

# Hear&There: An Augmented Reality System of Linked Audio

by

Joseph Michael Rozier

Submitted to the Department of Electrical Engineering and Computer  
Science

in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science in Computer Science and Engineering

and

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2000

©2000 M.I.T. All Rights Reserved.

Author .....  
Department of Electrical Engineering and Computer Science  
May 22, 2000

Certified by .....  
Judith Donath  
Assistant Professor of Media Arts and Sciences  
Thesis Supervisor

Accepted by .....  
Arthur C. Smith  
Chairman, Department Committee on Graduate Theses

# Hear&There: An Augmented Reality System of Linked Audio

by

Joseph Michael Rozier

Submitted to the Department of Electrical Engineering and Computer Science  
on May 22, 2000, in partial fulfillment of the  
requirements for the Degrees of  
Bachelor of Science in Computer Science and Engineering  
and  
Master of Engineering in Electrical Engineering and Computer Science

## Abstract

In this thesis, I designed a system for augmenting a space with linked audio. Using this system, individuals can associate audio clips with a location in real-world space. When an individual using the system passes through this augmented space, he or she can hear the audio clips that have been left by traveling through the associated locations. Furthermore, audio clips in the environment can be linked together, so that an individual listening to one audio clip can be directed to related audio clips in the augmented space. In addition, a method of allowing the user to navigate through the augmented space is presented.

Thesis Supervisor: Judith Donath

Title: Assistant Professor of Media Arts and Sciences

# Acknowledgments

I would like to thank the following people:

Judith Donath, who provided guidance and inspiration throughout the work performed on this thesis, invaluable advice, and an overall pleasant graduate year at MIT.

Arthur C. Smith, my academic advisor, who has looked after me since I was a sophomore.

Karrie Karahalios, who worked with me tirelessly at the outset of this project to get the GPS system up and running, and on many of the hardware issues involved in the project, and who helped me conceive and narrow the scope of the project.

Tuan Phan, an undergraduate researcher, who helped me design and implement the spatialized audio portion of the project.

Birendro Roy, another undergraduate researcher, who helped me interface the PalmPilot into the project.

The rest of the Sociable Media Group, for providing critiques and suggestions at our weekly meetings and late at night.

My Brothers and friends at Zeta Beta Tau, Xi Chapter, who have supported me throughout my five year journey at MIT, especially my Little Brother Gavin Miyasato and the entire Alpha Sigma class.

The rest of my close friends at MIT and long distance friends who have each provided an outlet for enjoying life while being a grad student.

Mom, Dad, Stephani, Jennie, Larry, Evelyn, and Crystal (my immediate family), and of course my aunts, uncles, and cousins who have encouraged me through my years before and at MIT.

And Ryan, Morgan, Alex, Kyle, and Madison (my nephews and nieces) and Natalie (my young cousin) who have provided inspiration and happiness, something I needed a lot of to complete my studies at MIT.

# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Previous Work</b>	<b>9</b>
2.1	<i>Touring Machine</i> . . . . .	9
2.2	<i>Spirit</i> . . . . .	10
2.3	<i>Audio Aura</i> . . . . .	11
2.4	<i>Automated Tour Guide</i> . . . . .	12
<b>3</b>	<b>Scenarios</b>	<b>14</b>
3.1	Scenario One: Fernanda and the Brazilian Network . . . . .	15
3.2	Scenario Two: Karrie the Explorer . . . . .	16
3.3	Scenario Three: Brian the Author . . . . .	17
<b>4</b>	<b>The Contributions of <i>Hear&amp;There</i></b>	<b>19</b>
4.1	Audio . . . . .	19
4.2	Authoring . . . . .	20
4.3	Navigation and Linking . . . . .	22
4.4	Location Accuracy . . . . .	22
<b>5</b>	<b>The Design of <i>Hear&amp;There</i></b>	<b>24</b>
5.1	The SoundSpot . . . . .	24
5.2	Activation Networks . . . . .	27
5.3	Field Use versus Desktop Use . . . . .	28
5.4	Authoring Software . . . . .	30

5.4.1	Field Authoring . . . . .	31
5.4.2	Desktop Authoring . . . . .	31
5.5	Playback Software . . . . .	32
5.5.1	Field Playback . . . . .	32
5.5.2	Desktop Playback . . . . .	32
5.6	Hardware . . . . .	32
5.6.1	Field Hardware . . . . .	32
5.6.2	Desktop Hardware . . . . .	36
5.7	Navigation . . . . .	36
5.7.1	Initiating the Experience . . . . .	36
5.7.2	Two Methods of Exploration . . . . .	37
5.7.3	PalmCompass . . . . .	37
5.7.4	The Use of Audio . . . . .	38
5.8	Spatialized Audio . . . . .	39
5.9	Storage of Data and Sounds . . . . .	40
5.10	Security and Privacy . . . . .	41
<b>6</b>	<b>Implementation</b>	<b>43</b>
6.1	Problems . . . . .	43
6.1.1	GPS . . . . .	43
6.1.2	PalmPilot . . . . .	44
6.1.3	Sound Issues . . . . .	44
6.2	Actual Implementation . . . . .	45
<b>7</b>	<b>Further Research</b>	<b>46</b>
7.1	Questions About Community and Identity . . . . .	46
7.2	Hardware . . . . .	46
7.3	Wider Range and Indoors . . . . .	47
7.4	Additional Content . . . . .	47
<b>8</b>	<b>Conclusion</b>	<b>48</b>

# List of Figures

5-1	The “field” in this implementation of <i>Hear&amp;There</i> . Because of technical limitations on the GPS, we limited this project to the Media Lab courtyard. . . . .	29
5-2	The graphical user interface for <i>Hear&amp;There</i> . Each circle represents a SoundSpot. This GUI can be used to create and edit SoundSpots, and to simulate the augmented environment. . . . .	30
5-3	The hardware for <i>Hear&amp;There</i> used in the field. The highly accurate GPS receiver is rather large, so we have attached it to a luggage cart. . . . .	33

# Chapter 1

## Introduction

Augmented Reality is a new area of research in which a virtual world is overlaid on top of the real world. Instead of the completely immersive environment of virtual reality, augmented reality attempts to use the flexibility of the digital world to enhance the environment in which we all live. Users of an augmented reality system are able to maintain the context of their surrounding environment, while still obtaining the benefit of additional sensory input and information.

Augmented reality can take many forms, most commonly visual and audio overlays on the world. In this thesis, I will use audio overlays to give individuals the ability to leave audio imprints in a real-world space. These imprints will be groups of related sounds that represent the author in some way. For example, an imprint could consist of an individual's recorded thoughts on a historical monument and some of the author's favorite music.

These imprints can then be tied to a particular location in a real-world space. Whenever someone using the *Hear&There* system walks in the vicinity of that location, he or she will be able to hear the imprint. In the example stated above, the author could leave his imprint about the monument near the monument he is describing. Then, whenever someone walked by that monument, he or she would hear the author's imprint.

I will create the infrastructure necessary to allow audio to be associated with a given location in space and to control the playback of the audio when someone travels

near the location. This will include software issues, including how to store and edit the audio, as well as hardware issues, such as how to find the precise location of an individual in a space.

Once this infrastructure has been completed, I will be focusing on the issue of how imprints can be linked together. There are many important issues to consider when providing a way to link imprints, which will be addressed in this thesis.

I begin this thesis with previous work that has been done in the area of augmented reality, with a particular focus on audio augmentations. I then delve into a few scenarios to familiarize the reader with the major concepts in this project. I then look at a few of the major contributions of *Hear&There* before presenting the design of the system, and then notes on the actual implementation. Finally, I finish up with a brief discussion of future work and a conclusion.



