Overall System Architecture 32
3.3.1 Audio Hallway 33
3.3.2 Audio Rooms 33
3.3.3 Spatialized Audio ToolKit 34
3.3.4 Automatic Braiding 34
3.4 Problems in the Initial Design of The Audio Hallway 34

3.4.1 Audio Braiding Length of Individual Recordings 35
3.4.2 Spacing of Hallway Doors 36
3.4.3 Perception of Speech Recording Location 37
3.4.4 Inconsistent User Position 40
3.5 Other System Implementation Problems 41
3.5.1 Looping Playback of Recordings 41
3.5.2 BSDI Errors 41
3.5.3 Dropped Packets of Audio 41
3.6 Summary Initial Problems in Browsing by Topic 42

Chapter 4 Selecting a Recording The Audio Room 43
4.1 Entering a Room The Problem 43
4.2 The Audio FishEye Lens A Model for Selecting Content 43
4.2.1 The Graphic Model FishEye 44
4.2.2 Audio FishEye Lens Simultaneous Presentation 45
4.2.3 Audio FishEye Lens Spatial Presentation 46
4.3 Modeling The Audio FishEye Lens 47
4.3.1 Magnification A Curved Path vs. a Linear Path 49
4.3.2 Setting a Scale Factor 50
4.4 Controlling The Lens User Interaction 50
4.4.1 Seamless Control over Audio Playback 51
4.4.2 Mouse Control 51
4.4.3 Head Movement in Yaw 52

Chapter 5 User Interaction 53
5.1 User Interaction The Audio Hallway 53
5.1.1. Graphic Representation of Hallway 53
5.1.2 Traditional Challenges of Audio Spatialization 54
5.1.3 Braiding Effects on Hallway Implementation 55
5.1.4 Hallway Navigation by Velocity 56
5.2 User Interaction The Audio Rooms 59
5.3 Integrated Hallway and Rooms 60
5.4 Practical Applications 62

Chapter 6 Conclusions and Future Work 64
6.1 Summary 64
6.1.1 The Idea of The Browsing System 64
6.1.2 Problems in The Initial Implementation 65
6.3 Contributions of This Thesis 68

References 69

Chapter 1 Introduction

1.1 Problems

Scenario 1:
Imagine that you would like to select a specific news story from a large database of news recordings. The recordings are unsorted. You listen to the beginning of the first news story; after a several seconds you realize your not interested in this topic, so you go on to the next news story. You are interested in this topic and would like to hear others related to it. You spend the next 10 minutes browsing the database looking for another related news story. You realize that audio is temporal in nature and slow to browse. At this rate you will miss your entire morning of work before finding and listening to the specific news stories you want from the many topics to browse through. You realize that if you could just listen to many news topics and stories at once while being able to control easily what you'r listening to, you would get to work on time.

Scenario 2:
Imagine sitting in front of your web browser to listen to many speech recordings. First you click on the first hyperlink to start playing the audio. An audio control panel appears displaying "play", "stop", "pause", "fast-forward" and "rewind" buttons. A second click on the "play" button starts the recording playing. You listen for a while and decide you have listened to enough of this recording. you click a third time to stop the file. You repeat this process to listen to the next 100 recordings. When your wrist hurts, you realize there has to be a better way to control audio playback and selection of the recorded audio. You read my thesis and realize there is a better way.

In browsing through text on a page it is easy to control one's visual navigation and change one's focus from one topic of interest to another. This occurs seamlessly and with little effort due to our innate motor control over the ocular musculature. The above scenarios suggest two problems this thesis will address: 1) the design of a system for more effective browsing/navigation of speech through an implementation of simultaneous listening of spatialized audio by topic segmented by topic and 2) the design of a system which provides an effective means for speech audio selection and playback

control by user's head movement and interaction with a metaphoric lens interface.

1.2 Simultaneous Listening and Spatial Mapping of Audio Content

Simultaneous listening of several audio recordings is a cornerstone of this thesis. The human auditory system provides one with the capability of shifting one's attention from one conversation toward another while listening to each simultaneously. This is known as the "Cocktail Party Effect"[Arons 1992]. As we hear a more interesting conversation our focus shifts. Simultaneous presentation of speech allows audio content to be scanned in a manner analogous to visually browsing text across a page. The system presented in this thesis uses simultaneous listening in two ways:

1) It allows the user to browse a "virtual acoustic hallway" in which news recordings, segmented by topic, are heard coming from doors along the hall.

2) It allows the user to listen simultaneously to several related news recordings all of the same topic and select the interesting one among them within an "audio room".

The user selectively changes focus from one topic to another in the hallway and from one recording to another in the rooms. This occurs through his/her innate ability to shift attention as well as through additional methods of navigating and interacting in each of these spaces.

Spatial representation of audio content is the other cornerstone of this thesis. Audio spatialization or 3D audio is the ability to place audio around the user's head so it is perceived as coming from a particular point in space. Audio spatialization in this work serves two main purposes:

1) It increases one's ability to selectively attend to one of several simultaneous sound sources.

2) It provides a means of mapping speech content to space for retrieval and audio playback control.

The system plays several news topics simultaneously within the hallway space. Separating these simultaneous streams spatially increases one's ability to focus on one recording over the other[Stifelman 1994]. Spatial separation is also applied for simultaneous listening of audio recordings within the audio rooms.

1.3 Navigating/Browsing Spatialized Speech

The initial premise of this thesis was to build upon speech browsing systems, developed earlier at the MIT Media Lab. Two of these in particular used spatialized audio and simultaneous listening. While audio spatialization has been shown to enhance simultaneous listening, as mentioned above, this thesis work also looked to extend the use of spatialization into the navigational interface. Mullins work, as we will see, explored simultaneous/spatialized speech for listening to a small number of recordings[Mullins 1995]. This work, in contrast, looks toward the design of a system to handle effective browsing of a hundred or more audio recordings.

Mullins used head position to control user focus upon three simultaneous audio recordings. This work extends Mullins work using a spatial metaphor, simultaneous listening and head position to focus one of (over) a hundred audio recordings. Many recordings can be browsed quickly by simply listening to speech audio while moving within a virtual Audio Hallway.

1.3.1 Initial Premise of The Audio Hallway

A hallway provides a natural metaphor for changing the volume of audio sources-- thereby improving simultaneous listening. While simultaneous listening is improved by spatial separation of audio sources, loudness differences can also strengthen audio stream segregation[Arons 1992]. A hallway metaphor offers an additional advantage in simultaneous listening-- not only are the sources spatialized but the volume changes as one gets nearer

to/farther from each source. As a result, the idea of an audio browsing hallway seemed a logical model for simultaneous listening. As a user's proximity to a recorded clip changes, so does the loudness of that recording. Thus, as users move through the virtual hallway space, they are able to control which recordings become loudest and therefore which recording becomes the primary one in the simultaneous listening experience.

A premise of the Audio Hallway is to map salient points about a particular news topic to a specific place in the hallway. Mapping topics into the space allows one to navigate toward one specific topic over another by moving through the hallway space. This builds upon Kobayashi's work, also developed at the Media Lab, which maps salient points in one recording to different places in space[Kobayashi 1996].

1.4 Representation and Selection of Specific Speech Recordings

The hallway environment provides a mechanism for browsing news by topic. At a doorway along the Audio Hallway one hears a "braided" collage of related news recordings of the same topic. Such a collage entices the user to listen to each of the recordings which make up the collage. Representing and selecting specific recordings is accomplished by modeling a fish eye lens.

An Audio FishEye Lens metaphor was developed and implemented to accommodate searching many recordings about a particular news topic. This metaphor provides the means for one to "focus in"(select) from a large number of audio recordings a few recordings to listen to simultaneously. The hallway metaphor could have been used for the purpose of selecting individual recordings which make up a braided collage. A user would travel down the initial virtual hallway then turn and enter into another virtual virtual hallway for listening to the individual recordings. Providing and additional hallway for the user to travel down, could induce additional challenges for the user in terms of orientation- "Which way is the original hallway of braided topics???". It seemed inappropriate to use the same hallway metaphor used for browsing topics for the purpose of selecting individual recordings which make up a braided collage. An alternative

metaphor for selecting the selecting individual recordings was developed in which the user left the hallway but remained in a fixed position with in a room. An audio room in which individual recordings could be selected decreased chance of a user from getting confused in navigating multiple hallway spaces and remaining consistent with the original hallway metaphor. Within each Audio Room, a FishEye Lens metaphor seemed appropriate to apply toward selecting simultaneous/spatialized recordings at a sub topic level. The lens metaphor allows one to be able to limit the number of simultaneous recordings per topic to four, a reasonable number. It also provides a means of focusing on (selecting) one recording while simultaneously listening to others in one's periphery.

Chapter 2 Overview and Related Work

2.1 Approach: The Basic Idea of The Audio Hallway Browsing System

Figure 1 illustrates the virtual acoustic space described in this thesis.

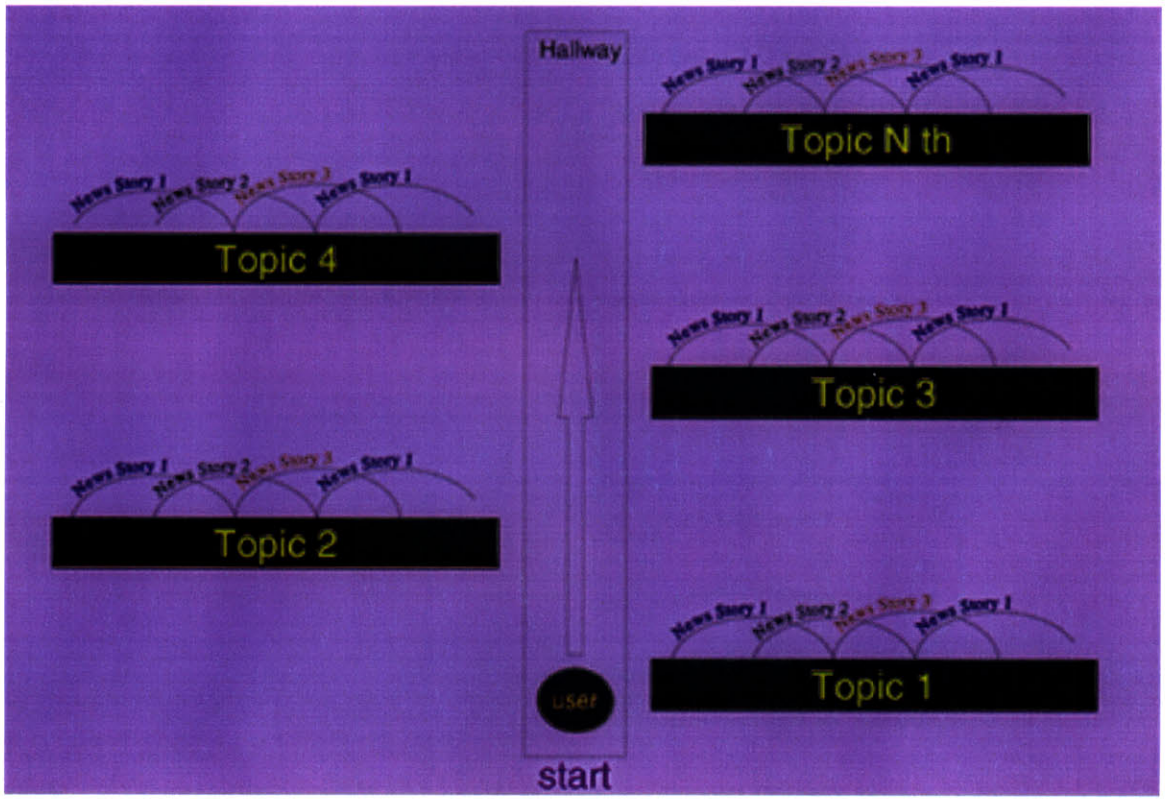

Figure 1: The Audio Hallway browsing space.
A user moves through the virtual hallway.
A braided collage of news recordings is heard at each doorway.

2.1.1 The Audio Hallway

The context for the browsing system is a "Virtual Audio Hallway." News topics originate from virtual doorways as a user progresses down the hallway. The user hears the three nearest topics at any given time. Like sound in a real hallway, the audio sources nearest the user's position are loudest. A limit of three topics was chosen to decrease the strain on the which is incurred from listening to simultaneous audio. Each topic is spatialized to a fixed position relative to the user-- a doorway. Head tracking ensures that each topic remains "fixed", providing a realistic experience even as a user turns his head (while wearing headphones)- an experience un-achievable through use of a joystick alone. While a joystick could have been used for navigation of the audio space, the intent was to research the viability of a hands-off navigational system. The result is that the user can replay an audio topic by revisiting the place along the hallway at which it was originally heard without having to use there hands.

(a) Braided Audio

Each audio topic in the virtual hallway is composed of an audio collage of many recordings, each related to the same topic. This collage or "Audio Braiding" increases the user's ability to browse many sources quickly by presenting many short segments or related news stories one right after the other[Figure 2].

Figure 2: A braided collage: Recordings of news stories of the same topic.

(b) Navigation

The user navigates the audio hallway space by leaning forward to go forward and leaning back to go back. Because the hallway is virtual, the user can traverse the hallway at a much greater rate than a real hallway. The farther the user leans forward, the greater his velocity through the space. A neutral position-- neither leaning forward nor backward, maintains the user's current position in the hallway. This effectively freezes the current volume of all the audio topics. The result is a sense of pausing the navigation and browsing mode. This pausing effectively brings the current nearest topic into focus for listening. Interesting topics in the Hallway may be selected by leaning to the left or right toward doorways. Doing so allows the user to enter an Audio Room for that topic for further listening. Navigation allows the user to control their velocity and direction. As well navigation allows one to focus on simultaneous speech, providing a foundation for browsing many audio sources rapidly and efficiently.

