Master of Engineering Thesis Proposal
Hear&There: An Augmented Reality System of Linked Audio

Joseph Rozier

9 November 1999

Abstract

This thesis proposal presents a research project on augmenting a space with audio. Using the system described in this proposal, individuals can leave audio imprints of their identity in a real-world environment for others to hear. Furthermore, audio imprints in the environment can be linked together, similar to the way web pages are linked together today. Previous research in the area of augmented reality is presented, the system I am proposing is described in detail, and several important questions are posed.

1 Introduction

Augmented Reality is a new area of research in which a virtual world is overlayed 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 members of a community 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 particular tree 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 walks in the vicinity of that location, he or she will be able to hear the imprint. In the above example, the author could leave his imprint about a tree near the tree he is describing. Then, whenever someone walked by that tree, 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. Several different approaches to linking will be proposed, a few of which will be implemented.

2 A Scenario

As an example of how this system can be used, consider a courtyard such as the one next to the Media Lab. There are many ways for people in the Media Lab community to tie their sense of identity to this space using audio. A researcher from Brazil, for example, could record her thoughts on how a New England autumn differs from the lack of seasons in Brazil. These thoughts could be located next to one of the trees in the courtyard. Then, to advertise the presence of her recorded thoughts, she could surround this audio with some Brazilian music.

In this way, the sound that the Brazilian researcher recorded has been contextualized to the space. She is not only revealing details about her identity, she is relating how her past experiences are comingling with her experiences at the Media Lab.

Over time, many members of the community could record similar "audio imprints" in the space, perhaps tied to a bench, a patch of grass, or one of the entrances to the Media Lab. Someone walking through the courtyard would hear the audio from a multitude of people, and could get a sense of what type of people work at the Media Lab.

Furthermore, the author of an audio imprint will have the opportunity to link his or her audio imprints to other audio imprints in the space. The Brazilian researcher may want to lead someone listening to her reflections on a New England autumn to her best friend's audio in another part of the courtyard. These sounds could then be "linked" together. Once a visitor has heard the Brazilian's audio, he or she would be made aware of where the linked audio is in the space.

3 Previous Research

Over the past ten years, there has been growing interest in the area of augmented reality. Much of the focus is on video instead of audio. In addition, I have been unable to find a system that proposes augmented linking as I propose in this thesis. However, the previous work is important in getting a sense for what types of systems have worked, and how other researchers approached general augmented reality problems.

3.1 Automated Tour Guide

Benjamin Bederson [1] developed a system that uses audio in an augmented reality system. His primary focus, however, is on the informational aspects of audio augmented reality. The prototype he built is a replacement for the taped audio tours common in museums today. Instead of having a linear tape, the visitor has a digital storage device that plays back sounds associated with a particular location in the museum.

This system uses transmitters above each piece of art on the tour, so that sounds are associated with a particular object instead of a particular location.

The project as presented is also not interactive. The content is created (presumably) by a professional. The museum visitor is the passive receiver of this information.

3.2 Audio Aura

Audio Aura [2] is another augmented reality system with sound. The goal of this project was to provide "serendipitous information via background auditory cues" to people as they move about their workplace environment. Here, the sounds are not really "authored" by an individual, but are actually iconographic representations of events such as receiving e-mail or a phone call. Some sounds are voice recordings, but these are for information purposes (such as indicating who a particular office belongs to).

This system uses "active badges" to determine where a person is. These can only determine proximity to a given location, so the designers must attach devices to the physical objects of interest.

In developing the system, the creators of Audio Aura used a VRML (Virtual Reality Markup Language) representation of the environment to simulate the different aspects of the system on a computer before implementing them in the real world. Although I do not intend to use VRML initially, the notion of a simulation environment to try out new ideas is an important one for my project.

3.3 Touring Machine

The Touring Machine [3], developed at Columbia University, is perhaps the closest system to the project I propose. The goal of the touring machine was to provide a content-rich augmented reality as someone toured an area (in this case, the Columbia campus). As such, content can be placed in arbitrary locations. However, this project is focused much more on visual information. Labels appear over buildings on the campus, and a handheld computer provides contextualized web pages about the different locations.