# Chapter 2

## Previous Work

Although Augmented Reality is a relatively new area of research, there have been several groundbreaking projects performed in the past several years. Here, I will describe the work that has been done previously. In Chapter 4, I will describe several ways in which *Hear&There* differs from these projects.

### 2.1 *Touring Machine*

One of the most important works in augmented reality to date is the *Touring Machine* (Feiner 1997), performed by Steven Feiner, et. al. at Columbia University. The focus of this project was in providing context-sensitive visual overlays as an individual walked through the Columbia campus.

After donning the necessary equipment (including a Global Positioning System [GPS] receiver, a computer, goggles, and a portable computer) the user was free to stroll around campus. As the user moves around the campus, the names and functions of the various buildings appear overlaid on the buildings in his or her head-mounted display. That is, the user is able to see computer generated names floating over the real buildings.

Whenever the user focuses on a particular building (indicated by pausing his glance on the building for a few seconds), additional information about that building appears in his view. This information can include what departments are located in a

particular building and which professors work in the building.

Additionally, context-sensitive pages on the world wide web (WWW) are be loaded on the handheld computer that the user carries around. This provides an augmentation to the world, but exists separate from it.

The *Touring Machine* is an important work that addressed many issues of augmented reality. The ability to determine where the user was standing and what he or she was looking at using a global positioning system receiver and a head mounted sensor is very important for visual augmented reality. Additionally, this information would be useful in an audio system using spatialized, or three-dimensional, audio.

One issue the *Touring Machine* brings to the forefront is the intrusiveness of an augmented reality system. Because the visual information is always displayed over the user's own visual field, the visual information could be perceived as intrusive. It could also be argued that this level of intrusiveness is appropriate for this particular task—providing an information-filled tour of a campus. Whether an augmented reality system should be this intrusive or not is dependent upon the application and the intended users, something any designer of an augmented reality system must take into consideration.

## 2.2 *Spirit*

*Spirit* (Uppsala 1997) is very similar to the *Touring Machine*. This unpublished work, performed by researchers at Uppsala University, provides context-sensitive information about locations around a city. The location is determined by a GPS system and transmitted to a central computer. The central computer then responds with appropriate information created by the designers in an audio format. Because this system relies on a relatively coarse notion of location (that is, near “building a” or next to “restaurant b”), the designers were allowed the “luxury” of a relatively inaccurate off-the-shelf GPS system that could work anywhere in the world.

A significant difference between *Spirit* and the *Touring Machine* is the lack of visual augmentation for the user. The context-sensitive information is provided via a

cell phone and a personal digital assistant (PDA). The visual field of the user is not occluded at all. This makes the system appear distinct from the environment, which may be less distracting for individuals.

One issue with *Spirit* is that it transmits the location of the individual to a central computer and then displays this information for others to see. This may seem like an invasion of privacy, but users of the system are aware they are being tracked. However, this knowledge may prevent people from using the system. This is an issue that I struggled with in *Hear&There*—whether to use a central computer or a decentralized system. The difficulty of a decentralized system is the effort involved in updating information across more than one device. This difficulty must be weighed against the potential privacy issues involved.

## 2.3 *Audio Aura*

*Audio Aura* (Mynatt 1995) was one of the first projects to deal almost exclusively with audio in an augmented reality system. The goal of this project, developed at Xerox PARC in 1997, was to provide serendipitous information to users as they travelled through their workplace.

This system provides information triggered by particular locations in the workplace. When passing by a vacant co-worker’s office, for example, the user of the system can tell how long the co-worker has been gone based on abstracted audio cues. The user can also receive an update on the number of e-mails he or she has waiting when standing by the coffee pot.

The use of audio in *Audio Aura* is particularly interesting because it relies primarily on abstract audio. Rather than a computerized or recorded voice, most of the cues are associated with sounds from nature. For example, the number of e-mails waiting may be represented by a flock of seagulls. When only a few seagulls are heard, only a few e-mail messages are waiting. If the user hears a large flock of seagulls, he or she may want to rush to check the large number of e-mails waiting.

Another important contribution of *Audio Aura* was the notion of prototyping.

Along with the augmented reality system that was created, the designers of *Audio Aura* created a Virtual Reality Markup Language (VRML) environment for testing out new ideas and new abstract sounds.

Rather than attempt to determine the precise location of an individual, *Audio Aura* focuses on information that can be associated with pre-determined points in space. Tags are placed at those locations, providing an easy mechanism of determining where the user is in the augmented space. This provides a very convenient mechanism and works well in the application the designers of *Audio Aura* envisioned. The difficulty with this approach is the scalability—every new augmented space requires more tags, and new technology would require the replacement of all of the old tags.

## 2.4 *Automated Tour Guide*

The purpose of B. Bederson's *Automated Tour Guide* (Bederson 1995) was to augment a museum space, so that visitors to the museum could hear pre-recorded commentary associated with a particular work of art. This system took as its inspiration the now common recorded museum tours. These early versions of an augmented reality system provided an audio tape to the visitor. As long as the visitor followed the path and pace the audio required, he or she would be given an augmented tour. However, the old approach did not allow individuals the chance to create their own paths.

The *Automated Tour Guide*, however, allows individuals to choose whatever path they desire through a museum. The visitor carries around a digital playback device that has stored audio commentary about the works of art in a museum. As the visitor passes in front of a piece of art, the appropriate audio is played back from the digital device.

This approach is similar to *Audio Aura* in that it associates augmented sounds with pre-determined locations in space. Again, this approach works well in the museum setting, because the triggers (the pieces of art) are readily identifiable and rarely changing.

This project is relevant to *Hear&There* because it uses the recorded spoken word. In this case, however, the audio is written by professional critics and is highly scripted. It also relies on the relatively static environment of a museum so that the audio is not updated frequently. The focus of *Hear&There*, on the other hand, is audio created by individuals in a changing environment.

# Chapter 3

## Scenarios

Before explaining the details of the project, it will be useful to provide a few high-level scenarios to show the ideas present in *Hear&There*. These are fictionalized accounts, but show many of the design aspects of *Hear&There* in a summary fashion.

As described in the introduction, the essential premise of *Hear&There* is that a physical environment (a courtyard, a park, a house, *et cetera*) has been augmented with audio. Authors have created sounds and “placed” them at particular locations in the physical environment. The sound, now associated with a location in space, can be heard by individuals passing through the physical space as long as they are using the *Hear&There* system. For people not using the system, the environment appears the same as it was before the augmentation—no speakers, wires, or tags are visible or even present. All of the data is contained inside the portable *Hear&There* system. The following sections will show more concrete examples of how this system works.

Note that due to technical limitations described later in this thesis, our implementation only works in the outside area surrounding the Media Lab. However, the concepts created and explored in this project are applicable to any space. With technical improvements beyond the scope of this project (such as improved GPS accuracy), all of the scenarios presented here will be realizable.

### 3.1 Scenario One: Fernanda and the Brazilian Network

Using *Hear&There*, Boston Common has become an augmented audio space. Fernanda decides to travel through the Common using the *Hear&There* system. She dons the headphones attached to small Walkman-like device that contains a GPS receiver and a miniature computer.

Once Fernanda has donned the equipment, she starts off on her augmented reality journey through the courtyard. On the PalmPilot that is connected to the *Hear&There* device, she indicates that she would like to be in “activation network” mode. On the screen appears a list of topics that other people have left behind in the Common. She sees topics such as “The Garner Family’s Thoughts on Boston Heritage,” and “Brazilian Boston.” Being from Brazil, she is immediately drawn to the “Brazilian Boston” topic. She chooses that category on her PalmPilot. On screen, arrows appear pointing in various directions. These indicate the location of a few of the topics directed to Brazilians in Boston.