2.1.2 The Audio Rooms

Two metaphors: The Audio Room and Audio FishEye Lens provide implementations for the user to explore further the audio content selected while in the browsing metaphor of the audio hallway. While in the Audio Hallway a user may become interested in listening to the entire length of each recording represented in the braided collage presented at each doorway in the hallway. To listen to each recording of a braided topic in detail the user must enter an Audio Room by leaning toward the nearest doorway(braided topic). The action of leaning toward a doorway allows the user to automatically enter the Audio Room. The user is carried into the metaphoric listening space of an Audio Room, leaving the sounds of the braided topics in the audio hallway behind.

(a) The Audio FishEye Lens

Within an Audio Room a user listens to the content of each recording of the braided collage selected by interacting with a metaphoric fisheye lens and a Graphical User Interface. A user listens to a recording by selecting it (by bringing it into focus above the other recordings) with a virtual lens. This

selection mechanism can be understood by the metaphor of a lens which is able to bring into focus and magnify one object over another. Furnas used a lens metaphor in his earlier work in selecting graphical representation of information[Furnas 1982]. The metaphor applied to the audio domain is a new metaphor called: The Audio FishEye Lens.

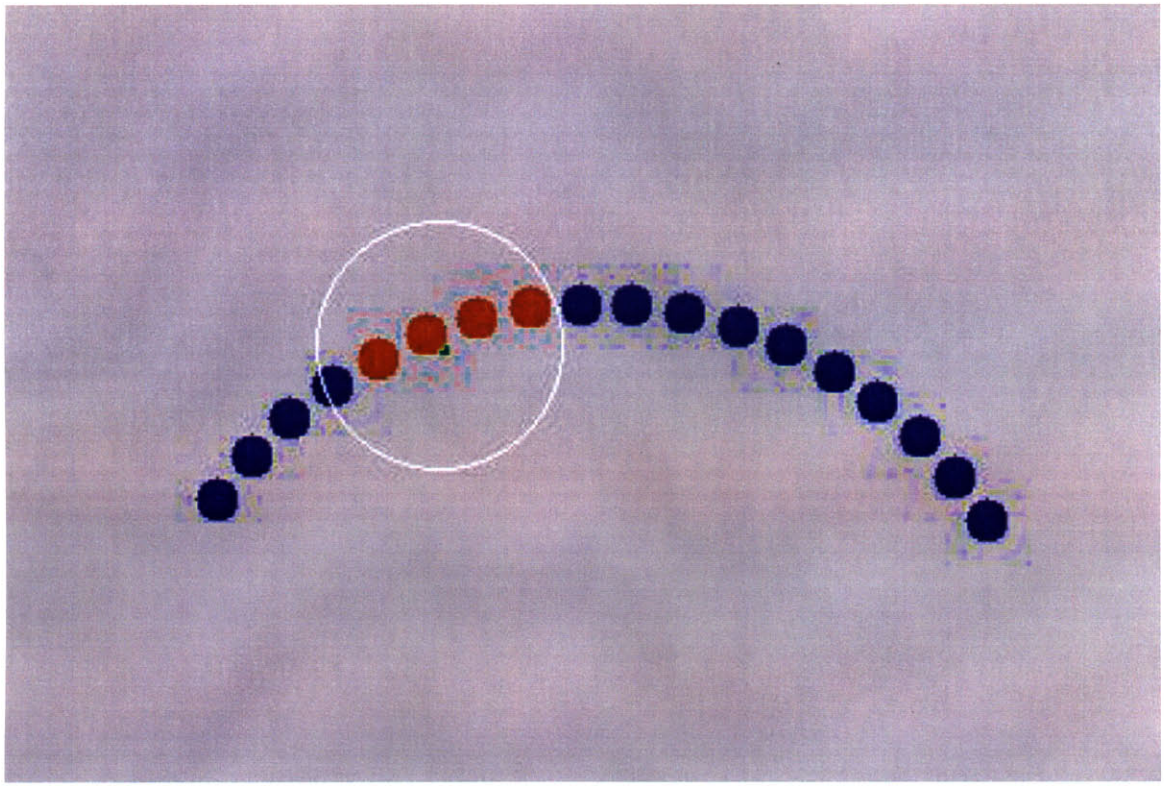

Figure 3: A virtual lens to select the recordings of a braided topic. The lens appears as the larger circle and the selected audio appears as the smaller circles wihin.

Within an Audio Room, an Audio FishEye Lens metaphor provides a method for selecting up to four simultaneous/spatialized audio sources from a group of several related audio sources which comprise a braided topic. These four individual recordings are selected using either mouse control over virtual lens represented by a graphical user interface or by the user changing his head orientation left and right without the use of a graphical user interface. In the former case, entering an Audio Room brings up a graphical user interface and a virtual lens. Moving the lens over dots which represent the audio recordings of a braided audio topic selects the recordings

under the lens[Figure 3]. The selected recordings, shown by the dots in the middle of the lens, are magnified into the virtual acoustic audio space of the room. In the latter case using head orientation, the user rotates his head left or right selecting the four recording from based on the direction of the user's gaze upon the virtual audio sources spread out in an arch around the users head.

2.2 Related Work

2.2.1 Simultaneous Listening Research

Earlier research also showed that subjects had little difficulty in listening to messages played in one ear while rejecting sounds in the other ear. Focus could be easily switched from one source to another at will [Norman 1976].

Early work in simultaneous listening was performed by Cherry in 1953. His results showed subjects presented with simultaneous sources presented to each ear (dichotic) could not restate much about the rejected source but did notice when the gender of the speaker changed or a 400Hz tone was played. Cherry also noted that accents, mean pitches, and speeds and subject matter also affect one's ability to filter out-- one simultaneous source over another[Cherry 1953].

2.2.2 Spatialized Audio Research and Applications

Early research in developing spatialized audio systems was performed in the mid-1980s[Wenzel, Wightman, Foster 1988][Begault 1994]. The NASA Ames Research Center developed one of the first digital systems for generating spatialized audio. The hardware implementation, called the Convolvetron Spatialized Audio, was based on three cues 1) interaural time difference 2) interaural intensity difference and 3) HRTF-- Head-Related Transfer Functions. HRTF's model the spectral filtering of a sound source which is affected by one's physiology.

2.2.3 FishEye View

FishEye View [Furnas 1982] outlines a methodology for generating a small display or representation of a large structure. This work provides a foundation on which the Audio Room and Audio FishEye View are based. The intent of both is to provide a means in which a user can attain an abbreviated "view" of a structure while achieving local detail as well as global context. There is a major difference between these works: 1) This thesis uses the FishEye View metaphor for representing audio sources. 2) The Audio Fish Eye Lens uses a different model for implementing the lens. The original FishEye View work provides insight into the design of the acoustic Audio FishEye Lens and its graphic component.

2.3 Related Work at The MIT Media Lab

Simultaneous Listening

Research on simultaneous listening has been conducted at The Media Lab. Barry Arons discusses several factors which increase the efficiency of simultaneous listening[Arons 1992]. He notes several methods of processing speech audio for effectively increasing one's ability to listen to simultaneous speech. These include: filter streams into separate frequency bands, use synthetic or recorded voices, pitch shifting voices away from each other, associating images the audio and provide spatial disparity between channels.

Lisa Stifelman also researched simultaneous listening. She explored listening comprehension of simultaneously presented speech[Stifelman 1994]. Her results show a clear decline in a subject's performance in comprehending simultaneous sources as the number of background channels increase.

Dynamic Soundscapes

Dynamic Soundscapes[Kobayashi 1996] creates an audio-only browsing environment using spatialized audio and simultaneous listening to enhance a listener's ability to listen to a single audio source. It presents the user with an audio source in which the audio content is mapped to spatial position around the user's head. The system allows the user to start other portions of the same audio source simultaneously at different positions around his head. This provides functionality for the user to be able to associate audio content with spatial position. Furthermore, it provides a means for efficient navigation toward previous portions of the audio source. Simultaneous listening provides a means for the user to change focus of attention selectively from one source to another.

Dynamic Soundscapes demonstrated the possibility for effectively mapping audio content of a single source spatially around the user. Through using a keyboard interface, pointing interface and/or a knob interface, the user controls the focus of simultaneous sources and navigation toward portions of the audio content.

This thesis is founded, in part, on the work of Dynamic Soundscapes. Spatial mapping and simultaneous listening is explored further in this thesis. The major difference between Dynamic Soundscapes and this thesis is: this work explores browsing several speech recordings instead of focusing on how to browse within a particular recording The focus of Dynamic Soundscapes is in providing a spatial/temporal mapping which provides the user with the ability to revisit a specific portion of an audio source. This work is focused on the a spatial mapping which supports navigation among several recordings and the selection of a particular recording out of the many.

Several issues remain the same in this work. Dynamic Soundscapes focuses on user interaction which enhances selective focus of simultaneous audio streams. Also several issues regarding the limitations of spatialized audio remain relevant.

AudioStreamer

AudioStreamer[Mullins 1995] also provides a foundation for this thesis by performing initial exploration into the use of simultaneous spatialized audio to navigate an audio space. This browsing system presents three sources to the listener simultaneously around the horizontal plane of the user's head. Each source is offset by 60 degrees. Head motion is used to accentuate a source's volume by moving in the direction of the source. By moving toward a sound source, the volume of that source is increased 10db thereby changing the user's focus toward that source. The gain of the source decays over time allowing the user to re-adjust his level of focus by continually moving toward the source of greatest interest. A 400 Hz 100msec tone is induced at the onset of each new topic to draw the user's attention toward the new source.

The major differences between AudioStreamer and this thesis are:

1) New topics arrive in the same place in space in AudioStreamer, but in this work, the new topics are spatially separated.

2) AudioStreamer uses active positioning of the user's head to control the volume of a source whereas, in this work, the volume of the source is controlled automatically as a function of the user's position in the hallway.

3) This work provides a mechanism for listening to previously heard audio sources; Audio Streamer does not.

Audio Notebook

The Audio Notebook [Stifelman 1997] is a paper notebook which may link audio conversation to notes on the printed page. Used while taking notes during a lecture, the system provides an intuitive way of navigating toward segments of recorded Audio by pointing at text and graphics drawn within the notebook. The Audio Notebook is relevant to this thesis work in that they both use spatial location to access audio. It is especially relevant to the user

interaction with the GUI representation of the Audio FishEye Lens in the audio room. Both use graphical representations to navigate an audio space. The graphical representations in this work, however, are not mapped to time.

Hyperspeech

Hyperspeech is a system developed for navigating a speech database by issuing voice commands[Arons 1991]. It uses voice recognition to navigate a database of audio segmented by topic. It is significant to the Audio Hallway in that the Hyperspeech system also faced challenges in providing context and "landmarks" to keep users from getting lost. Furthermore, it is suggested that a simultaneous/spatialized approach might be an appropriate interface for navigating an audio database[Arons 1991].

Nomadic Radio

Nomadic Radio[Sawhney 1997] implements a wearable system in which the user can access voice-mail, news, appointments, weather information. The location, speed and direction of spatialized audio sources signify content type, level of urgency, and characteristics associated with the audio content. This work is relevant to this thesis in that it explores how audio spatialization can be used as a means of navigating toward specific audio sources of interest to the user.

2.4 Overview of This Thesis

This thesis describes the design and implementation of the browsing system outlined in Chapter 1. The idea of this browsing system is based on these two hypotheses:

1) The metaphor of a virtual Acoustic Hallway utilizing Audio Braiding, simultaneous listening, and head gesture navigation provides an effective means for navigating a speech audio database.

2) The metaphor of a virtual FishEye Lens for selecting specific audio recordings is less burdensome than traditional "point and click" audio controls. It is also as effective as audio recording selection by head gesture.

Chapter 3 describes the initial design and implementation of the browsing system-- the Audio Hallway. It further describes the browsing system outlined in Chapter 2 and provides a context for discussion of the rest of the system.

Chapter 4 describes the initial design and implementation of "The Audio Room" metaphor for selecting specific audio recordings. The model of the Audio FishEye Lens is introduced.

Chapter 5 details the results of the user interface design decisions, described in chapters 3 and 4, including: user navigation within the Hallway and two methods for selecting specific recordings within the Audio Rooms. Other issues involving user interaction are also discussed.

Chapter 6 provides a review of this thesis and discusses directions for future work.

Chapter 3 Browsing By Topic: The Audio Hallway

Chapter 3 describes the design and development of the browsing system-- the Audio Hallway described in Chapter 2. The purpose of this implementation is to design a system to effectively browse many speech recordings grouped into related topics.

This chapter first describes two methods of reasoning which influenced the browsing system design: Audio Braiding and spatial interaction. It then describes the initial user interface and the overall system architecture which is manifested in the hallway browsing system. Finally, the problems of the initial design of the hallway browsing system are discussed.