This project differs significantly from my project in that it does not allow the user to become completely immersed in the environment. Rather, much of the focus of the user is on the supplemental information provided on the computer screen. The user has to shift focus from the environment to the screen, where the bulk of the information is stored. This is not an inherently bad system, but does not address many of the perceptual issues that I am interested in here.

One of the most important similarities betwen my thesis and the Touring Machine is that they both use GPS (Global Positioning System) to track the user. The Touring Machine will provide valuable lessons about the problems surrounding the use of GPS, including how to deal with it's inaccuracy and bulk.

4 Goals of Hear&There

The goals of Hear&There can be split into two major components.

First, I need to create the infrastructure to allow people to augment a space with sound. This includes an authoring component (in which people can leave and edit audio imprints) and a playback component (in which the appropriate sound is played at the correct location). At this level, there are many interesting questions to address. These include how to allow the author to quickly and easily make an audio imprint and the technical issues of how to find the precise location of a user in a space. There are many more details, which will be presented later in this thesis proposal.

Once the basic infrastructure has been created, I will begin to focus on the notion of linking audio together. There are many important questions here, such as how to indicate where imprints linked to a given imprint are located in space, and when an imprint becomes available for the user to hear.

Instead of choosing a single way to answer these questions about linking, my thesis will include a way of categorizing different approaches, as well as the foreseen benefits and drawbacks of each. I will then implement a few of the solutions to see which ones are most intuitive and interesting.

5 Infrastructure Description

Here, I provide a detailed description of what an imprint is, an overview of how imprints will be recorded and played back, a description of the software and hardware that I will be using, and a brief section on the issue of orientation in the augmented space.

5.1 The Imprint

Determining exactly what form an imprint will take is one of the most fundamental questions of this thesis, as much of the development work will arise out of this.

In the initial phase of this project, an imprint will have several rings of sound for the author to manipulate. These are circular areas around the location of the imprint. Whenever the user walks through one of these rings, he or she will hear the audio associated with the ring.

The center ring will contain the primary component of the imprint, such as a recorded message. Around the center, the author can place additional rings of sound. These additional rings of sound will provide cues to the user that there is an imprint nearby, and can also serve to pique the interest of the user navigating through a space.

These rings of sound can be thought of as three-dimensional envelopes of sound surrounding a single point. The volume of the sound can rise and fall within each ring, and the rings can overlap. In this way, as a user walks through the space occupied by an imprint, he or she will experience a dynamic combination of sounds.

It is important to note that most properties of these imprints are adjustable. How the sound rises and falls in a ring, where the maximum volume of a ring is, and even whether a ring exists or not is left up to the author. My intention is not to decide what an ideal imprint would look like and force it on the user. Rather, I intend to provide the underlying structure, and leave the form up to the person creating the imprints. This is similar to HTML (Hyper-Text Markup Language), in which the tools are provided for individuals to make a web page, but it is up to the author to make the web page interesting.

5.2 Recording and Playing Back Imprints

The infrastructure of this system consists of an authoring component and a playback component. This split is similar to the split that exists on the World Wide Web today–the author creates a web page using HTML (Hyper-Text Markup Language), and the user accesses the page with a web browser. In this system, I will create the authoring component to facilitate leaving audio imprints in space and provide enough flexibility to make those imprints unique. (The way the imprints can be made unique are discussed in the next section.) To allow the user to hear the imprints, I will design a system to play back the appropriate sound at the correct volume as the user navigates through a real space.

Both of these components can be accessed outside (by walking around) or inside (through an authoring environment/simulation).

Outside, the user will have a GPS receiver attached to a computer. The computer will have stored on it the location and sound files and real-world locations of all audio imprints. By comparing the current location of the user to the known locations of the imprints, the computer can play back the appropriate audio. The user will also be able to leave an audio imprint using the same system–the computer simply has to be set up to record.