As she moves her head around, the arrows on the screen change location. She realizes that the electrical compass on her headphones tells the system what direction she is facing, which aids in directing her through the environment.

Fernanda begins to walk through the environment, and notices the length of the arrows changing. As the length of one of the arrows decreases (indicating she is getting closer to the imprint), she begins to hear some audio that someone else has left behind. It takes her a moment to realize the sound is actually someone talking in Portuguese! The sound seems to be coming from a weeping willow tree, although she knows that in reality this is just an illusion. The other Bostonians walking through the area are unaware of the layer of sound that has been overlaid on the environment.

After listening to Brazilian’s description of trees in Boston as opposed to the native trees in Brazil, Fernanda decides she wants to hear more from this author. On her PalmPilot, she chooses “More From This Author.” Now, her screen shows new arrows. These indicate the location of sounds in the environment that are now

accessible to her. She follows the directions to a new imprint in space.

Once she reaches the new imprint and listens to the audio the author has left behind, she sees the category “Brazilian Oral History” on her display. Intrigued, she clicks it, and begins following the directions on her PalmPilot. She is surprised to hear the familiar voice of a famous Brazilian who passed away several years ago, who is talking about ways Boston and his hometown in Brazil have faced similar political struggles. She follows the directions on her PalmPilot to other imprints, and hears audio from many Brazilians who have been to Boston over the years.

After exploring the augmented environment for a few hours, Fernanda leaves. The familiar Boston Common has now become an even more personal environment that she would like to visit more often. She would even like to leave her own contributions to the oral history.

## **3.2 Scenario Two: Karrie the Explorer**

Karrie is a curious person. She has heard about *Hear&There*, but does not feel like exploring a path that someone else has set up for her to follow. Instead, she wants to wander through Harvard Square discovering what tourists, shop proprietors, and local Bostonians have left in the area. She dons the exact same system that Fernanda was wearing in the previous scenario. However, when she starts the system, she chooses “explorer mode” on the PalmPilot. The screen then becomes blank, and Karrie begins to wander around the square.

While she is walking by the garden shop near Brattle Square, she hears the friendly owner discussing his motivations for getting into gardening. She is very intrigued, so she walks inside and begins a discussion with the owner about how he started his business.

After purchasing some flowers, she walks back outside and wanders around, just as she would have if she were not using *Hear&There*. Now, though, she hears the laughter of kids from Tennessee who were visiting Boston three weeks earlier. She continues walking and notices an area near the Harvard Square subway stop where



the sounds are rather cacophonous. It seems as though twenty or thirty people have decided to leave imprints at this exact area. Although she cannot easily make out any individual imprint, she really gets a feeling for the variety of people who have passed through this spot. Looking around, she wonders where all of these people are now, since the area is quite empty.

### **3.3 Scenario Three: Brian the Author**

Brian decides he wants to add his own sounds to the augmented system that is being developed in the Media Lab courtyard. He dons the same hardware Fernanda and Karrie used, except he has the addition of a small microphone.

On the PalmPilot, Brian chooses the “author” option. He is prompted to enter his user name and password. He does, and begins to wander around the courtyard listening to the audio that others have left behind. He decides to record his audio on the bench that he likes to eat lunch at. He sits on the bench and taps the record button on the PalmPilot. He then records his commentary about how much he enjoys eating his lunch in the shadow of the MIT Media Lab and Medical Center.

Once he is done, he taps on the PalmPilot to indicate he is finished recording. The PalmPilot provides him with the options of hearing the audio he just recorded or saving it. If he chooses to save it, he can give it a name and a radius.

Brian then moves to various other locations to record other sounds in the space. Once he has finished, he decides to do some more refined editing of the sounds he created. He walks inside the Media Lab to the *Hear&There* desktop machine in the Sociable Media Group. He loads up the *Hear&There* program, and is presented with a map of the Media Lab courtyard. On the map are colored circles at various locations, each representing sounds that people have recorded. He logs into the system by entering his user name and password, and then finds the sounds he has just finished recording on the map. He makes some fine-tuned adjustments. On one sound, he decides to add a few braids of audio into the edge. On another, he moves the circle representing his sound slightly south of where he had placed it.

He then decides to link the audio that he has created together in a path, so that others will be able to follow it. He creates a “Brian’s Spots” category in each of the circles that he just made and, with a few clicks, adds all of his spots to that category. Additionally, he noticed that some of his friends had recorded audio that was related to the audio he just recorded. He creates another category called “My Friends,” and makes a few more clicks to add his friends’ SoundSpots.

Instead of going back out to the courtyard, Brian decides to try out the new audio by using the *Hear&There* simulation. By dragging the mouse around the map of the Media Lab courtyard, he is able to hear the way the audio would sound to a person walking around the actual courtyard.

Satisfied with his changes, Brian saves them to disk. Now, his audio has become part of the augmented landscape.

# Chapter 4

## The Contributions of *Hear&There*

*Hear&There* is an important addition to the field of augmented reality. It takes inspiration from many previous projects in augmented reality, while focusing on some significant areas that are necessary for moving augmented reality into new realms. These foci are the use of audio, the emphasis on authoring, the notion of linking and guided navigation, and the importance of precision. Adopted early in the design of *Hear&There*, these foci provided guidance in designing the system, and in choosing which features to add as the project developed.

### 4.1 Audio

Until now, most augmented reality research has focused on the use of visual information, with very little attention paid to the audio aspects of an augmented reality system. Therefore, I decided to investigate the use of audio in an AR environment, to explore areas that have not been explored before. There are many perceptual issues and challenges that an audio environment presents. For example, providing a way for the user to orient himself or herself in an audio only space is more difficult than providing visual cues. (That is, it is more difficult for a person to find the center of a “spherical” sound than the center of a visible sphere.)

The challenges, however, provide interesting areas to research. In this project, for example, I used spatialized audio to aid in orienting the user in an audio only

space. Using spatialized audio, the sound the user hears can “appear” to be at any location in 3D space. This exciting technology would have been interesting but perhaps unnecessary in a system that also used video. In this way, I was able to focus on perceptual issues involving audio that will be useful in future work in the AR field.

Another interesting aspect of audio is the fact that audio does not persist. Unlike visual media, such as text or graphics, as soon as audio is heard it disappears. Someone listening to audio must rely on his or her recollection of what has transpired, and must pay enough attention to the audio to glean whatever details he or she wants as the audio is playing. Although the audio may play in a loop, the author can never “hold” or examine the audio as a whole. This leads to some exciting opportunities for an author, as well as some interesting interface challenges.

## 4.2 Authoring

*Hear&There* differs from most other augmented reality systems in its focus on authoring. In general, augmented reality systems designed up to this point focus on the experience of the user traveling through the augmented space. Content creation was left to the designers of the system, or specialized authors. The Touring Machine, for instance, used faceless and authorless content, similar to the content that would be found on a professional web page.

In this way, augmented reality systems have tended to be updates to the familiar audio tours in museums, perhaps the earliest form of “augmented reality.” In these tours, professional art critics provide commentary on works throughout a museum.

*Hear&There* differs significantly from these approaches. From the initial conception of the project, I decided to make an augmented reality system in which anyone could make contributions. Instead of having a professional design all of the content, individuals would be able to add their comments and thoughts in an augmented space. Over time, the individual voices of a community would fill an augmented space, providing a more full view of the community than a professionally authored space would allow.