3.1 Simultaneous Listening: A Context for Braided Audio

One major problem in the design of the browsing system is how to reduce listening time yet provide enough information for the user to decide if the audio content is interesting. Traditionally, browsing speech takes considerable time as it is temporal in nature. One method, demonstrated by Arons at the MIT Media Lab, is to speed up the speech content[Arons 1993]. Although "speech skimming" proved effective, this thesis work builds upon Mullins's research[Mullins 1995]. Instead of time-compressing speech as in Aron's work, Mullins's work presents multiple channels of speech simultaneously and spatially to reduce listening time. Spatialization separates the speech recordings, increasing one's ability to listen simultaneously. Simultaneous presentation, of course, reduces listening time.

This thesis explores effective browsing of a dynamically changing speech databases based of a hundred recordings using a new method of processing

speech called Audio Braiding. While the spatial aspect of Mullins's simultaneous presentation seemed well suited for representing three recordings, this thesis looks to extend his work to an arbitrary number of recordings and to provide an appropriate context for Audio Braiding.

3.1.1 Audio Braiding: A Context for Browsing

The time it takes to browse lots of speech and consequently listen to a specific speech recording could be greatly reduced if the listener focused in on, and listened to, only the content of interest. Segmenting speech audio first by topic provides this functionality. Audio Braiding provides a mechanism to represent speech efficiently which has been segmented by topic.

3.1.2 The Audio Braiding Process

Audio Braiding is a process in which several audio recordings are concatenated together to form an "audio collage". The idea is to take speech recordings related by topic and present a small section of each one to the user. Audio Braiding, in effect, forms an audio collage in which the user hears a few seconds of one speech recording and then a few seconds of the next etc. The volume of each recording starts at zero then rises to a maximum then falls back to zero. Just as the volume of one recording decreases, the next recording starts to get louder[Figure 4]. This allows the user to hear some of the speaker's voice as well as content for each of the recordings. Some content from each recording can be heard in a short period of time. The purpose is to provide a mechanism for the user to decide how interested she is in a specific collage of speech recordings on specific news topics.

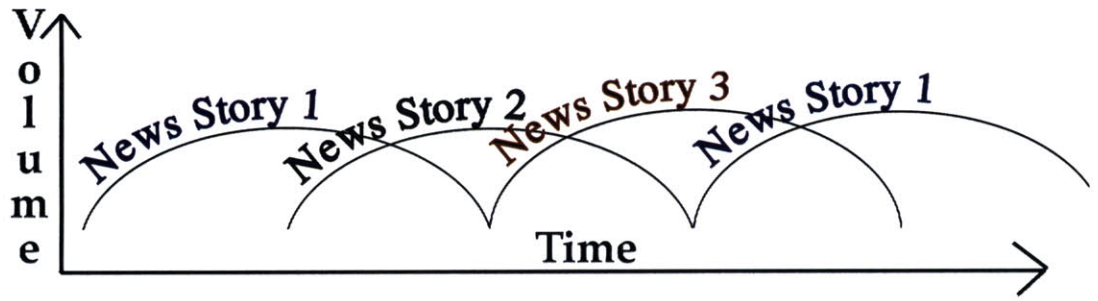

Figure 4: The conceptual image of a braided collage of news by topic.
(a) The volume of each news story starts low then rises, then falls as the next news story is heard.
(b) All of the news stories are on the same topic.

The browsing system receives news every hour from ABC News in the form of audio recordings and verbatim transcripts of these recordings. Speech is segmented by topic by taking the verbatim text transcripts and segmenting the text by topic through a system called SMART[Stalton 1981]. Text is segmented by topic based on word relevance among all the text for each recording for an entire day. Based on text correlation, the speech audio is related by topic and "Braided" as described above.

3.1.3 Audio Braiding: A Context for The Audio Hallway

Creating braided news topics is only an effective aid in browsing if the braided news is presented effectively to the user. This thesis proposes that an appropriate context for Audio Braiding can be established by extending Mullins's and Kobayashi's work in simultaneous/spatialized speech. This "context" has developed into the metaphors of an Audio Hallway and Audio Rooms. A virtual Audio Hallway provides a context familiar to the user in which he can browse simultaneous braided audio topics coming from each doorway. Moving down the hallway allows the user to hear a different collage of related news topics at each door. The "collage of recordings" not only intrigues the listener but also serves as an aid to simultaneous listening. Triesman showed that during simultaneous listening, words could be picked out more readily if all the words were of the same context[Triesman 1967].
A virtual Audio Hallway can also be infinitely long, allowing the system to be scaleable based on the number of news topics received each day. Furthermore,

the metaphor of a hallway allows for "Audio Rooms" which provide the implementation to select and listen to a single audio recording making up the collage.

3.1.4 Audio Braiding: Design Decisions

Early designs for the Braided Audio recordings produced braided audio in which the length of time of each individual recording was one second[Figure 5]. Additional lengths were tried including 5 seconds and 10 seconds for each individual recording making up the collage. Section 3.4.1 discusses the final implementation and reasons behind it.

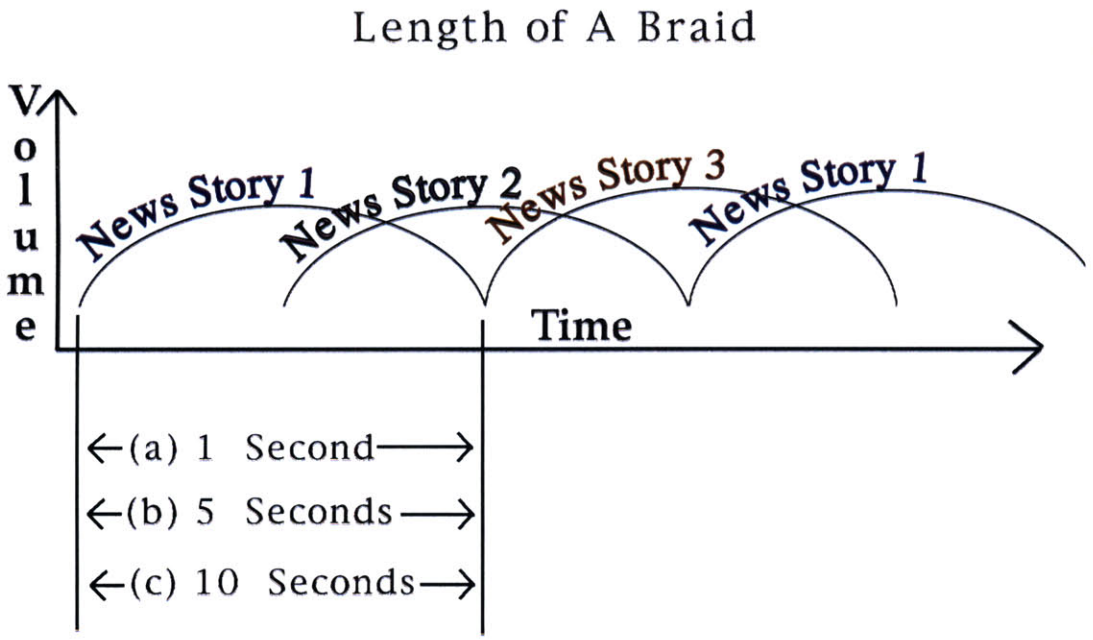

Figure 5: Initial lengths of each "Braid" tested in the making of a Braided Audio Collage.
(a) 1- second "Braid".
(b) 5- second "Braid".
(c) 10- second "Braid".

3.1.5 Audio Braiding: Design Decisions in The Hallway

Section 3.1.4 discuses the decisions around choosing the right number of different news stories put into the braided audio which is heard apart from

the hallway listening environment. This section discusses design decisions made regarding audio braiding in the context of the hallway. With a braid of a length of three seconds(section 3.4.1), it was found a user within the hallway way space could adequately get an idea of the category of news while staying at one of the doorway for typically 3 to 9 seconds(i.e. listening to one to three braids).

The system could present one story at each door(i.e. no braiding). This was in fact implemented during development. While using the browsing system with only one story at each doorway was effective, it defeated the purpose(and metaphor) of being able to browse through many recorded stories of the same topic. If each door only presented one recording the user would have to listen to several doors of the same topic and subsequently another set of doorways for the next topic: The hallway would become extraordinarily long. Even worse, though, would be the fact the audio coming from some doors would be related by the same topic and others would be related by a different topic. In such a listening space, where the audio is segmented both by topic and by recording, it would become difficult for the user to comprehend the space and navigate it.

A design decision was made in which the hallway space would be used for browsing many braided audio topics. Each of the topics, having many related recordings, would be presented to the user in a manner in which they would be able to preview many of the recordings of the same topic. Audio Braiding met this criteria. A design decision was also made such that the Audio Rooms would be a metaphor for browsing and selecting each of the detailed recordings which comprised the braided audio. Here in the Audio Rooms the user would be able to browse and select from any of the audio recordings of the same related topic- but only up to four simultaneously.

3.2 The Audio Hallway: Spatial Interaction

Decisions of the spatial configuration of the hallway and methods of navigation were based on user interaction within the space. Since one moves through the space to listen to audio, the methods of navigation and dynamics of the space are related.

3.2.1 Spacing of Doors

The hallway space was designed so that its length would grow in accordance with the number of new topics/braided recordings generated each day. A lower limit was established such that the system would not have less than four doors, since browsing fewer topics would be of little use. On the high end, the number of doors is unlimited, ensuring a listening environment for all the news topics of the day regardless of the number. The spacing between the doors, however, is pre-determined. Initially the doors were set to be thirty inches apart. This was later changed to fifty inches apart. As we will see, interaction or movement past doors spaced at these intervals is problematic.

3.2.2 Initial Navigation Design: Keyboard

The initial navigational design used keyboard interaction and was a precursor to the head tracking interface. Using the Keyboard interface, one would push the "f" key to go faster; the "s" key to go slower; the "r" key to reverse direction; the "d" key to enter a room; and the spacebar to pause/continue motion through the hallway[Figure 6].

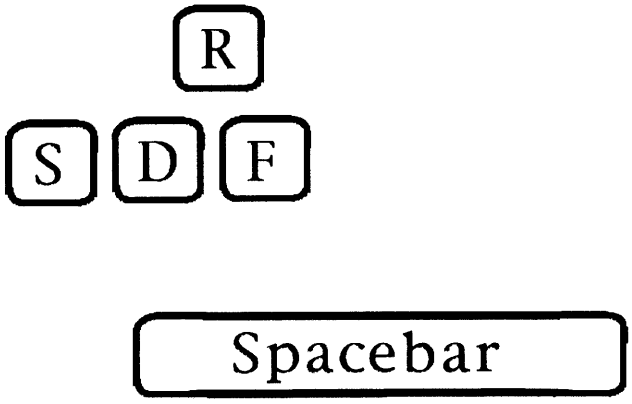

Figure 6: Initial Hallway navigation using keyboard.

(f) faster
(s) slower
(r) reverse direction
(d) enter room
(spacebar) pause/continue motion

The system automatically moved one through the hallway at a constant rate which one could adjust with the keys. However, the keyboard design proved to be cumbersome. Adjusting how fast one moved down the hallway required locating the correct key and keeping a mental count of how many clicks or adjustments one had made. Reversing direction also required finding the correct key. A more intuitive control interface was devised based on head position.

3.2.3 Initial Navigation Design: Head Tracking

The initial implementation also used head tracking to control: movement through the hallway space, rate of movement, direction of movement and selection of braided topics. This control interface builds upon the work of Mullins who used head position to control prominence of an audio recording[Mullins 1995]. Latter work by Kobayashi also measured the direction in which one leaned, to increase the loudness of a particular recording, and ". . . . it enables more efficient and easier selective listening than in the natural environment"[Kobayashi 1996].

The system used a Polhemus FastTrack tracker to (electromagneticly) track the user's head position and orientation. Yaw or azimuth tracks head orientation left to right providing a realistic experience of the audio. Tracking yaw makes it possible to keep the audio emanating from a doorway in a fixed position even while the user has turned his head while wearing headphones.

3.2.3.1 Navigation: In The Hallway

The length of the hallway was mapped to the user-- front to back-- such that looking straight ahead is looking down the hallway when the system starts. Browsing is accomplished by changing head position: The farther one leans forward or backward, the faster he will move in that direction[Figure 7,8,9]. Approaching the initial upright position slows the user's movement. Several iterations on this control method were tested(see section 5.1.4)

Figure 7: This image illustrates leaning forward to navigate in the forward direction.

Figure 8: This image illustrates sitting in the up-right position to pause/stop motion through the hallway.

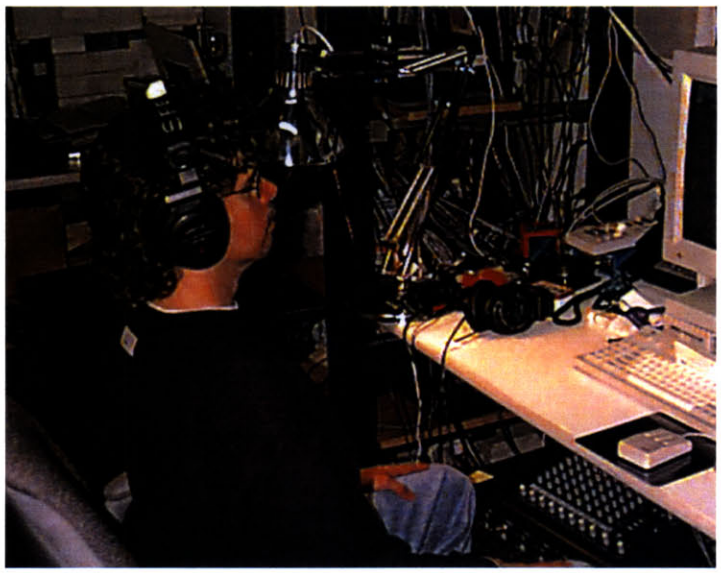

Figure 9: This image illustrates leaning back to reverse direction and review topics previously heard.