Inside, there will be a GUI (Graphical User Interface) authoring environment. This will provide a map of the augmented space, and a graphical representation of the audio imprints that have already been left. The author will be able to link imprints together, add new imprints, move the existing imprints, change the distance from the "source" that the sound will be audible, and other audio properties of the imprint. Additionally, the authoring environment will allow the user to simulate walking through the augmented space to hear what the audio imprints will sound like.

5.3 Software

Most of the augmented reality system will be implemented in software. The primary language will be Java, although C or C++ may be used in some critical components.

There will actually be two different programs–one to provide an authoring environment and simulation, and one to control recording and playback in the field. These two programs will share a single set of data (the descriptions of the audio imprints).

5.3.1 The Authoring Environment

The authoring environment will be an easy-to-use GUI allowing the author significant control over the placement and playback of the audio imprints in a space. Each imprint in the system will have a graphical representation that can be moved around a map of the environment to be augmented (in this implementation, the MIT Media Lab). The author will have access to existing imprints that were created in the field, as well as imprints created while using the authoring environment.

I have decided to make the imprint highly customizable (just as web pages are highly customizable). Each imprint has several related rings of sounds. The author can control how the sound rises and fades within each ring of sound and how the audio loops over time, among other audio properties. It will be important to provide an interface that provides all of these options to the author without being overwhelming.

The authoring environment will also serve as a 2D simulation of the augmented space. By dragging a representation of a person through the augmented space, the author can hear what a user will hear at any given location. This provides a chance for the author to try out any new imprints and layouts without having to go into the environment.

5.3.2 In the Field

The playback of the sounds in the field will be very similar to the simulation environment at its core. To focus the attention of the user on the audio, however, I will avoid using a GUI.

The software in the field should, in fact, be invisible to the user. To make a truly augmented space, the user should feel as though the sounds actually belong in a given location. The fact that the computer is doing a simple comparison between the user's location and the imprint's location and then playing back an audio file is something the user should not think about at all.

Allowing a person to record an imprint in the field is an important goal of this project. Because the audio imprints are contextualized to the space, it is reasonable for a user to create an imprint while standing or sitting in that space.

To make adding imprints to a space as easy as possible, the interface for recording a sound in the field will be very simple. As an initial idea, the user will have a microphone with a "record" button. By pressing the record button, the user can make an imprint. The imprint will have have default characteristics (such as volume ramping) that can be adjusted later using the more complex GUI environment.

5.4 Hardware: GPS

Determining the location of an individual is one of the most important tasks in this project. I have decided to use a GPS system because it allows for arbitrary positioning of sounds. That is, there is no need to associate sound with a physical object such as an IR tag or a radio beacon. Using GPS, a sound can be placed over water, on a busy street, or at an arbitrary location in a field.

There are a few problems with GPS, though. One is accuracy. It is very important for this augmented space that the location of the user be known with precision on the order of centimeters, instead of the meters that a standard GPS receiver provides. The reason is that the audio imprints will be closely connected to the space. The author will be given precise control over where an imprint is placed and how large an area the sound covers.

As an example of why accuracy is important, consider the case of a man who wants to leave his thoughts on a particular tree in the Boston Public Garden. If the precision is only 10m (which is typical for an off-the-shelf GPS), the author cannot place his imprint directly in front of his favorite tree. Rather, he can only guarantee that his imprint will fall in the vicinity of that tree, and it may even end up in front of a completely different tree (perhaps his least favorite tree).

There are several systems that provide better accuracy, however. One such system is the Z-Sensor system from Magellan. This system employs a base station and a mobile receiver. Both the base station and receiver obtain signals from the GPS satellites. Using a set of calculations (the details of which are beyond the scope of this project), the system is able to remove much of the error that affects single-unit GPS receivers. Additionally, the Z-Sensor uses a proprietary technique to expose the encrypted government signal (which is much more accurate than the civilian signal).