As a result, instead of only focusing on the technical aspects of navigation through an augmented space, this project required a focus on how individuals would add content to the system. Several important design considerations had to be taken into account on how best to enable community authoring of an augmented space. The first and most important consideration was where the authoring would take place. It is important to provide a way of adding content at the point in the augmented space to which the author was adding information. However, it is also important to provide a more powerful authoring environment that may be too cumbersome to use in the field. A more detailed discussion of these design issues and the approach I used appears in a Chapter 5.4.

Additionally, the notion of allowing authoring in a public space brings up the issue of etiquette. Unlike the web, in which anyone can create a page without impinging on the web page of another individual, the authors in a *Hear&There* environment are sharing a limited resource of a public space. For example, the Media Lab courtyard contains a finite amount of space. As more people add sounds to the augmented space, the courtyard will become crowded and overlapping of sounds may become necessary. Under these conditions, a “space hog” may try to fill the space with his or her own sounds, and a vandal may place several small sounds within another author’s sound area.

These potential problems could be fixed by placing programmatic constraints on the system, by limiting the size of a sound area and by preventing overlap. However, I have chosen to avoid this route so that the notion of etiquette in the augmented space can be explored and to give ultimate control to the authors. To this end, I have created a system of Activation Networks, in which different groups of SoundSpots can be activated or deactivated as the user directs, providing a certain level of filtering. See Chapter 5.2 for a description of this method.

## 4.3 Navigation and Linking

Most previous augmented reality systems used one of two different methods of guiding a user through the content. In one approach, the user was left to his or her own devices in choosing where to travel in a space with no direction provided by the system (as in the Touring Machine and Audio Aura examples). In the other type of navigation system, a user was made to rigidly follow a given pattern, as in the case of the audio-taped guided museum tour.

When designing *Hear&There*, I took the world wide web (WWW) as part of my inspiration. I wanted to create a system in which the authors could link his or her additions to the augmented space to the existing collection of augmentations. Then, after listening to the authors' addition to a space, a visitor could be directed to other augmentations that would be of interest. Just as on the WWW, the visitor would not be forced to follow any particular path. In this way, the visitor is provided guidance, but at the same time allowed flexibility, the best features of the existing approaches to navigation. This approach to designing an augmented reality space led to several challenges, which will be discussed in Chapter 5.2.

## 4.4 Location Accuracy

The goal of most augmented reality systems is approaching a high level of accuracy in determining the user's location in a space. Determining the accurate position of the user is important because the AR produces output to the user based on his or her location in space. Unfortunately, technology has not allowed the accuracy most AR designers would consider ideal. As a result, most systems have made compromises or been designed with the realization that the desired accuracy could not be achieved.

This design issue can be seen in projects like the Touring Machine. The Touring Machine was designed without the need for a highly accurate location. Instead, the project provides AR output on the scale of buildings. As long as the general location of the user is determined, the building the user is near can easily be determined.

Another approach to the current limitations of technology is to make the location of the user context-sensitive. That is, the location of a user is determined as “in room x” or “near painting b.” This can be seen in Audio Aura (Chapter 2.3) and the Automated Tour Guide (Chapter 2.4), in which standing outside a particular room or near a particular painting determined what content was going to be played for the user. Because these systems rely on indicators to be placed in every room or near every painting of interest, they do not scale very well, and are not appropriate for generic uses in which content may not be tied to a significant object.

Accuracy is very important to *Hear&There*, as is scalability. One goal of this project is to allow individuals to leave audio imprints tied to a particular location in space (such as by a tree, on a bench, et cetera). Additionally, I intended to explore some perceptual issues as the user navigates through an audio space, which would only be possible if the location of the user could be accurately determined.

In *Hear&There*, I attempted to overcome this limitation of accuracy by using some of the latest GPS technology. This technology had an advertised accuracy on the order of 1cm, compared with the 10-20m accuracy of traditional GPS systems. The GPS system we use does create some significant design issues that will be discussed in Chapter 5.6, yet it also allows me to explore the possibilities of an AR system with high accuracy. I decided that it was important to begin exploring these issues, so that when the technology improves to the point of providing high accuracy with smaller devices, the groundwork of these issues will have already been laid.

# Chapter 5

## The Design of *Hear&There*

*Hear&There* provided many interesting design issues. Here, I present the design I chose to implement, as well as my reasoning for choosing it.

In the first section, I present the design of the SoundSpot, the basic unit for adding sound to an environment. I then discuss how SoundSpots may be linked together using an approach we termed Activation Networks.

Following this basic design information, I provide a discussion of the two different environments in which *Hear&There* is used (outside in the field, and inside at a computer), as well as the different methods of authoring and playback possible with the system. I have also included a description of the hardware needed for this system.

I follow this up with the navigation methods that we have used and a discussion of the use of spatialized audio in *Hear&There*, followed by a description of the method of storing the augmentation data. Finally, I conclude this section with information on how the issues of security and privacy affected the design of this project.

### 5.1 The SoundSpot

The basic unit of augmentation in *Hear&There* is the SoundSpot. This represents the audio imprint that an individual has left in an area. The definition of the SoundSpot provides the framework for the augmentation of an environment, so it must be defined with great care and reasoning.



For this implementation, I decided that a very basic SoundSpot would help us investigate the issues of augmentation. Therefore, each SoundSpot can be thought of as a sphere with its center at a single location in space (a longitude/latitude coordinate). The sphere has a radius specified by the author (typically on the order of meters). This sphere is filled with a sound that an author associates with it. Because the center of a SoundSpot is defined to be a two dimensional coordinate on the surface of the earth, the SoundSpot can usually be thought of as a hemisphere (with a matching hemisphere below the ground, and therefore inaccessible).

As a user reaches the outer edge of this hemisphere, he or she will begin to hear the sound that the author has chosen. This sound can be anything—spoken word, music, sound effects, *et cetera*. Through the use of spatialized audio, the sound appears to the user to be emanating from the center of the sphere. (See Chapter 5.8 for a more complete discussion of the implementation of spatialized audio.)

In order to give the author more control over his or her SoundSpot, this design calls for a mechanism to “braid” audio into the outer edges of the sphere. The effect of braided audio is to allow several sounds to coexist in one space. Each sound moves to the forefront of the user’s perception for a brief period of time, before receding to the background while other sounds move to the forefront. By providing braided audio, we give the author more creative control over the SoundSpot, allowing him or her to create more personalized imprints.

Braiding provides the benefit of allowing the “main” sound (the sound emanating from the center) to be heard throughout the SoundSpot. On the edges, the sound will be competing with the other braids, but it will still be audible. This provides the user with an orientation mechanism. Because the sound “appears” to be coming from the center of the SoundSpot, the author can use it as a beacon to navigate to the center.

As an example of a SoundSpot, consider the Boston Garden. An author may wish to leave a SoundSpot on one of the small walking bridges in the garden. The central sound could be the author’s ruminations on the beautiful trees around him. He could then add braids of sound to the outer edge of the SoundSpot with his favorite classical

music. In order to completely associate the bridge with his SoundSpot, the author could place it in the center of the bridge, and make the SoundSpot's sphere extend to the bridge's entrances. Now, any user walking across the bridge would first encounter classical music with barely perceptible speech appearing to come from the center of the bridge. As the user continued across the bridge, the music would fade away, and he or she could hear the author's recorded words.

The way I have chosen to define the structure of a SoundSpot is certainly not the only way to allow sound augmentation to a space. The SoundSpot could have been defined to allow any shape, or to use rings of sound instead of braids, or even to allow for a "moving" SoundSpot that floats through an environment instead of being located at one coordinate. However, I decided that the basic SoundSpot layout that has been defined allows for the creation of sufficiently individualized SoundSpots without the need for a very complicated definition. Future work may involve the creation of different definitions of a SoundSpot.