3.2.3.2 Navigation: Into a Room

When the user becomes interested in a Braided Topic, she may listen to a specific recording which is part of the braid. When the user leans toward the door of the topic of interest, the head tracking system senses her head position causing her to enter a audio room[Figure 10]. There, in the metaphoric room, the user may select and listen to each recording of a "braided collage"(Chapter 4). When the user is finished in room, a click of the mouse brings her back to the hallway for further browsing of the topics. Ideally the system should track the user head position with in each Audio Room allowing the user to reenter the hallway using head gesture in a similar manner to the way they entered the Audio Room. This interface was not explored although future systems should provide a consistent interface between enter and leaving an audio space.

Figure 10: This image illustrates leaning toward the doorway of an "interesting" braided audio collage. Leaning far enough causes one to enter an Audio Room.

3.3 Overall System Architecture

The overall system hardware configuration is shown in [Figure 11]. For clarity it is easiest to think of the system as two entirely separate systems for rendering spatialized audio: one for the hallway and one for the rooms.

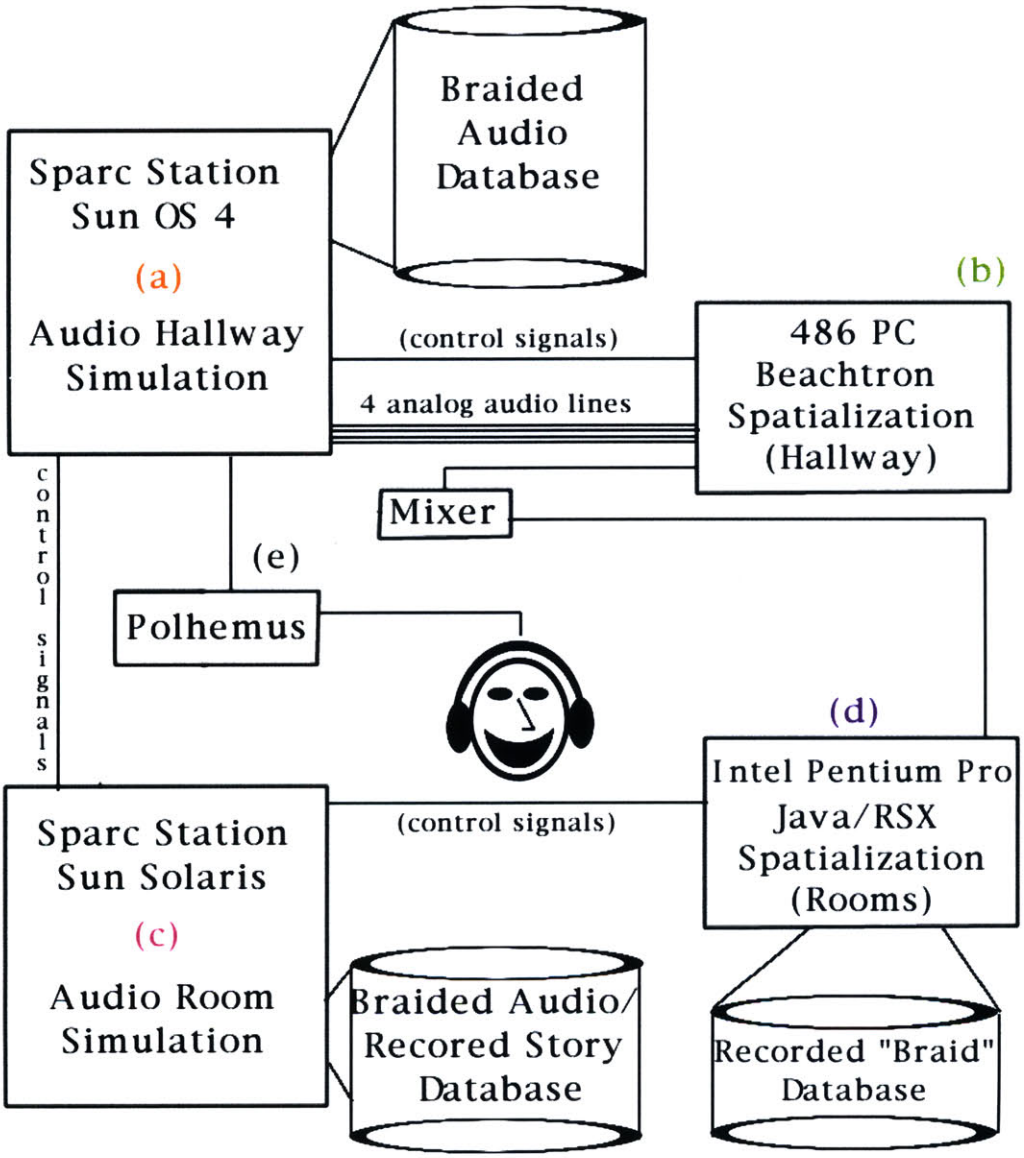

Figure 11: The System Architecture for the Audio Hallway and Rooms.

(a) The Audio Hallway Simulation System.

(b) The Audio Hallway Spatialized Audio Subsystem.

(c) The Audio Room Simulation System.

(d) The Audio Room Spatialized Audio Subsystem.

(e) The Polhemus Head Tracking System.

This dual system is due to two primary factors: 1) The Graphic component of the Audio Rooms required a different OS and programming language from the Audio Hallway system. 2) Memory limitations of the Hallway system prompted use of a second audio subsystem in the Audio Room system. In total, the system runs on four hardware platforms/OS's, uses one external device for head tracking, six client-servers, three databases, a web browser and two methods of audio spatialization.

3.3.1 Audio Hallway

The Sparc Station 10 [Figure 11 (a)] runs the Audio Hallway simulation, playing multiple analog audio signals through its audio ports. Since the Sparc Station has two stereo outputs, at most four monaural audio signals can be set to the hallway audio subsystem.

This "Hallway" audio subsystem, for rendering spatialized audio, is a PC 486 which contains two Crystal River Beachtron audio cards [Figure 11 (b)]. The PC receives four channels of analog audio from the Sparc Station and control signals for specifying the position of each audio source and the user.

A Polhemus sensor mounted on the headphones measures the direction and orientation of the user's head[Figure 11 (e)]. Head tracking helps in the ability to localize audio by providing a change in the audio stimuli as the user moves her head. Head tracking is particularly important in reducing ambiguities between an audio source perceived as coming from in front of the user or from behind. These "medial" front-back ambiguities are reduced by providing computer control over the HRTF filtering[Kendall 1995]. Head position and orientation information is sent to the hallway audio subsystem for controlling audio rendering.

3.3.2 Audio Rooms

The Sun Solaris System manages the Audio Room simulation[Figure 11 (c)]. This system receives the current topic/collage the user had heard and selected in the hallway. It determines the corresponding recorded "Braids" associated with the collage and draws the GUI interface. This system manages which

recordings should be played as a result of collisions with the virtual lens and/or head orientation.

Control signals are sent to an Intel Pentium Pro System-- audio subsystem for spatially rendering the recordings selected in a room. The audio room spatialization uses Intel's RSX audio spatialization paradigm. A Web Browser and server are used to circumvent application execution and security restrictions. Another Web server is used to access audio recordings dynamically[Figure 11 (e)].

3.3.3 Spatialized Audio ToolKit

A general purpose toolkit in C for creating audio virtual environments was created. It features a number of components-- high level API's for managing various aspects of development. These components include an Audio Manager and an Interactivity Manager. The Audio Manager allows for interacting with sound (see section 3.5.1 looping playback of recordings) on the Sun Architecture. The Interactivity Manager provides an API to allow for user interactions (in the form of code) to be dynamically added or taken away from the Scene Graph by calling function names or ID.

3.3.4 Automatic Braiding

A set of software was developed to produce the Braided Audio dynamically on a daily basis. This included Audio Braiding software and several other components. These other components included code to prevent audio content with artifacts from being included in the braiding process. Error-checking software was also created to test final Braided Audio. In all, this included several programs in C, Pearl and Java.

3.4 Problems in the Initial Design of The Audio Hallway

The initial system described in this chapter manifests an implementation in which braided audio is placed in a virtual space, and one must listen

simultaneously to other recordings while navigating through the space. Section 3.1 describes in detail the design rationale behind many of these decisions. This section reports the problems inherent in the initial system and offers methods for resolving them.

3.4.1 Audio Braiding: Length of Individual Recordings

Early designs in which the length of the individual recordings were a half of a second and one second proved to be too short for one to be able comprehend the content of the recording. The recording would go from a low volume to high and back to low before it was possible to hear enough of the story to recognize what it was about. Longer lengths of 5, and 7, and 10 seconds provided too much information. An informal evaluation judged these to be too slow to browse quickly and efficiently. An individual recording length of 3 seconds was informally determined to provide an adequate balance between recognizing content and reducing listening time. [Figure 12].

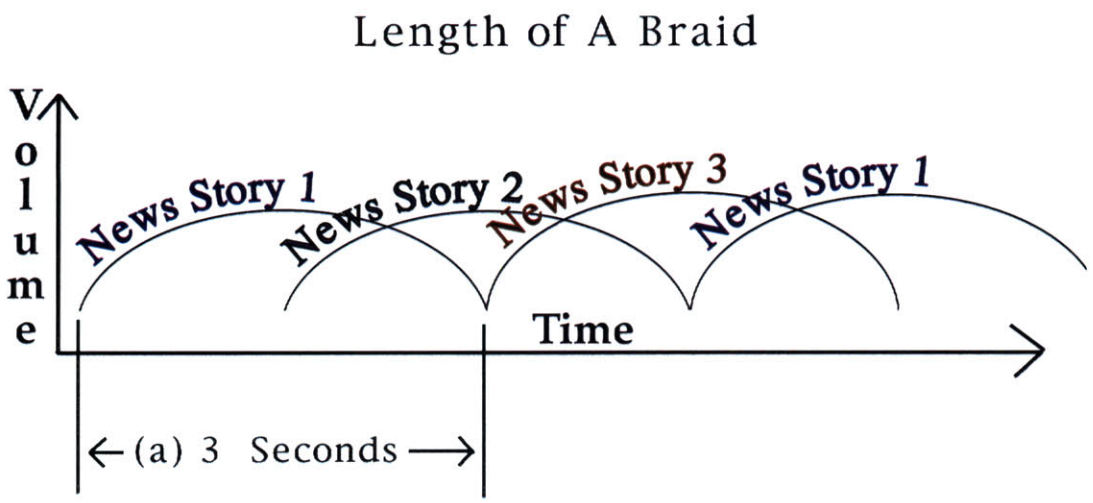

Figure: 12 Final length of each "Braid" of a Braided Audio Collage.
(a) 3- second "Braid".

3.4.2 Spacing of Hallway Doors

In subjective tests, the initial 30 and 50 inch spacing between doors proved inadequate because the doors were too close together. One moved down the hallway so quickly, it was impossible to hear enough of each braided recording to determine its content; one moved from one topic to the next too fast. A spacing of 100 inches was finally determined to be acceptable[Figure 13].

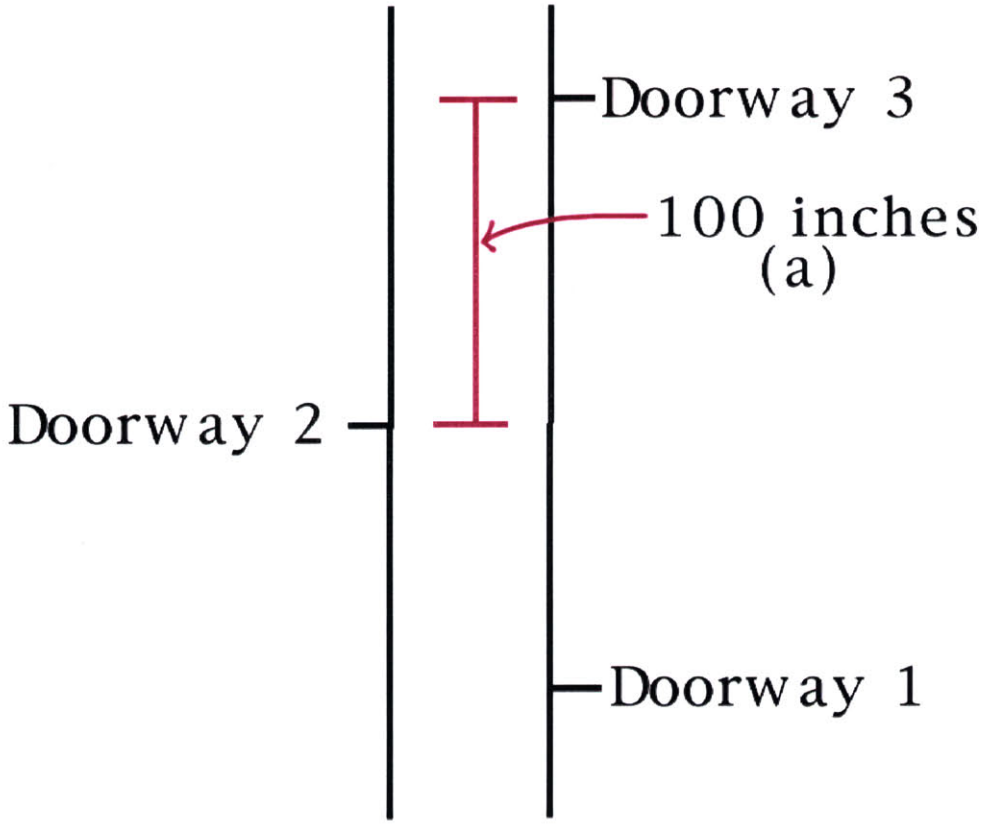

Figure 13: Final length between doorways.
(a) 100 inches between doors.

