Guided by Voices: An Audio Augmented Reality System

Kent Lyons, Maribeth Gandy, Thad Starner
College of Computing
801 Atlantic Drive
Georgia Institute of Technology
Atlanta, GA, 30332 USA
+1 404 894 3152
{kent, maribeth, thad}@cc.gatech.edu

ABSTRACT

This paper presents an application of a low cost, lightweight audio-only augmented reality infrastructure. The system uses a simple wearable computer and a RF based location system to play digital sounds corresponding to the user's location and current state. Using this infrastructure we implemented a game in the fantasy genre where players move around in the real world and trigger actions in the virtual game world. We present some of the issues involved in creating audio-only augmented reality games and show how our location infrastructure is generalizable to other audio augmented realities. according

Keywords

Audio, augmented reality, wearable computing, context-awareness

INTRODUCTION

This paper presents a lightweight and inexpensive infrastructure for augmented realities that uses a simple wearable computer. Whereas most traditional augmented reality systems overlay graphics onto the user's environment, this system employs only audio. Furthermore, we have created a positioning infrastructure using radio frequency (RF) transmissions that is easy to deploy, provides approximate location and uses little computational overhead. As a result, the system we created is much less expensive and computationally complex than traditional augmented reality (AR) systems. "Guided by Voices" is an AR game developed to use our prototype infrastructure that allowed us to explore the unique issues involved in creating an engaging augmented reality audio environment. We first present other systems that use personal audio in the environment followed by the infrastructure we created for our audio AR. We next describe the prototype application and describe the user's experience. Finally we present some of the issues involved in creating sound for an audio-only AR and show how our infrastructure can be generalized for use in other augmented reality systems with varying levels of user interaction.

RELATED WORK

Guided by Voices explores a different design space than previous audio augmented realities. Bederson [2] describes a related nonlinear sound playback system triggered by location using infrared beacons. The sounds presented provide detail on real life artifacts the user observes. Our system extends this concept to utilize state information on a wearable device to play different sounds for different situations at the same location.

Nomadic Radio [5] utilizes sound to inform the user of incoming messages and events. It further permits the user to control the audio's intrusiveness level. Audio Aura [4] explores several different soundscapes for displaying ambient information. Both systems provide a hint of the information content and significance through ambient sounds. They try not to burden the user or deflect the user's attention away from his/her current task unnecessarily while presenting information. In contrast, Guided by Voices explicitly engages the user in the audio environment through narration, sound effects and ambient audio. It exploits player location and history to play contextually appropriate sounds and uses lightweight and inexpensive wearable computers that require minimal explicit user input.

AUGMENTED REALITY INFRASTRUCTURE

The augmented reality system consists of a positioning system and wearable device that plays contextually appropriate sounds. Users of the system carry a simple wearable computer that receives location information from transmitters placed in the environment. Each transmitter sends out a unique ID over a limited range creating a small cell. The wearable then uses the cell's ID to infer location. The current location is used in conjunction with the list of previously visited locations and the order in which they are visited to play contextually appropriate sounds from the augmented reality.

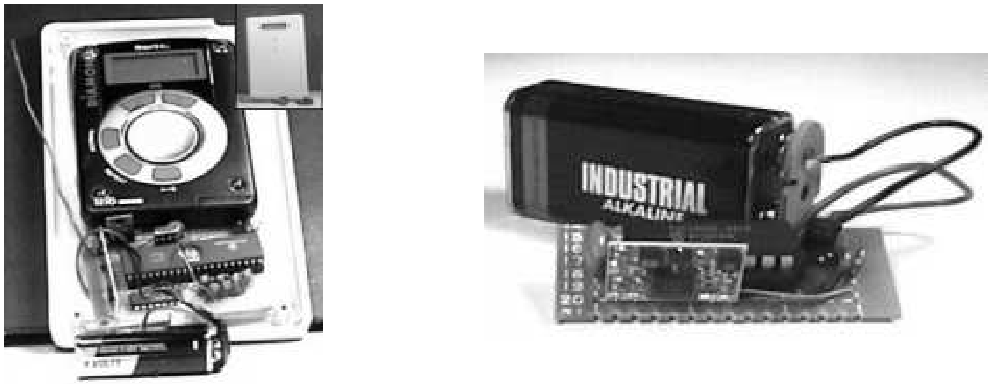

Figure 1: **Left** - The inside of the wearable device showing the Diamond Rio^{TM}, the RF receiver, the PIC microcontroller and a 9V battery. The inset is the closed unit. **Right** - The RF transmitter attached to a 9V battery.