A SoundSpot may be designated as a *StartSpot*. These are special SoundSpots, because they are always present in the environment. In some methods of navigation (see Chapter 5.7), most of the SoundSpots are initially hidden until they are activated (see Chapter 5.2). A StartSpot, however, is a SoundSpot that an author designates as a particularly important or interesting SoundSpot, so that it will always appear. (This of course has the potential to be abused. See Chapter 5.10 for a discussion of how such vandalism is resolved.)

The SoundSpot is a Java class that encapsulates the audio, size, and other characteristics. Each object contains a file handle that points to the appropriate sound file on disk. Currently, the file formats supported are the WAV format and the AU format. All of the SoundSpot objects in the system are stored in a single Vector, serialized, and saved to disk. Whenever any part of *Hear&There* runs, the SoundSpots are loaded into memory and, if necessary, saved to disk when instructed. All of the *Hear&There* components use the same representation of the SoundSpot.

## 5.2 Activation Networks

Although the SoundSpot provides the basic unit of augmentation, there is another authorable unit in *Hear&There*—the activation network. Essentially, an activation network is a group of SoundSpots that an author decides forms a cohesive unit. For example, one activation network may be “Fernanda’s Friends,” which contains all of the SoundSpots that Fernanda’s friends have created. Another activation network may be “Media Lab History” which contains SoundSpots about the history of the Media Lab.

Activation networks provide two major benefits. First, it allows the authors to create a richer environment as they guide users to related imprints in an area. Secondly, it provides a way of managing the inevitable clutter that results in a limited and shared resource.

The inspiration for this notion of linking comes from the world wide web (WWW). On the web, an author can create a home page. On the home page, he or she can link to other home pages around the world. When a visitor views the author’s home page, he or she can follow a link at any point to jump to a new page.

The notion of linking in *Hear&There*, however, has many differences from the linking done on the web. Primarily, the difference is the importance of physical location. As described above, the physical location of a SoundSpot is an important component of the SoundSpot’s design. That is, an author places a SoundSpot in a physical location with some significance to that SoundSpot (such as a personal description of a favorite landmark placed next to that landmark, or musings on the origin of the Media Lab placed at the entrance to the Media Lab). Whenever the visitor wishes to follow a link in *Hear&There*, he or she must physically move to a new location. The *Hear&There* system must provide a way of directing the user to the linked SoundSpots. This is an important focus of this thesis, and will be explored in more detail in Chapter 5.7.

*Hear&There* goes beyond the simple notion of linking by including what we have termed *activation networks*. The author of a SoundSpot can create many links to

other SoundSpots, and organize these links into categories. For example, I could create a SoundSpot with links to imprints related to my fraternity, imprints created by other members of the Sociable Media Group, and imprints about the Media Lab. I would be able to create three distinct categories for links in my imprint.

These categories serve a useful purpose in the environment as a whole—they help to reduce clutter. As indicated in Chapter 5.1, only a few SoundSpots (termed StartSpots) initially appear in the environment under the most common method of exploring an environment. Once inside a StartSpot that has link categories, the user is presented with a list of all of the link categories that the author has created in that imprint. The user can then activate a particular category. At that point, the linked SoundSpots appear in the environment. The user can then wander into those SoundSpots, which may have link categories of their own.

### 5.3 Field Use versus Desktop Use

The *Hear&There* system is designed to be used in two distinct ways. One, which I term the “field use,” is the situation in which a user dons the hardware described in Chapter 5.6 and travels through the real world space under consideration. The other, which I term “desktop use,” is the situation in which an author uses a graphical user interface to edit and create SoundSpots, typically done on a desktop computer (although it could just as easily be performed on a laptop).

Each environment (field and desktop) provides a method of authoring SoundSpots, and a method of hearing the SoundSpots. However, each combination of environment and function provides different advantages and disadvantages.

In the field, playback is of the utmost importance, because it is what provides the actual environment augmentation. Authoring in the field is cumbersome, but it gives the author the ability to create imprints while in the actual environment. (See Figure 5-1.)

On the desktop, authoring can be performed using a GUI, which allows fine-tuning and large scale editing. (See Figure 5-2.) Playback on the desktop is useful



Figure 5-1: The “field” in this implementation of *Hear&There*. Because of technical limitations on the GPS, we limited this project to the Media Lab courtyard.

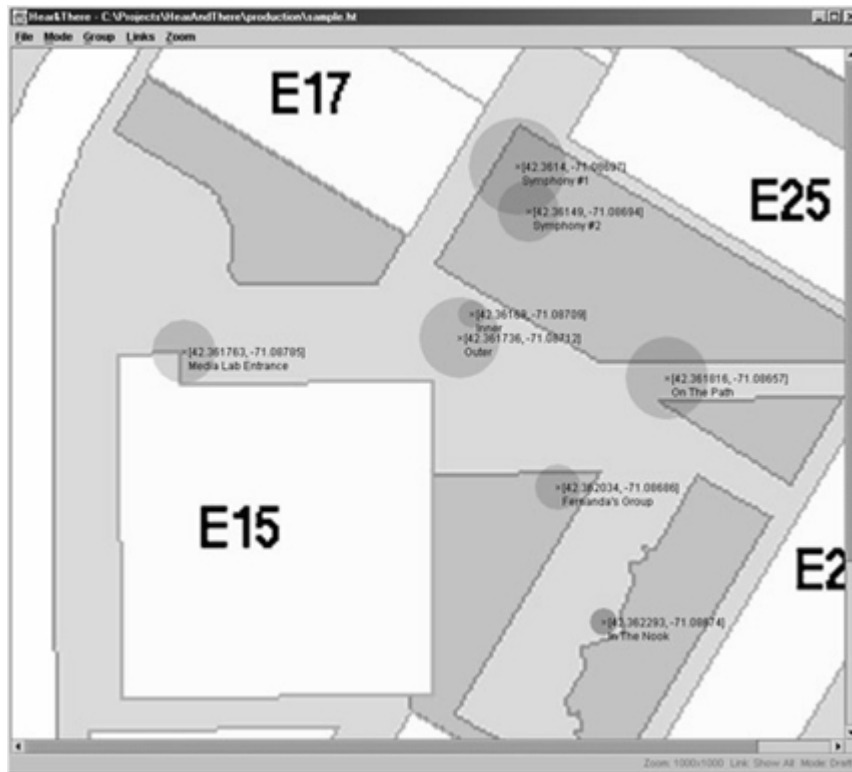


Figure 5-2: The graphical user interface for *Hear&There*. Each circle represents a SoundSpot. This GUI can be used to create and edit SoundSpots, and to simulate the augmented environment.

for simulation so that the author may hear what a modified environment will sound like.

More complete descriptions of the difference between field use and desktop use can be found in the following sections about Authoring (Chapter 5.4), Playback (Chapter 5.5), and Hardware (Chapter 5.6).

## 5.4 Authoring Software

There are two separate pieces of authoring software in *Hear&There*, one for the field use and one for desktop use. Each is written in Java 1.3, including the Java3D API.

### 5.4.1 Field Authoring

In *Hear&There*, the Field Use authoring is primarily intended for “quick-and-dirty” additions to the augmented space. The author dons the Hardware described in Chapter 5.6, and moves about the augmented environment. Whenever the author decides to add a new SoundSpot, he or she uses a PalmPilot interface and a microphone. The author is able to provide basic characteristics for the new SoundSpot, such as its name and radius.