3.4.3 Perception of Speech Recording Location

One challenging issue involved the perception of the direction the audio sources moved past a user as she moved down the hallway. In short, the accuracy of the spatialized audio came into question. Three out of ten listeners, reported the topics, or doors, passing from left to back rather than from front to back as they would in a real hallway. There are many possible reasons for this:

1) Non-individual HRTF's are used.
2) There were no graphics to correlate with spatialized audio.
3) When topics move front to back, medial spatialization is most difficult.

These factors influencing audio spatialization are covered in detail in Chapter 5.1. 2.

There are several other potential factors affecting perception specifically related to this implementation:

1) Braided audio "motion" causes perceived motion of audio around head.
2) Each Braid of a the Braided Audio changes as well as the Braided Audio topics themselves change as one moves through the space.
3) Special attention must be made toward designing effective simultaneous listening in this complex audio space.

These factors influencing audio spatialization, specific to this thesis, are covered in detail in Chapter 5.1.3.

Some users also perceived audio moving in a circular motion around their heads[Figure 14 (a)]. As one moves past the doorways the audio should be perceived as continuing to move straight back behind the user[Figure 14 (b)]. To counteract the perception of circular motion, Braided Audio topics were rendered so that the Braided topics farthest from the listener were rendered even farther to his/her left and right. The Braided Audio which was farthest from the user's position in the hallway [Figure 15 (b)] was rendered to be even farther from the user's head (i.e. farther to the left or right). This processing seemed to counteract the rounding effect[Figure 15 (d)].

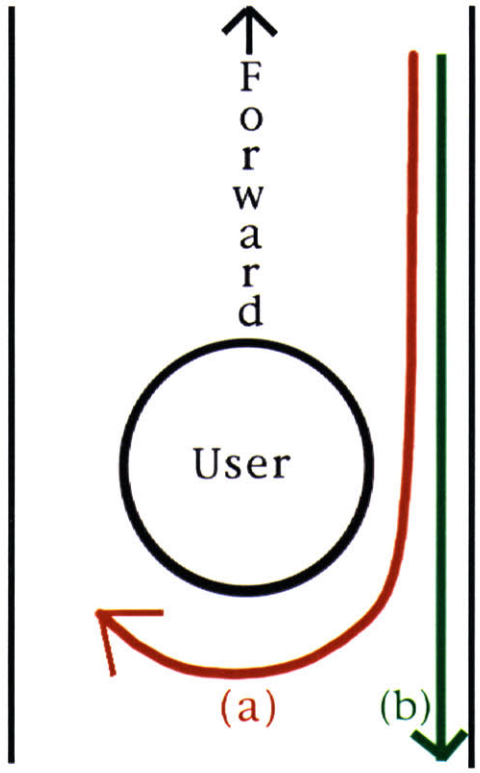

Figure 14: Shows perceived spatialization of audio in hallway.
(a) Inappropriate "rounding" of audio moving around head.
(b) Depiction of how the audio should be perceived.

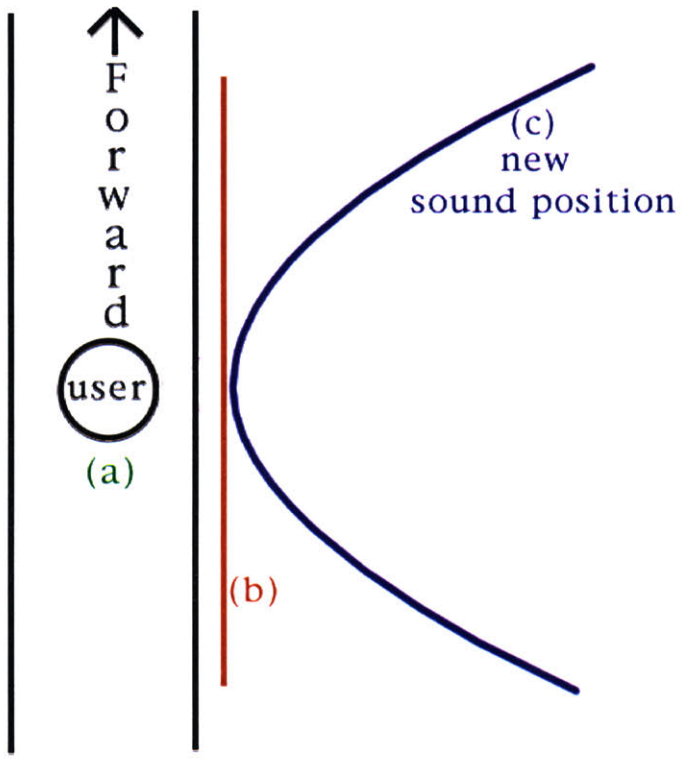

Figure 15: Shows special processing applied to counteract the perceived rounding of audio around a user's head.
(a) direction of user motion through hallway.
(b) original path of spatialized audio.
(c) new spatialization to counteract "rounding". Topics farthest from the user are rendered even farther off to each side.

3.4.4 Inconsistent User Position

In the initial implementation using keyboard navigation, the user's motion through the space was generated automatically. Beyond this, the user could increase or decrease the initial rate at which the system "carried him along". It was observed that, for a given rate user motion (i.e. position) was inconsistent. These variations had the result that in some computationally expensive processes took longer to process than others. The result was a drop in frame rate. Switching over to a time-based model for rendering a frame of audio solved this problem. In short, Distance = Rate/Time was used to calculate the actual distance the user moved even if Time varied due to demand on computational resources.

3.5 Other System Implementation Problems

Several other implementation issues challenged the design and development of this browsing system. Many of these issues resulted from building upon a hardware/software architecture developed in the Speech Group which is still growing toward robustness.

3.5.1 Looping Playback of Recordings

Each Braided Audio recording should loop so that the user is always hearing the topics emanating from the doorways. Looping of spatialized audio proved to be problematic. The Speech Group's Network Audio Server- NAS provides services such a playing audio for the Audio Hallway application. The NAS and the earlier implementation for playing multiple streams audio in Mullins's work[Mullins 1995] were not compatible. In short, NAS did not implement an API which included looping of audio files. An extended set of software needed to be written to provide this basic functionality.

3.5.2 BSDI Errors

Every so often, NAS server crashes. As a result the Audio Hallway Application--which is implemented on top of NAS- crashes. This is highly inconsistent as well as problematic. The cause of the low level Byte Stream Manager errors has not been determined.

3.5.3 Dropped Packets of Audio

Quite often the system would become overloaded. Unable to process the audio data, audio would be dropped resulting in an incoherent listening experience. This presented a challenge as the limits of the system were constantly being pushed.

3.6 Summary: Initial Problems in Browsing by Topic

3.6.1 Inadequate Spacing between doors.

The initial spacing between the doors provided poor listening experience in which it was difficult for a user to comprehend the Braided topics and maintain a sense of orientation within the hallway.

3.6.2 Initial Navigation Design Keyboard

The initial keyboard interaction proved to be too cumbersome for navigating the hallway space. An alternative method using head tracking was devised.

3.6.3 Audio Braiding Length of Individual Recordings

The length of each braid which comprises a Braided Audio topic was inadequate in providing the right amount of information to peak a user's interest in the audio content.

3.6.4 Perception of Speech Recording Location

The user's perceived a "rounding effect" of the braided topics moving around his/her head while moving past the doorways. A special processing was employed to counteract the perceived effect.

3.6.5 Inconsistent User Position

The user's position while moving through the hallway was inconsistent. To counteract the user's random motion through the hallway, the simulation loop was switched to being based on time rather than per frame.

3.6.6 Other Implementation Challenges

Several other implementation challenges were experienced in the development of the initial system. These included looping the recordings, low level system errors and dropped packets of audio.

Chapter 4 Selecting a Recording: The Audio Room

Chapter 4 describes the design and development of the Audio Rooms described in Chapter 2. The purpose of this implementation is to design a system for selecting specific recordings within a braided collage in each doorway of the hallway.

This chapter first describes the reasoning behind the Audio FishEye model for representing speech audio. The Audio FishEye model in relationship to simultaneous listening and spatialization is presented. It then describes the implementation of the Audio FishEye Model and two methods of user interaction: Head Yaw Interaction and Virtual Lens Interaction. Initial problems in the design of Audio Rooms are also discussed.

4.1 Entering a Room: The Problem

The Hallway model presented the users with topics of Braided Audio at each doorway. Leaning toward the door allows one to enter a audio room for that specific topic. The problem is how to present the user with a methodology for selecting a specific audio recording-- from each Braided Audio Collage-- given the problems related to navigating speech covered in Chapter 3. More specifically, the problem is how to choose a small subset of the recordings to listen to simultaneously. The Audio FishEye Lens offers such a methodology.

4.2 The Audio FishEye Lens: A Model for Selecting Content

Each stream of Braided Audio contains a collage of many audio recordings. Section 3.1 discussed the afordances of using simultaneous listening for browsing. Section 3.1 also discussed the afordances of simultaneous listening

within a spatial environment. It was realized Furnas's work in representing information through a FishEye Lens could be extended from the graphic domain into the audio[Furnas 1982]. This thesis presents the selection of audio by the Audio FishEye Lens metaphor. It extends Furnas's work and builds upon the research in simultaneous/spatialized listening by Mullins and Kobayashi. The Audio FishEye Lens presents an alternative method for browsing/navigating and selecting recordings.

4.2.1 The Graphic Model: FishEye

The Audio Rooms and the Audio FishEye Model for selecting and listening to recordings has its roots in the graphic FishEye work by Furnas. His work outlines a methodology for generating a small display of a large structure.
The idea behind his work is:

> A very wide angle, or fisheye, lens used at close distance shows things near the center of view in high magnification and detail. At the same time, however, it shows the whole structure--with decreasing magnification, less detail--as one gets further away from the center view[Furnas 1982][16] .

Figure 16: This image illustrates the distortion which occurs as objects near the lens's edges are spread out relative to objects in the center of the lens.

Furnas's fisheye representation provides a foundation for the Audio Room and Audio FishEye Lens. The intent of both is to provide a means by which a user can attain an abbreviated "view" of a structure while achieving local

detail as well as global context. The major difference between his work and this thesis is that the Audio FishEye explores the audio domain and uses a different method for modeling the lens.

4.2.2 Audio FishEye Lens: Simultaneous Presentation

It was realized that the notion of a "lens" could provide an effective metaphor for selective listening of one audio recording from many. Entering an Audio Room brings up a GUI. Each recording which comprises a braided audio topic is represented as a dot. The system automatically draws-- in the form of an arch-- as many dots as there are recordings in the braided topic. A white circular "lens" is controlled by mouse movement or head yaw. The user moves the lens over the dots of interest to him in order to play the recording associated with that dot. The physical constraints are such that only four dots fit under the lens at any given moment. This constraint limits the number of audio recordings which can be played simultaneously to four [Figure 17].

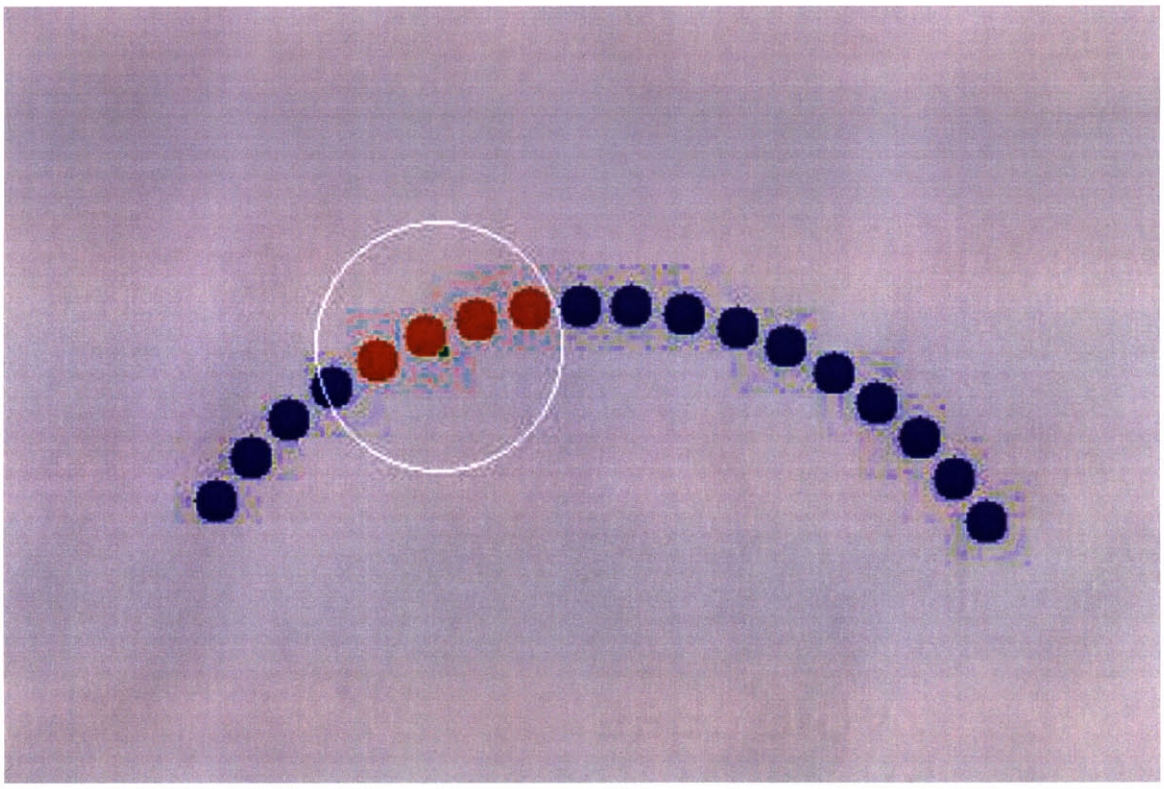

Figure 17: A virtual lens to select the recordings of a braided topic. The lens appears as the larger circle and the selected audio appears as the smaller circles wihin.