The users carry a simple wearable computer that is responsible for tracking the user's location and controlling the playback of sounds. The wearable consists of an off-the-shelf MPEG 1 Layer 3 (MP3) Diamond Rio^{TM} digital audio player, a Microchip PIC microcontroller, and a RF receiver (Figure 1 Left). The PIC decodes information from the RF receiver into location information and controls the MP3 player, including selecting the correct track and starting and stopping the device at appropriate times. The PIC is responsible for tracking the user's position over time and updates the user's state in the augmented reality. That information is then used to select and play appropriate sounds stored on the MP3 player.

The location beacon consists of a PIC and RF transmitter (Figure 1 Right). Each beacon transmits a unique ID approximately once every second. The range of the transmitters was deliberately limited to create a small cell with a diameter of between 5 to 30 feet that largely depends on antenna configuration. The beacons are small, low powered and inexpensive. Currently each beacon should last approximately one year on a 9V battery but eventually we hope to environmentally power the beacons to avoid batteries altogether. We plan to deploy the beacons as an economical infrastructure for general use to gather location information for a variety of wearable devices [6].

The use of the the beacon system greatly reduces the complexity of gathering the position of the user for the augmented reality. Other techniques such as vision or magnetic tracking devices typically require much more computational power, are more expensive and cover a smaller range. Other beacon systems have been created using infrared (IR), but they require either line of sight, limiting the area covered by a beacon or flood an entire area increasing the cell size to a whole room. Using RF we eliminate both problems and we are able to create a relatively small cell.

The largest problem with RF beacons from a technical standpoint is the error caused by noise and interference. It was found that the MP3 player generated a lot of interference while playing a sound, but very little performing its other functions. Collisions of transmissions between beacons is another source of error. Although the system is designed such that transmitters are spaced apart into cells, there are bound to be areas of overlap when trying to get complete location coverage. To minimize this type of interference the beacons use the transmitter for as little time as possible. Currently a single byte provides more than enough unique IDs. Most error detection or correction techniques are designed for systems that send much more data (on the order of kilobytes) and typically they either double the amount of data sent, or add on extra bytes with error detection information. In either case using these techniques would greatly increase the amount of data being sent and correspondingly increase the chance of a collision between transmissions. In our system we found that the simple heuristic of waiting to update position until the same ID was received twice was sufficient to block most errors. Once we deploy this system for general use we will have to increase the ID space and then more general error detection/correction schemes may become a better option.

THE GAME

To demonstrate our lightweight infrastructure we decided to implement an augmented reality game. A game was easy to prototype because we could change the rules of the game to adapt to our developing infrastructure. The game also engaged the user focusing the player's attention on the AR experience and not the devices used for implementing the augmented reality.

The setting we chose for Guided by Voices is a medieval fantasy world where the player wanders through a soundscape of trolls, dragons, and wizards. Various sounds from the setting are presented as the player explores the environment passing through a dragon's lair and a troll's bridge, meeting friends or foes and gathering objects like a sword and a spellbook. The game play is similar to traditional role-playing games such as Dungeons and Dragons^{TM} (D&D) where the player is a character on a quest. As in D&D the player does not directly interact with computers or graphics. Instead, the wearable observes the player's actions and functions as the game master. All of the game's input comes from where the player chooses to walk in the real world environment. The narrative of the story changes depending on the paths of the players. The wearable keeps track of the player's state checking if he/she has acquired the necessary objects for a specific location. Rich environments can be created because each location can serve multiple purposes and play different roles in different situations.

One could look upon this audio-only environment as being less immersive than a traditional computer game with graphics. However, we see several advantages to the Guided by Voices game play. For one, instead of moving an avatar in the computer world the player physically moves through a real world space. This aids in immersing the player in the game play. Also, many people can play this game simultaneously in the same physical space creating a real life social aspect to the game. Unlike most multi-player PC based video games, Guided by Voices brings people together to the same geographical location. The game could be extended easily to capitalize and enhance this social interaction. The audio-only nature of the game takes advantage of the player's imagination drawing them further into the game. For example, instead of being presented with a rendered dragon that may or may not be frightening, the player gets to interact with his/her idea of what a frightening dragon should look like. The sound elements of the game are designed to aid this imagination process.