Because of these features, the Z-Sensor provides an advertised accuracy of 2cm. This is sufficient for this project.

The downside of the Z-Sensor system is its bulk. Together, the "mobile" components weigh approximately 15 pounds and include a meter long antenna and a second, disk shaped antenna. This bulk makes the system less than ideal. For this project, however, the important questions can still be answered with the Z-Sensor.

5.5 Orientation

An important issue to consider in this augmented reality environment is how a user listening to the audio will orient himself or herself. This can actually be divided into several parts. First, how will the user become oriented inside a particular imprint so that he or she can find the center (where the primary sound is located)? Second, how can the user orient themselves in an environment to find the audio imprints?

For the first issue, the initial implementation will use rising and falling volume to indicate nearness to the center. As a user walks through an imprint, the sound in each ring will increase to a point and then fall away. This may be enough information to lead the user to the center of the imprint.

If I find that the volume does not provide enough of an orientation cue, I will investigate the use of spatialized audio. Using spatialized audio, a more realistic 3D representation of sound can be created. This will, however, require knowing the orientation of the user's head.

The second issue is something that will be a focus of my thesis. Providing a means of indicating where linked imprints are and directing the user to them is a large question. In my thesis, I will propose several ideas and implement some of them.

6 Scope

The implementation of this project will be limited to the MIT Media Lab courtyard for practical reasons. These include the range of the GPS system that we use and the need to choose an appropriate map for the authoring environment. However, this system will be developed in such a way that it can be used anywhere with an appropriate GPS setup.

Additionally, I will be desiging the system so that it is not reliant on GPS. Using the object-oriented approach of Java, I will allow any system that can provide relative coordinates to be "plugged into" the system.

7 Software and Hardware Required

This system will use the following software and hardware components:

- Java 2, JDK 1.3
- Audio Software for creation of sound files
- Desktop computer for Authoring/Simulation System
- Notebook computer for In-Field System
- Palm computer for simple In-Field UI
- Magellan Z-Sensor GPS for location tracking
- Microphone for recording audio
- Headphones for In-Field listening
- Backpack for carrying In-Field components

8 Schedule

As of early November, I have completed much of the work on the authoring environment, including imprint creation, linking, and simulation. Because there have been delays in installing part of our GPS system on the roof of the Media Lab, I have not yet been able to test the system in the field.

Following is a list of work still to be completed, organized into a month-by-month schedule.

- *November 1999.* Make the GPS system functional and experiment with it; calibrate the system to use the actual GPS measurements; write a basic program for playback in the field
- *December 1999.* Investigate spatialized audio; investigate coding and communication on the PalmPilot/Visor unit
- *January 2000.* Begin categorization of links; begin implementation of one linking method.
- *February 2000.* Finish implementation of one linking method; begin implementation of second linking method.

- *March 2000.* Finish implementation of second linking method; start field trials of the system with real users; begin writing thesis
- *April 2000.* Work out bugs in the system; continue writing thesis.
- *May 2000.* Finish thesis.

9 Conclusion

I have presented my idea for an augmented space that allows users to leave audio imprints for others to hear and to link them together. I presented several issues that are important to the infrastructure of the project, and laid the groundwork for approaching the issue of how imprints in an augmented space can be linked together.

10 Acknowledgements

Judith Donath is my thesis advisor on this project and, as such, has provided guidance in choosing an appropriate topic and helping me to narrow the scope of my research.

I have worked closely with Karrie Karahalios, a Ph.D. student in my group, on the infrastructure for this project. We have had frequent meetings to discuss the technical approach we will take in this project. Additionally, she has provided much know-how on the hardware issues of GPS and other types of sensing that may become necessary as the project develops.

References

[1] Bederson, B. (1995) Audio Augmented Reality: A Prototype Automated Tour Guide. *Conference Companion on Human factors in computing systems, 1995*, ACM, pp. 210-211

[2] 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*, ACM, pp. 211-212

[3] 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