4.2.3 Audio FishEye Lens: Spatial Presentation

The recordings corresponding to the dots under the lens are played simultaneously. The challenge is to spread the audio out spatially to increase one's ability to listen simultaneously. A lens which can "magnify" and spread out the content over which it is focused provides a metaphor for doing the same in the audio domain. The dots which are under the lens are selected, and their audio is "magnified" or spread out into the space of the Room[Figure 18].

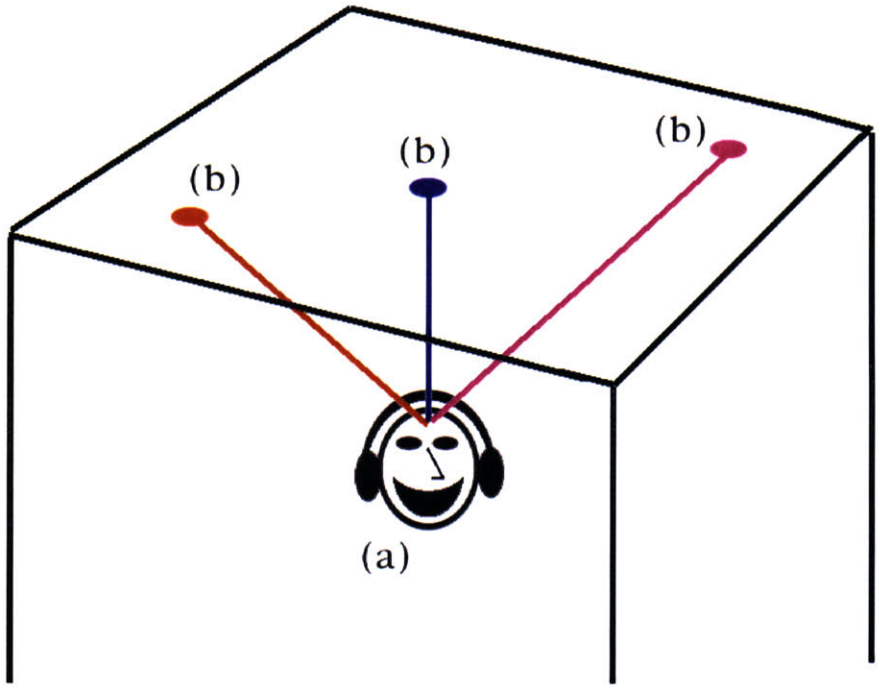

Figure 18: Shows the perceived "magnifying" of news recordings into the space of an Audio Room.
(a) user in a "virtual" audio room.
(b) spatial "spreading-out" of recordings.

The dot directly under the center of the lens is brought into strongest focus and its corresponding recording is heard louder then the recordings corresponding to the off-center dots. This allows one to listen to other recordings simultaneously and shift one's focus easily by changing which dot is under the lens.

3 Modeling The Audio FishEye Lens

In his work, Furnas, closely models graphic information representation to a fisheye lens. He describes the three essential factors for modeling his interface as: 1) a focal point 2) a distance from the focal point for each object and 3) the level of detail (LOD) i.e. importance or resolution of an object being viewed. His model describes how graphic objects at the center of the "lens" can be

scaled and transformed in size. Applying these transforms allows graphic objects at the center of the lens to become bigger and more prominent than objects in the periphery of the lens[Furnas 1982].

Furnas's model of magnifying graphics is such that:

1) Graphic entities nearest the center of the lens move out of the plane of the screen toward the user. Objects in the center of the lens "bulge out" more than those in the periphery[Figure 19].

2) Objects are rendered along a curved path out from the center of the lens[Figure 19].

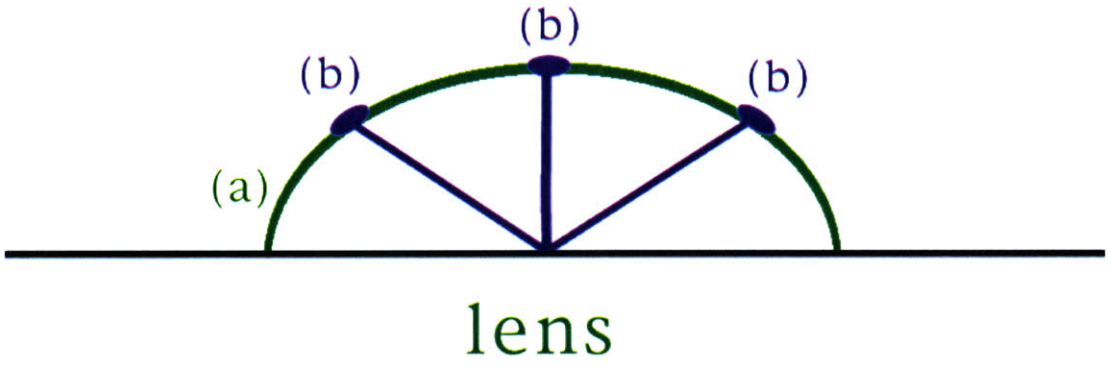

Figure 19: Shows objects rendered to the curved path along the surface of a lens.
(a) the lens
(b) objects rendered along a "curved" surface of a lens.

These two characteristics describe a model which renders the graphics along a curved path. The research in this thesis began development toward this "curved path" model when it was realized that a simpler "linear path" model would suffice for the audio domain.

4.3.1 Magnification: A Curved Path vs. a Linear Path

Initial work was done to model the lens magnification such that the audio would be rendered along a curved path from the center of the lens as demonstrated by Furnas's work.

One specific implementation of the curved path involved magnifying or rendering the audio into (no more than) four distinct positions in a curved path around the user's head. It was realized that in the audio domain the same experience could be achieved through a linear model of rendering the audio.

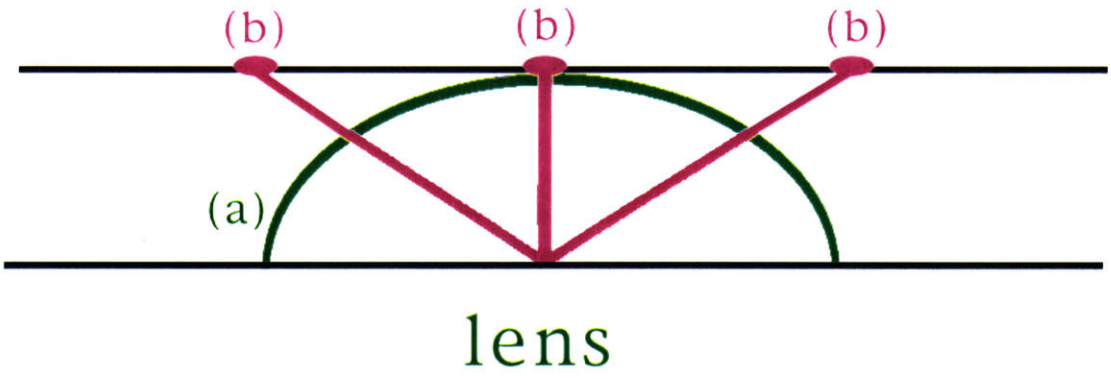

Figure 20: Shows objects rendered to the curved path along the "linear path".
(a) the lens
(b) objects magnified and rendered along a "linear" path.

In the audio domain, the distance of a sound source from a user affects its loudness. For two objects emitting sounds of equal intensity, the one farthest away will be perceived to have a lower volume. A linear model for Audio FishEye was developed based on this notion. It was decided that the curved path model would not be followed explicitly. Instead audio is rendered out from the center of the lens along a linear rather than curved path[Figure 20]. Spatialized audio naturally renders audio farthest away from the user as more quiet or with less prominence than audio nearest the user. This affordance of spatialized audio allows one to achieve the same kind of scaling of graphics that Furnas developed with his curved model. The essential factor needed to make the "linear" Audio FishEye work in the same way as Furnas's graphic

model is setting a scale factor. This is comparable to setting the focal length of a lens. A scale factor needs be determined for controlling how much an audio recording at the periphery of the "lens" is perceived: In essence how low it is relative to the recording directly in focus under the center of the lens.

4.3.2 Setting a Scale Factor

The initial mapping which mapped or magnified the audio into the room space was insufficient. The audio was not spread out enough to provide adequate spatial separation for the recordings when played simultaneously. An informal observation noted that scaling the placement of the audio by a factor of three proved effective in providing spatial separation for simultaneous listening. This allowed the user to easily select a specific news audio recording from as many as seventeen others at one time.

Controlling The Lens: User Interaction

The user interaction of controlling the lens (i.e. what recordings one focuses on), provides seamless control over audio playback. The lens metaphor of selecting recordings for listening also provides user control in one of two ways: mouse interaction or head movement in yaw[Figure 21].

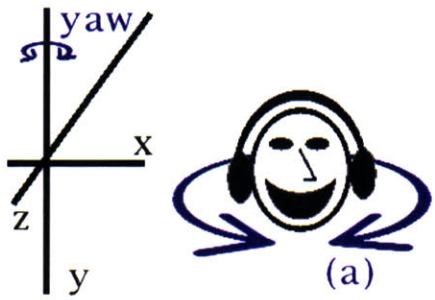

Figure 21: Shows head movement from left to right in yaw(azimuth) orientation.
(a) Yaw orientation

4.4.1 Seamless Control over Audio Playback

The lens is a compelling metaphor for selecting audio content. By definition, a lens brings into focus information from a background of uniformity. Much work has been done in developing graphical interfaces which have properties of guiding the user's attention seamlessly toward some desired task[Mackinlay, Rao, Card 1995]. Interaction in the Audio Rooms attempts to model this flow of focus demonstrated by earlier graphical information browsing systems[Colby, Scholl 1989][Rennison 1994]. In GeoSpace the..."most important information is displayed at a higher level of opacity, and related information is related with medium translucency[Lokuge, Ishizaki 1995]. Accordingly, the user is able to play several recordings simultaneously and seamlessly shift from one recording to another as if shifting focus to the translucent information in GeoSpace. The result is that recorded audio fades in and out in an aesthetically pleasing manner under user control.

4.4.2 Mouse Control

In this method the user controls the lens movement (i.e. the ability to select dots for listening) by using the computer mouse. Much work has been done in the design and development of graphical interfaces for interacting with audio through the use of a mouse. Specific work has been done on accessing

salient points within an audio recording[Hindus, Schmandt, Horner 1993][Kobayashi 1997]. This work looks toward using the mouse effectively for selecting among many audio recordings. Moving the mouse changes the position of a virtual lens to select audio in a GUI interface. Effects of this interaction are discussed in Chapter 5.

4.4.3 Head Movement in Yaw

A second method of selecting audio recordings involves using head orientation in yaw. Moving or turning one's head to the left or right controls movement of the lens. The purpose of this method is to utilize a hands-off interaction paradigm.

Chapter 5 User Interaction

Chapter 5 describes the results of the user interface design decisions described in Chapters 3 and 4. This chapter first covers user interactions in the Audio Hallway, then in the Audio Rooms. The final section covers the user's experience of interacting with the system as a whole: integrated hallway and rooms.

5.1 User Interaction: The Audio Hallway

The Audio Hallway presented a browsing experience which was both effective in enabling user's to choose one of fifteen different news topics. At times, However, some users were challenged. The following sections examine these experiences in detail.

5.1.1. Graphic Representation of Hallway

The Audio Hallway has been presented to approximately 100 people, about 80 percent of whom have suggested that a graphical representation of the hallway would greatly help their ability to navigate through the hallway space. Many users, at times, expressed experiences of being lost or confused in the hallway space. Research shows that limitations in audio spatialization, both in modeling and perception, can be improved by correlating graphical cues to spatialized audio events:

> An example of the interrelationship is that the grouping of sounds can influence the grouping of visual events with which they are synchronized and vice versa. . . . the tendency to experience a sound coming from a location at which visual events are occurring at the same temporal pattern (the so-called ventriloquism effect) can be interpreted as a way in which visual evidence about the location of an event can supplement unclear auditory evidence[Bregman 1990 p.653]

A graphical interface for the hallway was not developed from the beginning as the system was initially intended to be an audio only listening environment. Future work may look toward a graphic representation of the Audio Hallway.

5.1.2 Traditional Challenges of Audio Spatialization

Users experiencing the Audio Hallway sometimes report a sense of being confused by how the audio sources are moving around them. An understanding of the hallway model is helpful; however, it is not always adequate. Modeling an Audio Hallway poses particular challenges inherent to audio spatialization itself.