GAME PLAY

To play Guided by Voices the user first puts on the headphones and the wearable. The player then enters the game space, which is simply an area where the RF beacons have been set up (convention hall, gallery, computer lab, etc.). For this particular game we used 10 beacons creating 9 different locations in the augmented reality. There is also one beacon that the player passes as he/she enters the space that triggers a sound clip of the game narrator. This clip explains to the player the backstory of the game and how to play. The player is told that he/she is about to embark on a quest to rescue his/her elfin friend who was kidnapped by an evil wizard. The narrator goes on to explain that the player will be exploring a medieval fantasy landscape and will encounter both friends and foes as he/she travels through this world. Along the way he/she will find virtual items that will help to battle enemies and release the imprisoned elf friend. The narrator explains that to play the game, the user must simply walk through the gamespace, visiting the sites of RF beacons that are also marked in the real world by small props that hint at what will be encountered at that site.

As the player enters a new site the wearable receives the unique ID from the beacon in that location and updates its internal position information. The current game state, location information and rule of the game are then used to select and play the appropriate sound clips and in turn update the game state. For example, the player starts the game with no items. If he/she walks over to the beacon marked by a stuffed rat, he/she will hear a narrative sound clip that indicates that he/she has encountered a rat and has picked it up and has placed it in his/her inventory. Later on, if the player walks over to the beacon representing a knight one of two narrative sound clips will play. If the player has previously procured the rat the narrative will play out such that the player slays the knight and steals his sword. The player can then continue his/her quest having successfully defeated a enemy and having won another item vital to the completion of the total quest. However, if the player has not procured the rat the narrative will play out such that the knight defeats the player. The narrator then explains to the player that he/she must start the quest again. In another instance the player may find the beacon that represents the cage where his/her friend is imprisoned, but will not be able to virtually unlock the door until he/she has procured the magic spellbook from another site. The state diagram (Figure 2) shows all the beacons that are present in the Guided by Voices game space. Below each space is the logic that controls the sound clips that will be played and how the game state will be effected when a player comes within range of the beacon. There is more than one winning path through the set of beacons, ranging from a path that consists of visiting only three beacons to one that visits all nine beacons.

One interaction problem that exists in the current implementation of Guided by Voices results from the range of the RF transmitters used at each site. These transmitters have a wide range compared to the small props at each location and the boundary of the transmitter is not apparent to the user. Therefore, it is not clear to users when they will be in range to trigger the sounds for a particular site. This can result in users inadvertently entering sites they had intended to avoid. One way to fix this problem would be to modify the transmitters to have a much smaller range so that the site is not entered until the user is right next to the beacon and prop. However this would require developing a new radio transmitter instead of using a commercially available one. Another solution that would also result in richer game play would be to incorporate time into the system. Currently the game state is updated once the player enters a new space and accordingly sounds are played immediately. The game could be redesigned so that the player is first informed that they have entered a new space. The player could then choose to leave with nothing happening, or he/she could stay triggering the system to update the game state and play the sounds for that location.