By limiting the characteristics that the user can modify in the field, the authoring system avoids the need for a complex graphical user interface (GUI). *Hear&There* thus provides a means for adding content while in the augmented space, without requiring the additional use of hardware (to display a GUI).

### 5.4.2 Desktop Authoring

Once the SoundSpot has been created in the field, the author can move inside to fine tune all of the SoundSpot qualities, including the addition of braids and the creation of activation networks. This is done in a graphical user interface (GUI), which is shown in Figure 5-2. The GUI contains a map of the augmented space, and graphical representations of all of the SoundSpots in the space.

Additionally, the GUI may be used to create new SoundSpots from scratch, without the need to return to the field. Conceptually, the author need never enter the augmented space.

Note that we make approximations in the GUI so that the spherical earth can be represented on a flat monitor. However, because the system will typically be used for a small area (less than one square kilometer), this approximation approaches reality just as the scale on a city map approaches reality.

## 5.5 Playback Software

The playback software is divided into two parts, one for playback of audio in the field, and the other for simulating the augmented space on the desktop. Both components are written in Java 1.3.

### 5.5.1 Field Playback

The playback software for the field is relatively simple. The software need only compare the location of the user (obtained through the GPS) to the known locations of the SoundSpots. The correct audio file is played at the appropriate volume. (As discussed in Chapter 5.8, the playback of the audio is actually spatialized.)

The software for field playback provides navigation cues to the visitor via the navigation system described in Chapter 5.7.

### 5.5.2 Desktop Playback

In order to provide a complete GUI for the desktop use, *Hear&There* includes a simulation environment. By dragging the mouse over the map, the author can simulate walking through the augmented space. He or she can therefore test how the additions to the space will sound without having to enter the augmented environment.

## 5.6 Hardware

Because there are two different environments used in *Hear&There* (the field and desktop environments) there are two wholly different sets of hardware used.

### 5.6.1 Field Hardware

In order to use *Hear&There* in the field, the user must carry around the necessary equipment described below. Because this equipment is quite bulky, it is mostly attached to a luggage cart that can be wheeled around. We briefly considered using





Figure 5-3: The hardware for *Hear&There* used in the field. The highly accurate GPS receiver is rather large, so we have attached it to a luggage cart.

a backpack, but determined that the equipment would be too heavy for the average person to carry around. Although it would be preferable to have a more lightweight system, we were constrained by current Global Positioning System (GPS) technology (as discussed later in this section). See Figure 5-3 for a picture of the *Hear&There* field hardware.

The system consists of several hardware components. The primary component which coordinates all of the activity is a laptop computer. For this implementation, a Sony VAIO PCG-F390 with a Pentium III processor and 128 megabytes of RAM was used. We required a powerful machine to perform the rapid user position updates and to control many different streams of spatialized audio. As a side effect of the bulky GPS, we were not as constrained to find a lightweight laptop since it could be mounted on the same luggage cart.

The laptop contains all of the sound data and the software necessary to run the *Hear&There* components. The following hardware is attached to the computer to provide the functionality necessary to the system. More complete information about some of the hardware is included in the indicated appendices:

- a GPS unit, which provides the location of the user in longitude/latitude coordinates with an accuracy on the order of 10cm.
- a microphone, for recording the imprints in the field
- headphones, for audio output to the user.
- an electrical compass physically attached to the headphones, to obtain the orientation of the user's head in space. This information is necessary for spatialized audio and some navigation methods.
- a PalmPilot to provide start points and some navigational information, as well as simple commands.

The GPS receiver that we use consists of several components in order to provide a more accurate signal than possible with typical off-the-shelf GPS systems. A dis-

cussion of how GPS works is beyond the scope of this thesis. However, a few details are appropriate to defend the use of such a bulky system.

The main drawback to GPS is its lack of accuracy. Typical receivers provide a location with an accuracy on the order of 10m. This accuracy is fine for a ship at sea, but is not acceptable for a system in which the an imprint can be less than a meter large. (Note that when I began this thesis project, the government deliberately inserted inaccuracy into the GPS system for security purposes. As of May 2000, the Clinton Administration removed this inaccuracy for civilian purposes. However, this alone only improves accuracy to about 5m.)

The Ashtech Z-Sensor system we use consists of a base station and a roaming station. The base station sits on the roof of the Media Lab. It receives signals from the GPS satellites to determine its own location, and after a time it converges on an accurate value. It then calculates an error correction to the signals and transmits this correction to the roaming receiver, using radio waves. The roaming station uses this error correction and some kinetic calculations, along with a signal from the GPS satellites, to arrive at a more accurate location measurement.

The error correction signal from the base station can only be reliably received within five kilometers in urban areas, limiting the area that can be augmented to areas within five kilometers of the base station. Therefore, our implementation has focused on the Media Lab courtyard, an area very close to the base station.

The Precision Navigation brand compass is an electrical compass that provides numerical output between 0 and 360, where 0 is due north. We attach the compass to the user's headphones, so that we can determine the head position of the user.

Both of these measurements (location and head position) are polled ten times a second by the *Hear&There* system. The results are then buffered, and the system then averages the past ten user positions to dampen large fluctuations (which are most likely the result of error).

Note that these components of the *Hear&There* system were selected for the accuracy they provide. In the future, a new GPS system or a better compass could easily be plugged into the system instead. There is nothing magic about the components

that we have selected; they merely provide the highest accuracy that we can achieve today.

## 5.6.2 Desktop Hardware

The requirements on the desktop hardware are very small. Essentially, any modern PC will suffice. Because the software is written in Java 1.3, and uses some advanced graphical and audio techniques, relatively fast hardware makes the software easier to use. (The test machine I used was a Pentium III 500 with one gigabyte of RAM.)

The hardware needs to have a microphone (for recording audio) and speakers or headphones for the simulations.

## 5.7 Navigation

Allowing the user to navigate an augmented audio space is one of the fundamental issues of *Hear&There*. The system is designed to allow a visitor to wander around a real-world space and hear the augmented audio associated with different locations. Here, I present the method we chose to use to lead the user through the *Hear&There* environment.

### 5.7.1 Initiating the Experience

When the visitor starts the system, he or she is presented with two options. One option is the “explorer” option, in which all of the sounds in a space are activated and the visitor is free to wander around. The other is the “activation networks” option. After this option has been chosen, the user is presented with a listing of StartSpots, as described in Chapter 5.1. The author provides a brief (several word) description of the StartSpot. The user can then choose which StartSpot to head towards using one of the methods described in Chapter 5.7.2.

Whenever the visitor enters a SoundSpot (whether using the activation network method or the explorer method), a list of link categories (as described in Chapter

5.2) is provided. The author may choose to activate any of the categories, which will activate the SoundSpots in the desired link category. The user is then provided the locations of the linked SoundSpots as described in Chapter 5.7.2. The author may at any time choose to follow one of the links by simply walking to it.

At any point while exploring, a link category that has been activated can be deactivated. This allows the user ultimate control over their experience in the augmented environment.

## 5.7.2 Two Methods of Exploration

The first step in allowing a path is deciding where to start the path. One possibility is to allow the user to wander around until he or she stumbles onto a sound that he or she finds interesting. This option is unappealing for a few reasons. Primarily, it would make many users feel “lost” for a while as he or she explores the space without any direction. Although this may be desirable to some users in some circumstances, it is probably an uncomfortable situation to most.

Because my design philosophy is to allow as much control to the user as possible, though, I allow the visitor to choose how to use *Hear&There*. He or she can choose to wander throughout the space (explorer mode), or to follow activation networks as described in Chapter 5.2.