The ability to perceive audio spatially is a complex process which is not fully understood. Much research, however, has been done in this field. Our ability to localize audio is primarily due to our ability to process interaural time differences (the times it take for the same sound to reach each ear) and interaural intensity differences (the differences in intensity of the same sound at each ear)[Wightman 1989]. Perceptual errors can be categorized into three areas: 1) Localization errors in azimuth and elevation, 2) Reversal errors (hearing a virtual sound source at its mirror position in the rear hemisphere instead of in the front, or vice versa) and 3) Distance errors [Begauldt 1992].

Problem 2, above, has the biggest impact on the use ability of the Audio Hallway system. The problem is the difficulty for users to localize audio in the medial plane (front-to-back plane)- such as in a hallway. interaural time differences and interaural intensity differences are cues which provide little help in this orientation[Hebrank, Wright 1974]. Spectral cues provided by the pinne are essential for localization in the medial sagittal plane[Butler, Belendiuk1977]. Filtering also occurs due to reflections and diffractions from the human head and torso. These anatomical differences should be modeled into the HRTF used by the spatializing system. Because of the difficulty in obtaining individualized HRTF's, "averaged" HRTF's are often used (including this system)[Moller 1995]. Individualized HRTF's more accurately model the complex spectral shaping which occur based on individual anatomical differences. Use of individualized HRTF's in the Audio Hallway

system should help decrease medial plane ambiguities[Wenzel 1993]. The audio environment may be modeled specifically to correct for these limitations as described earlier in Section 3.4.3.

5.1.3 Braiding Effects on Hallway Implementation

An informal evaluation of 10 users presented with braided audio recordings reported mixed reviews. Some felt that it was simply confusing to listen to while others did not feel confused.
Aside from the traditional problems associated with audio spatialization mentioned above, this system presents it's own challenges.

The user's ability to make sense of the Audio Hallway seemed to vary related to:

1) one's understanding of the hallway model.
2) the nature of the recorded content.
3) the juxtaposition of the place in the braided audio stream in relationship to the surrounding topics.

In this system there is a complex relationship between simultaneous presentation, audio spatialization, dynamically changing content and user interaction. Each of the above factors affect the others and one's overall listening experience.

It was found that providing the user with a background of the system helped him avoid a sense of being lost or confused. Explaining the metaphor of the hallway, how the braided audio is rendered into hallway, and how the user's motion directly affects his listening experience provides an understanding necessary to use this browsing system effectively. The improvement in one's ability to localize audio by providing context is described by Begault. Expectation and familiarity affect localization particularly with speech[Begault 1994].

It was also found that users could better make sense of the browsing experience when the topics or recorded segments which make up a braided collage alter the gender of the speaker. This makes sense in light of Bregman's

research in which simultaneous audio streams which have similar acoustic features are perceived as forming a single auditory stream[Bregman 1990]. This explains why users have difficulty segmenting one part of a collage from another as well as one topic from another if the speaker is the same in the recordings or if different speakers are of the same gender.

The user listens to news topics simultaneously from each doorway. A single braided topic from a doorway is also iterating through each of the recordings which make up that particular collage. Clearly, there is a lot of simultaneous audio to comprehend. Stifelman reports that increasing the number of simultaneous streams decreases one's ability to attend to multiple simultaneous streams [Stifelman 1994]. As a result, Mullins suggested reducing the user's cognitive load by reducing the amount of information presented on each channel. Another method which might improve the ability to listen simultaneously within the hallway is to present cues which signal one to switch attention or focus from one simultaneous source to another[Mullins 1995].

5.1.4 Hallway Navigation by Velocity

Navigation through the hallway space is controlled by the position of one's head. The user leans forward to go forward and leans back to go backward. The farther the user leans, the faster he will go in either direction. An upright position stops or slows movement through the space.

Two navigational interfaces were developed:

1) Box method.
2) Gradual method.

The difference between the two is the position in real space at which the user's velocity starts to increase. Figure 22 depicts the mapping between user position in real space and velocity within the hallway space.

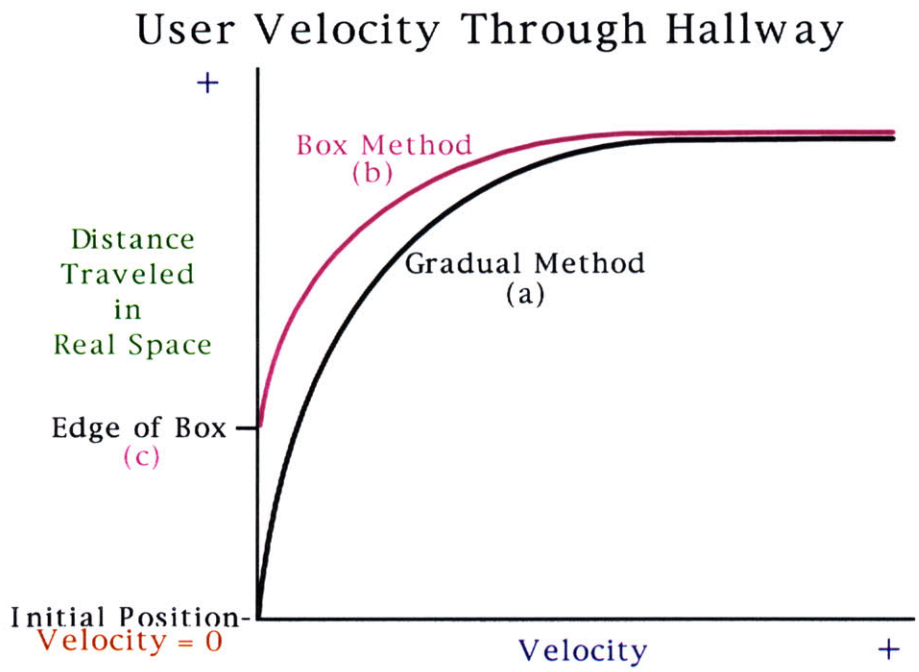

Figure 22: Shows the mapping between user position in real space and velocity within the Audio Hallway.

(a) Velocity mapped in the "gradual method."

(b) Velocity mapped in the "box method."

(c) Velocity starts to increase at edge of "box."

(a) The box method

In this implementation the user has a "zone" or area around her which, as long as the user remains in this zone, motion through the hallway is paused[Figure 23 (a)]. This allows the user to have an area of "free motion" in real space which does not affect the hallway application. More importantly, it provides a distinct perceptual cue when user motion goes from paused to un-paused and vice versa. Specifically, it provides feedback about the user's state: browsing (moving through the hallway) or (paused listening to a topic at a doorway). After the user reaches the edge of the "pause" area she slowly begins to pick up speed in the forward or reverse directions.

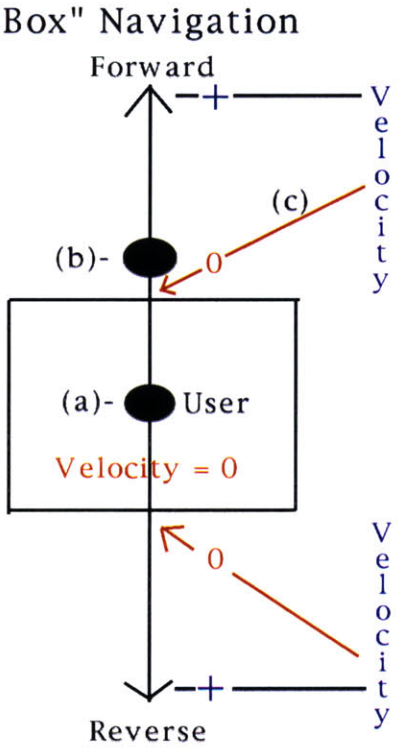

Figure 23: Box method for mapping user motion to space

(a) While in box or "zone" around user-- Velocity = 0

(b) User's velocity starts to increase when leaning forward, past edge of "Box".

(c) Velocity is 0 at edge of box and increases "non-linearly" as user leans farther forward.

Informal experiments showed that this method was effective at providing helpful information about the user's position. Several users commented that they were not as "disoriented" and had a better sense of control since this method is less sensitive to user motion.

(b) The gradual method

In the gradual method the user's motion through the hallway is paused only when he remains in one specific place: the position of the user when the system started[Figure 24 (a)]. If the user is not in this specific place he is either moving forward or backward with an increasing velocity. Gradually speed increases from the initial starting point[Figure 24 (b)].

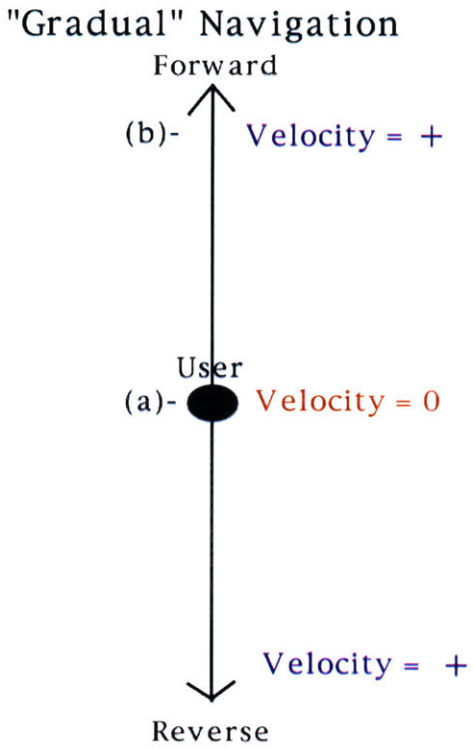

Figure 24: Gradual method for mapping user motion to space
(a) Initial user position: Velocity = 0
(b) User's velocity starts to increase when leaning forward past the initial starting point.

Informal experiments showed that this method was more effective at providing a continuous flow of information. Users could easily go from focused listening, while moving slowly, to browsing and moving through the hallway quickly. It was found that more experienced users preferred this implementation of navigation. Overall, the gradual method of head tracking provides a much more effective and continuous browsing experience.

5.2 User Interaction: The Audio Rooms

The Audio Rooms provide a listening environment for selecting one recording over many others through simultaneous listening and audio spatialization. Two methods of selection were designed. One utilized selecting recordings by moving a virtual lens through mouse interaction. Another

method utilized selecting recordings by rotating one's head left and right in the azimuth plane or yaw orientation.

Section 4.4.4 described using the mouse to move the position of a virtual lens. Performing this action allows the user to select specific audio recordings which comprise a Braided Audio topic. The GUI implemented in this system was intended to be a prototype for an implementation which used a real lens for audio selection. While the virtual lens prototype is far from the tangible/physical afordances of a real lens, the current system does offer insight into the acoustics and use ability of future systems.

5.2.1 Selection of Recording by Lens

Out of twenty five users, almost all reported that the lens metaphor and the simultaneous/spatialized presentation of audio provided an implementation with the following utility:

1) The lens allowed the user to listen while having control over which simultaneously sources they listened to, and for how long they listened. Putting control over the simultaneous audio into the "hands of the user" helped decrease disorientation while listening to several sources simultaneously.

2) The acoustical environment provided by the audio spatialization allowed each audio recording (represented by a dot) directly under the lens to be distinctly louder. Louder sources directly under the lens provides a means for the user to easily distinguish one simultaneous source from another and shift focus from the most prominent recording to others in the periphery.

5.3 Integrated Hallway and Rooms

The fully integrated system allowed users to select news stories by topic in the hallway and listen to specific news recordings in the Audio Rooms. When the Audio Hallway and Audio Rooms were integrated, the twenty five or so

users who experienced the system reported the following about the system as a whole:

1) The system Audio Hallway and Audio Room metaphors were simple and an easy to understand means for browsing and selecting specific news recordings.

2) Users liked using the head motion for navigating the hallway space and particularly liked being able to easily and quickly focus in on specific recordings. Furthermore, users liked the facility of being able to enter back into the hallway and continue browsing in the familiar context.

Visually we are able to see and browse a collage of many graphic elements, sorting by size, color, transparency, position and aesthetics. Our visual acuity enables us to process all of this information seemingly at once. The equivalent in the audio domain it would seem to involve a tremendous mixing of audio information, decreasing intelligibility and one's ability to browse. The hallway space does both a low-level signal mixing sense with a high-level content browsing sense. Two-thirds of the users found the system quite usable even though they were browsing through a complex audio information space of multiply mixed recordings.

The Audio Room environment faired much better. About ninety-five percent of those who tried the audio rooms and Audio Fisheye Lens metaphor were able to simultaneously browse multiple audio recordings which were mixed together. This success is due to the fact that the audio space is less complex than the hallway as no braided audio is used here. Furthermore, the sound model used a much more focused and directed presentation of the audio content(it could be projected in a narrow band in a specific direction). The Audio Room environment, while gave the user a far greater ability to browse audio recordings which involve a low-level signal mixing, still falls short of the browsing ability of many graphic elements which are possible in the graphics domain.

The system has been used successfully for browsing up to 12 news topics in the hallway and browsing and selecting among 17 detailed recordings in the audio rooms using the Audio Fisheye Lens metaphor.

5.4 Practical Applications

This thesis presents a system which uses the Hallway, Room and Fisheye metaphors for interacting with speech. This system focused browsing speech news information. One question which arises is: When is it appropriate to use each of these for navigating a new information space as well as other domains?