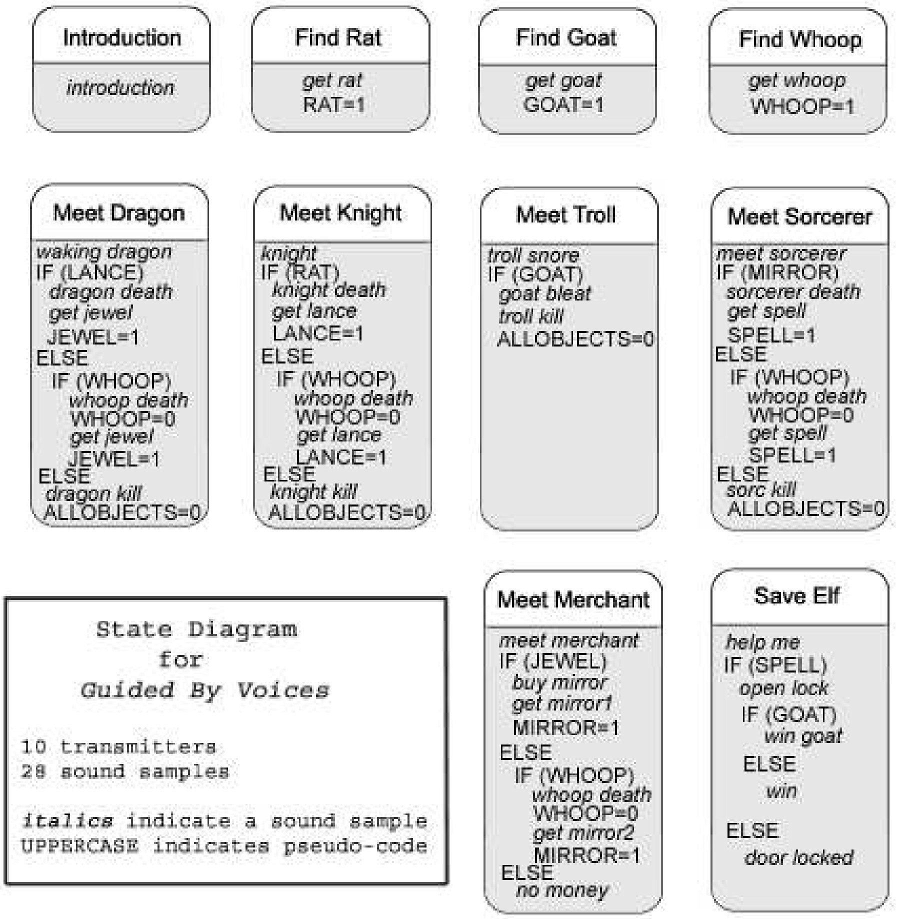

Figure 2: State diagram for Guided by Voices. Each box represents a beacon in the environment. The text indicates the sounds played and how the state of the augmented reality is used and updated.

For example, if the player entered the site that contained the dragon narrative, he/she would first hear the heavy breathing of the dragon and the cave ambiance. If the player remained in the range of the beacon he/she would wake up the dragon and end up fighting it. However, if the player quickly retreated from this location the narrative would stop and the player would have avoided the encounter with the dragon.

SOUND DESIGN

There were several challenges to be overcome by the sound design for this game. Our first goal in the sound design was to address the key issues of creating a narrative audio space. This involves answering the questions: Where am I? What is going on? How does it feel? What happens next? [1]

When creating the sounds for this environment "real" sounds were not appropriate. It is not enough to simply record a sword being drawn from a sheath. Instead, a sound effect must match the listener's mental model of what a sword should sound like [1]. This is especially important in this game which lacks visual cues. It is known that if a sound effect does not match the player's mental model, no matter how "real", he/she will not be able to recognize it [3]. Therefore, creation of a caricature of the sound is important [1]. The setting of this game is also in a fantasy world so there were seldom "real" sounds available. Instead we tapped into the cultural conditioning of the player to make the narrative clear.

Another design challenge lay in the fact that the game is a nonlinear narrative. This could cause for a weakening of artistic control due to the fact that the user is in control of the narrative flow [1]. Special care was focused on content design so the story's elements would be consistent throughout the various paths of the game. A player's actions weave small narrative elements (sound events) together to form a larger story. An additional potential obstacle to audio design is that this game might be played in a noisy environment through which the player is trying to navigate at the same time as play the game. The sounds are therefore especially clear and exaggerated, and all speech is well enunciated.

Each sound event has three layers: ambient sounds, sound effects, and narration. We feel that for the game to be immersive, entertaining, clear and not frustrating all of these elements are necessary.

- Ambient Sounds: The ambient sounds at each location give the player an impression of the fantasy setting they have just entered, from the echoing water drip of a cave, to the wind and animal sounds of a forest setting. These ambient sounds need to be loud enough to be heard by the player but should not overwhelm the sound effects or narration. As a result, the ambiance is often faded out slightly when a character speaks.
- Sound Effects: The sound effects present in a sound event range from a knight riding up to the player, to a dragon blowing a deadly fireball. The sound effects also include the speech of other characters in the game world. The speech and other effects were modified to reflect the acoustics of the fantasy setting and exaggerated to match the player's mental model.
- Narrator: Guided by Voices uses the narrator to reinforce the elements stemming from ambiance and effects. The narrator keeps the gaming experience from being frustrating by explaining events that are not obvious from the sound effects alone. For instance, when a player dies in the game, the narrator informs the player what object would have helped him/her overcome the foe for the next time they play. Likewise, the narration makes it clear to the player what actions have taken place, such as killing an enemy or picking up an object. It is important that the narrator voice be distinct from the characters in the game. Therefore, the narrator speech is not subject to the acoustics of the setting and is louder and consistent throughout the locations. This makes the narrator into an omnipotent ever-present character.