## 5.7.3 PalmCompass

In *Hear&There*, the primary use of navigation is to direct the user to SoundSpots in the environment. This involves indicating the general location of SoundSpots in the environment, and also providing the location of specific SoundSpots that the user is interested in finding (such as SoundSpots that have been activated as part of a link category).

To do this, we use a PalmPilot. The navigation screen, which we have termed the PalmCompass, has a circle at the center of the screen to represent the user. Around the circle are arrows pointing to the location of active SoundSpots in the environment.

The length of the arrows indicates the distance to the specified SoundSpot. The direction of the arrows change whenever the user's location or head position change.

Which SoundSpots are displayed on the PalmCompass changes depending on what mode the user is in. If he or she is in explorer mode, then all of the SoundSpots should be shown. If the user is using the activation networks, then only the activated SoundSpots should be shown.

#### 5.7.4 The Use of Audio

One consideration when designing a navigation system for an audio augmented reality system is whether to use *only* audio. There are many benefits and costs to this approach.

A significant benefit of using only audio is that the system would be very accessible to the visually impaired. More than that, environments specifically defined for the blind could be developed (such as a mechanism for providing cues about hazards in the environment).

Even for the unimpaired user, an audio only interface provides some advantages. The user would not have to focus on a textual display, but can instead pay attention to the augmentations in the environment. At the same time, this problematically combines two functions into one interface—the content and the navigation both being provided through audio. It could become quite confusing.

In order to provide the benefits of an audio only interface with the flexibility of visual interfaces, we have attempted to keep the visual interfaces to a minimum. Only very basic options such as the type of navigation desired and the name of new SoundSpots use visual prompts on the PalmPilot. (They are so basic that future work may involve using voice recognition to provide the option input.)

The explorer mode described in Chapter 5.7.2 is an attempt to compromise between the desire for an audio only interface and the benefits of visual cues. In explorer mode, the PalmPilot only needs to be used for the basic options described above. In the activation network mode, however, the visual interface of the PalmCompass described in Chapter 5.7.3 becomes very important.

## 5.8 Spatialized Audio

One issue that we discovered while implementing the system was the need for a better way of navigating through an audio landscape. Without visual cues, the system must rely entirely on audio to indicate the presence of SoundSpots and to direct a user into the center of a SoundSpot to hear the primary audio of the SoundSpot (the audio that the author may have added braids to). In initial implementations, I had several rings of sound that would guide the user towards the center. This method of orienting the user in a sound, though, does not work very well. It works only if the user is already walking toward the center of the sound—the rising and fading sounds provide cues that he or she is on the right track. Unfortunately, if the visitor is passing through the SoundSpot in any other way, the cues can become disorienting.

To alleviate this problem, we chose to investigate the more complex system of spatialized sound. Using spatialized, or 3D sound, a sound can “appear” to be in any location in 3D space. As the visitor moves his or head, the system adjusts the volume and timing of the sound in each ear, which gives the sound a sense of location in real world space.

To accomplish this spatialization, however, the sound emanating from the center must be audible throughout the entire SoundSpot. This is one reason we chose to use a single sphere instead of the rings of sound.

Using the spatialized audio, the user wandering through a SoundSpot will have a very real sense of where the sound is supposed to be located. This is important because the author may have placed the sound at a particular location in space for a reason.

In order to implement the spatialization using Java 1.3, we used the Java 3D API. Although the 3D API is generally used for three dimensional graphics, it does provide a mechanism for spatializing sound. The primary use of the spatialized sound is for virtual environments. Therefore, it was necessary to create a mapping from the real world onto a virtual world for the 3D API. Once this mapping is created, adding sounds and providing the correct spatialization become much more simple.

## 5.9 Storage of Data and Sounds

While designing this system, it became clear that we would need to decide on how to store the data about the SoundSpots (such as position and sound files). Because we were developing the system to be used by many people concurrently, it would not make sense to store the data on the machine the user was carrying around. Although it would probably lead to much faster recall of the data, it would cause a tremendous syncing problem. If one user created a SoundSpot, how would others be alerted about it? Additionally, for large augmented environments, the device would have to have a huge amount of storage. Also, a single *Hear&There* system could conceivably be used anywhere on Earth, but realistically it could only be used in places in which the SoundSpot data had been loaded on the computer.

Therefore, storing the data on a central server that sends the data to the devices when needed would make much more sense. Using simple algorithms, the server could broadcast sounds that the user was about to encounter to limit the lag while sounds download. Additionally, as soon as a user creates a SoundSpot in the field, it can be uploaded to the server and then distributed to all of the devices in the field. Ultimately, this would aid our vision of a universal system, because a *Hear&There* device could be used anywhere a “*Hear&There* server” existed to broadcast local SoundSpot data.

This approach requires the use of a wireless connection to the server, but current technology (in which anyone with a cellular phone can connect to the Internet) makes this a minor point. The more significant problem with this approach is the privacy issues that are created. Because the user’s device would have to transmit its coordinates to the server, it could conceivably be used to track a user in an environment. However, encryption could be used to make sure that only the server has access to this data. Additionally, since the user does not “log into” the system (see Chapter 5.10), the system only knows that “a user” is at a spot, but not the identity of that user.

For testing and development purposes, though, we chose not to implement this



system. Rather, we chose to focus on the basic issues of augmentation using the more simple notion of local data storage. The more complex use of a central server, though desirable, is not central to the notion of augmentation and has therefore been left to future work.

## 5.10 Security and Privacy

An important issue to consider in any shared space is user privacy and the issue of security. These were both important considerations from the outset of *Hear&There*.

The issue of security arises primarily in the authoring component of the system. We had to determine a method of controlling who may edit and create the SoundSpots in an area. This poses interesting problems in *Hear&There* because a limited resource is being used by many people. Over time, the area will become quite crowded, and “outdated” imprints will become clutter. However, allowing any individual to modify the imprints that others have left behind is not acceptable, as intentional or accidental vandalism could become a problem.

To prevent the malicious destruction or vandalism of imprints, we have implemented a very basic authentication system. Whenever a person wishes to begin creating SoundSpots, he or she creates an account with a password. Then, he or she must log in using that account before the system will allow him or her to author or edit a SoundSpot. That person may modify only those imprints that he or she has created. Additionally, there is a “superuser” who can modify the imprints that anyone has left behind, which helps address the issue of clutter (although defining what is “clutter” and what is merely old is a more difficult question).

A more complete system of authentication, involving the assignment of rights to an imprint to groups or particular members, would be more ideal. However, we have chosen to keep the authentication system as basic as possible so that we may focus on more important research issues. Future work on this project could involve the creation of a more complete authorization system.

The issue of privacy is also important in a shared space like *Hear&There*. The

overall goals of the system greatly affect the amount of privacy available. For example, we could have created the system so that an individual could only allow certain other individuals to hear his or her imprints. However, our vision for *Hear&There* is that of a community-based system in which everyone is able to hear all of the imprints. Therefore, individuals using the system must be made aware that the imprints that they create are open to the public. Additionally, making the imprints open to everyone removes the need for non-author users to ever log in, or to even create access accounts.

# Chapter 6

## Implementation

As is the case with most engineering projects, the implementation of *Hear&There* was not entirely smooth. Several problems cropped up along the way, which prevented a complete implementation on schedule. However, we were able to get a significant amount of implementation done to show the worth of continuing the investigation into the ideas of *Hear&There*

Here, I will describe several of the problems we ran into and how they affected the outcome of the project. I will also describe what we *have* implemented.