First, it is important to understand which parts of the system did not work. One clear area where the system fails is in providing a coherent transition between the Audio Hallway and the Audio Rooms. Problems in a lack of a graphical representation in the hallway were covered in section 5.1.1. Graphic Representation of Hallway. However a larger problem remains in that the user transitions from a non graphical interface in the hallway to a graphical one in the rooms. This doesn't seem to make sense to most users. Either the hallway and rooms should have a graphical interface or be an entirely audio only interface. Furthermore, it is clear that if a graphical interface is to be used, it could be improved upon greatly. Perhaps, if a graphical interface is used, the virtues of spatialized audio could be used in conjunction with better filtering of the audio content along with corresponding visual representations of how this filtering occurs. Such a system, in regards to news, might consist of many other graphic representation(other that simple dots representing each audio recording) which one could dynamically manipulate to filter news by time, known speakers, geographic origins etc. Combining this visual filtering along with further segmentation of audio spatially could provide a much more powerful means of navigating news if a graphical interface were to be used effectively.

Could the system presented in this thesis be extended toward building an audio-only Web interface that is as search able as AltaVista and as browsable as HTML? I believe this work proves that building a system such an audio-only Alta Vista is plausible. Such a system would have to provide an

improved mechanism for browsing over an encumbering head tracking system with a high learning as the one presented here. Much greater thought would be needed into how the audio could be segmented appropriately for a spatialized audio search engine. Could one build an audio-only interface as browsable as HTML? I think this is quit plausible. The success of HTML is in it's simplicity and universality. As we better understand how to interact with spatialized audio and what spatial paradigms are successful for achieving various tasks, I believe it will be necessary to take advantage of a spatialized audio protocol as simple as HTML.

Chapter 6 Conclusions and Future Work

This chapter reviews this thesis and its implementation, discusses user feedback, and shows directions for future research.

6.1 Summary

This thesis describes an audio browsing system which uses spatialized audio and simultaneous listening to browse and navigate an audio database of fifteen topics and one hundred news recordings.

6.1.1 The Idea of The Browsing System

Browsing audio is not as efficient as browsing text because of the temporal nature of sound. The browsing system in this thesis is based on two key ideas: 1) simultaneous presentation of audio 2) a spatial representation of audio for enhancing simultaneous listening and providing a means of navigating for audio selection and playback control.

Simultaneous presentation of braided audio collages allows one to browse several news topics by switching one's focus from one topic to another. Many speech recordings which comprise a braided topic can also be scanned quickly by shifting one's focus among a few select simultaneous recordings.

Spatial presentation maps news segmented by topic into specific places along a virtual audio hallway. Such mapping allows one to control which topics a user hears. By tracking the user's head position in real space, the user controls his/her navigation along the hallway. Navigating the hallway space allows one to select and play a braided news topic from the nearest doorway to the user. Navigating by one's head motion subsequently allows one to play, fast forward, rewind and/or pause the audio. Audio spatialization also allows

recordings which comprise a braided collage to be magnified and spread out into the space of Audio Rooms through an Audio Fisheye Lens metaphor. This metaphor allows for selective listening of each recording of the collage.

6.1.2 Problems in The Initial Implementation

The initial system presented several challenges which needed to be overcome. These challenges include: 1) The spacing between doorways was too short. 2) The length of a braid (recording) for each braided topic was either too short or long. 3) The user perceived the audio as moving around his head instead of straight back behind him. 4) The user's position within the hallway space was inconsistent over time. 5) Scaling of audio recordings into the Audio Rooms was inappropriate.

These problems were addressed by 1) setting the spacing between doors to 100 inches, 2) setting the length of a braid to three seconds, 3) developing algorithms which spread the audio which had been farthest from the user even farther apart in an effort to counteract the perceived motion of audio moving around a user's head, 4) calculating a user's position based on elapsed time rather than position corrected for inconsistent user position, and 5) using a scale factor of three which was determined to give the best magnification of audio into the Audio Rooms.

6.1.3 Challenges in User Interaction

User interaction with the browsing system demonstrated in this thesis presented numerous challenges. The system presents the user with braided audio, localized to doorways along a hallway. The braided audio itself is composed of a collage of news clips of which the volume is continuously rising and falling. While braided audio alone is not difficult to comprehend, adding it to the hallway space does make it more challenging to decipher. Not only does each braided audio collage change its content but each collage comes into and out of hearing as one progressed down the hallway. While the braiding definitely helps a user to browse the topics and peeks the user's interest in a topic, at times it is less effective. Braiding is especially less effective when either: a) The user does not understand the hallway model or

why a pause in movement in the hallway space does not pause the braided topics (i.e. keep them from changing) or b) Each braid is in the voice of speakers of the same sex, making it difficult to separate the news content.

Two methods for using head navigation were presented. The box method was shown to be most effective for new users as it provided a clear distinction between pausing mode and browsing mode. The gradual method, however, was preferred by the more experienced user as it continuously increases the velocity through the hallway space. The result was a more intuitive and easier to use interface for quick browsing by those with experience on the system.

Users found the Audio Fisheye Lens implementation to be effective allowing one to select an audio recording from a group of seventeen others. Of particular interest was the ability to preview some of the other recordings content before having to shift focus to that recording. The lens metaphor allowed users to grasp the context of the audio quickly and to use this method of selecting audio recordings easily.

6.2 Future Work

The audio hallway has few cues to give the user feedback as to where he is in the space. Future systems might include a graphical representation of the hallway or other audio cues to give the user a better sense of orientation within the hallway.

The Audio Rooms used a graphical user interface and a virtual lens. The Audio Fisheye lens metaphor could be better understood if the audio and visual environment both distorted information in a similar manner. The current system provides a continuity and uniformity of both audio/graphical space due to the fact that the user receives direct auditory feedback from his interactions with the graphical interface. Subtle movements to the virtual lens easily informs the use of the relationship between the two spaces.

While the dots under the lens do not spread visually and bend while being magnified, the audio does get spread out into the acoustic space of the room.

This tight relation ship between what the user see's and hear helps reduce the discrepancy that the visual field does not in fact warp. It is proposed that a final system could accomplish this visual magnification by using a real physical lens over a panel of LCD's. Such a system would give an accurate and real sense visually of how the audio is analogously magnified and spread out.

Some users noticed a poor sense of the spatialization of the audio. Future systems might include individual head-related transfer functions to improve spatialization. Such an addition would provide a better sense of the position of the topics and improve intelligibility among the topics and the corresponding recordings which make up each collage.

Future systems might also include alternative methods of navigating and selecting audio recordings, rather than using head position. These might include manipulation of physical objects such as a real lens or manipulation of physical objects which represent each audio recording and allow for playback functionality based on their physical orientation.

On the other hand, it is conceivable that an audio only system could be designed for listening an browsing news recordings and other speech information within an automobile. Improvements to head tracking technology and further refinements in spatialized audio interfaces could speech browsing while one is driving.

Presenting speech simultaneously and spatially helps a user to browse more audio information. Achieving the same level of efficiencies in browsing a complex visual information space is at present time difficult to match within the audio domain. A much more intelligent approach for future spatialized/speech interfaces is to provide better filtering of the audio content to increase the users ability in navigation and selection.

6.3 Contributions of This Thesis

The major contribution of this thesis is the design of an audio browsing environment based on spatial navigation and simultaneous listening of speech audio. The design includes a metaphor for browsing speech segmented by topic called the Audio Hallway. In the context of the hallway, a new method of processing speech called Audio Braiding was designed to improve browsing of simultaneous sources. This work builds on the strong foundation of work in simultaneous/spatialized listening developed previously at The MIT Media Lab by Mullins and Kobayashi.

An interface, using head position, was developed for navigating the hallway space. This head-driven interface along with the presentation of topics along the hallway provide an effective method for shifting one's focus from a browsing mode to focusing in on specific audio recordings. This research of a methodology for shifting focus from the general to the detailed contributes to a larger body of work in the design of information retrieval systems.

An Audio Room metaphor provides a context for the implementation of the Audio Fisheye Lens. The Audio Fisheye Lens is a method of representing and selecting audio for listening, building upon the earlier work by Furnas. It extends his work by exploring implementation issues relevant to the domain of audio. This thesis concludes that in order to achieve the perceptual experience of the audio being spread out (as if being magnified) a system utilizing spatialized audio does not need to model the curved path that a lens produces.

Finally, this work contributes a system which is effective for browsing up to fifteen news topics and selecting one from among a group of seventeen news recordings. In total, the system allows one to select a specific recordings from a total of one hundred news recordings from varying topics. Through simultaneous listening and spatialization it provides a means for seamlessly interacting with speech which does not burden the user with traditional playback control buttons.

References

[Arons 1991]. Arrons, Barry, Hyperspeech Navigating in Speech-Only Hypermedia, HypertextProceedings 1991.

[Arons 1992]. Arrons, Barry, A Review of the Cocktail Party Effect. Journal of the American Voice I/O Society(1992).

[Arons 1993] Arons, Barry, Speech Skimmer: Interactivly Skimming Recorded Speech UIST'93, AMC

[Begault 1992] Begult, Durand, Perceptual Effects of Synthetic Reverberation on Three-Dimentaial Audio Systems, Journal of the Audio Engineering Society, Vol. 40, No. 11, November 1992.

[Begault 1994] Begult, Durand, 3D Sound for Virtual Reality and Multimedia, Academic Press,Inc.

[Bregman 1990] Bregman, A.S, Auditory Scene Analysis: The Preceptual Organization of sound. MIT Press, 1990.

[Butler, Benlendiuk 1977] Butler, Robert, Belendiuk, Krystyna,Spectral cues utilized in the localization of sound in the medial sagittal plane, Journal Acoustic Society of America, Vol 61, No. 5, May 1977.

[Cherry 1953] E.C.Cherry. Some experiments on the recognition of speech, with one and two ears. Journal of the Acoustic Society of America, 25: 975 - 979(1953).

[Colby, Scholl 1989] [Colby, Scholl, Laura, Transparency and Blur as Slective Cues for Complex Visual Information, Visible Language Workshop, Media Laboratory, Massachusettes Institute of Technology 1989.

[Furnas 1982] Furnas, G, The FishEye view: a new look at structered files, Bell Laboratories, 1982.

[Hebrank, Wright 1975] Hebrank, Jack, Wright, D,. Spectral cues used in the localization of sound sources on the medial plane, Journal Accoustic Society of America, Vol. 56, No. 6, December 1974.

[Hindus, Schmandt, Horner 1993] Hindus, Debby, Schmandt, Chris, Horner Chris, Capturing, Structuring, and Representing Ubiquitous Audio, ACM Transactions on Information Systems, Vol. 11, No. 4, October 1993.

[Lokuge, Ishizaki 1995] Lokuge, Ishantha, Ishizaki, Suguru, GeoSpace: An Interactive Visualization System for Exploring Complex Information Spaces, CHI 1995.

[Kendall 1995] Kendall, G, A 3-D Sound Primer Directional Hearing and Stereo Reproduction. Computer Music Journal, 19:4, pp.23-46, Winter1995. Massachusetts Institute of Technology.

[Kobayashi 1996] Kobayashi, Minoru, Design of dynamic soundscape: mapping time to space for audio browsing with simultaneous listening, Speech Interface Group, Media Laboratory, Massachusettes Institute of Technology 1996.

[Kobayashi 1997] Kobayashi, Minoru, and Schmandt, Chris, Dynamic Soundscape: mapping time to space for audio browsing. CHI 1997

[Mackinlay, Rao, Card 1995] Mackinlay, Jock, Rao, Ramanna, Card,Stuart, An Organic User Interface For Searching Citation Links, CHI 1995.

[Moller 1995] Moller, Henrik, Head-Related Transfer Function of Human Subjects, Journal of the Audio Engineering Society, Vol. 43, No. 5, November 1995.

[Mullins 1995] Mullins, Atty Thomas, AudioStreamer: Exploiting Simultaneity for Listening CHI '95 Short Papers.

[Norman 1976] D.A Norman. Memory and Attention. John Wiley an Sons(1976).

[Rennison 1994] Rennison, Earl, Galaxy of News: An Approach to visualizing Expansive News Landscapes, ACM Multimedia Demonstration Proposal 1994.

[Sawhney 1997] Sawhney, Nick, Nomadic Radio: A 3D Environment for wearable Computing, Speech Interface Group MIT Media Lab.

[Stifelman 1994] Stifelman, Lisa J, The Cocktail Party Effect in Auditory Interfaces: A Study of Simultaneous Presentation, MIT Media Laboratory Technical Report

[Stifelman 1993] Stifelman, Lisa J, VoiceNotes: A Speech Interface for a Hand-Held Voich Notetaker, CHI 1993.

[Stifelman 1997] Stifelman, Lisa J, The Audio Notebook: Paper and Pen Interaction with Structured Speech, Speech Interface Group, Media Laboratory, Massachusettes Institute of Technology 1997.

[Triesman 1967] Treisman, A. and Geffen, G, Selective Attention: Preception or Responce? The Quarterly Journal Experimental Psycology, XIX(1):1-17,1967.

[Wenzel 1993] Wenzel,Elizabeth M, Localization using nonindividualized head-related transfer functions, Journal Acoustic Society of America, July 1993.

[Wenzel, Wightman, Foster 1988] Wenzel,Elizabeth M, Wightman, Fedric L, and Foster, Scott H, A Virtual Diaplay System for Conveying Three-Dimential Acoustic Information, Proceeedings of The Human Factors Society 1988 (1988)

[Wightman 1989] Wightman, Federic, Headphone simulation of free-field listening. I: Stimulus synthesis, Journal Acoustic Society of America, Febuary 1989.