GENERALIZING THE INFRASTRUCTURE

The prototype location infrastructure can be configured in several ways that facilitate different forms of interaction between the users and the environment. By controlling the placement of transmitters and which devices record beacon IDs, various levels of privacy can be achieved.

1. No state: Sounds are determined by user location and neither the wearable nor the environment records any information.
2. State only on wearable: The wearable records the presence of other transmitter's IDs and the order they were received. By mounting transmitters on players as well as in the environment, multiple players can interact in the same augmented reality.
3. State in the environment and on the wearable: The environment records the IDs of transmitters that pass nearby and the beacons can therefore respond differently to individual users. This configuration allows the users to share one virtual world if all of the environmental beacons are networked together permitting more complex and rich interactions.

In the first system the audio is directly triggered by the current location. The system used for Guided By Voices is of the second type but has no transmitters on the wearable. Each user can share the same physical space, but the augmented realities are completely separate. Placing transmitters on the wearables would allow the users to interact in the virtual world linking the personal augmented realities together. An example using this additional infrastructure would by to extend Guided by Voices such that there is a location that requires two players to cooperate to get a particular item. Likewise, by using the third type of infrastructure and having the environment track the users, Guided by Voices could be extended to allow locations to have a memory of who visited over time and across users. If one player's actions in a location destroyed a monster, when another player visited they would encounter the remnants of the battle. This infrastructure would further link the augment realities of the users and extend the abilities to cooperate and compete with other users.

The location information provided by the positioning infrastructure poses some privacy concerns for users of the augmented reality. With the current Guided by Voices configuration only the environment transmits IDs and all location information is local to each wearable. By mounting transmitters on other wearable systems (the second configuration) the users of the augmented reality can track their proximity to other users. Although the location of a wearable is no longer private to each person, only the other people physically present at the same time and place receive the other user's location. In the third configuration the environment tracks the users position and records that information. This poses the greatest privacy concerns because other people could potentially query the environment about a user's location and a great deal of effort would be needed to make sure that the environment is secure and trusted. For the last two configurations that have a transmitter on the wearable the user could, in theory, selectively reveal himself/herself to the environmental infrastructure if such exposure would add desired functionality. Selecting the correct type of infrastructure to support augmented realities must weigh the concerns about the privacy of user's location against the added functionality achieved.

CONCLUSION

We have created a low cost, lightweight solution for audio augmented realities and a prototype role-playing game that utilizes and demonstrates this infrastructure. Our sample application required special attention to the construction of the sound environment. We found that three layers of audio (ambient sounds, sound effects and narration) are important in creating the immersive augmented reality game. Although we have used this system to implement an immersive gaming environment, the infrastructure we created has many possible uses. It can be configured to support applications that range from simple reactive audio tour guides to rich systems that allow for complex interaction between the users and environment.

ACKNOWLEDGEMENTS

Thanks to the AudioLab for the help with the creation of the great sounds and the Broadband Institute for their support.

REFERENCES

1 Back, M. and Des, D. Micro-Narratives in Sound Design: Context, Character, and Caricature. *Proceedings of ICAD '96*, November 4-6, 1996.

2 Bederson, B. Audio Augmented Reality: A Prototype Automated Tour Guide. *Proceedings of CHI '95*, May 1995, pp.210–211.

3 Mynatt, E. Designing with Auditory Icons. *Proceedings of ICAD '94*, November 7-9, 1994.

4 Mynatt, E., Back, M and Want, R. Designing Audio Aura, *Proceedings of CHI '98*, April 1998.

5 Sawhney N. and Schmandt, C. Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging. *Proceedings of CHI '99*, May 15-20, 1999.

6 Starner T., Kirsch D., and Assefa S. The Locust Swarm. *Proceedings of ISWC '97*, October 13-14, 1997.