### 6.1 Problems

#### 6.1.1 GPS

The most distressing problem we encountered involved our GPS system. The first problems were purely bureaucratic—attempts to get the GPS base station on the roof of the Media Lab met with some resistance. The base station, which provides error corrections to the roaming receiver via radio waves, needs to be at a high location. Getting the necessary permission to put the base station on the Media Lab’s roof from the Media Lab Facilities office and the MIT Department of Facilities took several months. Once we received permission, we discovered that we had incorrect connecting cables, and had to spend more time trying to get the correct cables from Ashtech,

the company that makes our GPS system. Once everything was put together, we discovered that there was a significant amount of radio noise on the Media Lab's roof, from numerous other research projects.

Once all of these difficulties were overcome, we were finally able to receive GPS signals and obtain our position with an accuracy of about 2 meters, although the system advertised a sub-meter accuracy. One possible reason is the number of buildings in the area which limited the number of satellites the GPS receiver could "see." We decided to continue using the system that we had, although it would not allow the precision that we initially desired.

In future implementations of this project, it is unlikely that GPS will be used. The initial benefits of it that we envisioned do not counterbalance the low accuracy of the system. (See Chapter 7.2 for a description of one proposed system.)

### **6.1.2 PalmPilot**

Because we could not obtain the accuracy that we desired, the creation of a navigation system on the PalmPilot did not make much sense. The navigation system we envisioned relied on a highly accurate system to display directional arrows for the user. We therefore chose to revert to using a map on the laptop that the user could carry around, showing his or her approximate location. Because we were using a map on the laptop, we moved the entire interface over to the laptop. This is a problem because we initially wanted to make the system as immersive as possible. It is difficult to be immersed in an environment when you are carrying around an open laptop. With a more accurate system, we could move back to the PalmPilot interface.

### **6.1.3 Sound Issues**

While implementing the sound in *Hear&There*, we discovered that Java3D is extraordinarily bulky. Therefore, much time was spent on getting the spatialized sound to work, without adding the additional sound effects that we wanted (such as braided audio). We feel that these are important to the project and should definitely be

pursued in the future.

## 6.2 Actual Implementation

With the above problems, it may seem that very little was implemented. In actuality, we achieved significant, and have laid important ground work for others who are pursuing augmented audio environments.

Using our implementation, people are able to add a sound to the environment at a particular location. Users of the system are then able to hear it while they are roaming around. As long as this augmentation is larger than a couple of meters, our implementation works well. That is, if the author attempts to place a SoundSpot within 10cm of a tree, our system will not provide the necessary accuracy. However, if the author puts a four-meter wide sound at the center of a two-meter wide monument, it will work rather well.

The activation networks were also implemented, allowing authors to tie groups of related SoundSpots together.

We did not attempt a rigorous user test of this system, because we were not able to implement all of the ideas that we initially included. However, informal user testing primarily using the GUI suggests that people enjoy exploring an augmented sound environment, and enjoy adding their own personal augmentations for others to hear. The spatialization worked well, and is something that should certainly be included in future augmented sound environments.

# Chapter 7

## Further Research

There are many interesting questions and issues that have been created by the *Hear&There* project, some of which will require future research.

### 7.1 Questions About Community and Identity

The goal of this thesis was to create an augmented reality system that works. Although they were present to a certain extent, issues of community interaction and individual expression using this system were not fully investigated. Using this augmented system as a framework, the issue of how audio and technology affect and reflect a community can be examined.

### 7.2 Hardware

Because of the limitations of current technology, the system developed in *Hear&There* is rather bulky. As a result, many perceptual issues are difficult to address. For example, what happens when a person moves through the environment rapidly could not be addressed, because of the weight of the equipment and inconvenience of the luggage cart. As GPS technology improves, and as computing technology becomes smaller, the issue of bulk should disappear. At that point, the system could address more perceptual issues, and could become a more widely used system.

In the near future, when GPS technology is unlikely to improve orders of magnitude, it may be desirable to find a more accurate way of locating a user. We initially chose to use GPS because of its universal nature (it can be used virtually anywhere on Earth) and because it does not require any changes to an environment. However, after implementing *Hear&There*, I am no longer convinced that the benefits of such a system outweigh the higher accuracy of other systems.

One approach suggested by my advisor, Judith Donath, would be a combination of tags and accelerometers. Upon walking up to the tag, the system would know precisely where the user is located. Then, using the accelerometer to measure how fast and far the user is moving, it can provide highly accurate results about where the user is in a particular environment. Using this system, the perceptual issues involving audio I have presented in this thesis could be more easily investigated.

### **7.3 Wider Range and Indoors**

Because of the need to use a stationary GPS system on the roof of the Media Lab, *Hear&There* is currently limited to use in the Media Lab courtyard. Moving beyond this GPS technology so that the system can be used anywhere indoors and outdoors is an important next step for this project. Once this is made possible, new issues arise, including the need to consider three-dimensional space (rather than merely longitude and latitude).

### **7.4 Additional Content**

Currently, the augmented realm we implemented consists only of audio created by authors. Other types of audio could be included, such as audio that changes based on environmental conditions, or a continuous broadcast area that broadcasts a radio station into a particular physical space. Additionally, new modes of communication could arise, in which one space is augmented with current sound from another space.

# Chapter 8

## Conclusion

*Hear&There* is a system that provides many important additions to the field of augmented reality. These include the focus on audio (including spatialized audio), the focus on authoring in the space, the notion of linking components in an augmented space together, and the focus on highly accurate location information.

Our implementation fell short of our initial goals, generally as a result of problems surround the GPS system we were using, as well as the bulky nature of Java3D. However, we were able to implement enough to augment an environment with user-created audio and to allow navigation through the environment. We were also able to develop of system of linking audio augmentations together.



## Bibliography

- Bederson, B. 1995. Audio Augmented Reality: A Prototype Automated Tour Guide. *Conference Companion on Human factors in computing systems, 1995*, pp. 210-211.
- Brewster, S. A., Wright, P. C., and Edwards, A.D.N. 1993. An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. *Proceedings of ACM CHI '93*, pp. 222-227.
- Card, Stuart K., Robertson, George, and York, William. 1996. The WebBook and the Web Forager: An Information Workspace for the World-Wide Web. *Proceedings of ACM Conference on Human Factors in Computing Systems, 1996*, pp. 111-117.
- Donath, Judith. 1996. Inhabiting the Virtual City: The design of social environments for electronic communities. PhD Thesis, MIT.
- Feiner, S., MacIntyre, B., Haupt, M., and Solomon, E. 1993. Windows on the world: 2D windows for 3D augmented reality. *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 145-155.
- Feiner, S., MacIntyre, B., Hollerer, T., and Webster, T. 1997. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Proceedings of the International Symposium on Wearable Computers '97*, pp. 74-81.
- Ishii, H. and Ullmer, B. 1997. Tangible Bits: Towards Seamless Interfaces between People, Bits, and Atoms. *Proceedings of CHI'97, ACM, March 1997*, pp. 234-241.
- The Java 3D API Specification v. 1.2. 2000. Sun Microsystems.
- Lyons, K., Gandy, M., and Starner, T. 2000. Guided by Voices: An Audio Augmented Reality System. *Proceedings of the International Conference on Audi-*

*tory Display 2000*, pp. 57-62.

Mynatt, E., Back, M., Want, R., Frederick, R. 1997. Audio Aura: Light-Weight Audio Augmented Reality. *Proceedings of the 10th annual ACM symposium on User interface software and technology*, pp. 211-212.

*Spirit '97*. February 1998. Uppsala University. <<http://www.csd.uu.se/projects/spirit97/>>.

Whyte, William H. 1988. *City: Rediscovering the Center*. New York: Doubleday.