Designing a Narrative-Based Audio Only 3D Game Engine

Timothy Roden *, Ian Parberry
Department of Computer Science & Engineering
University of North Texas, P.O. Box 311366, Denton, Texas 76203
http://www.cs.unt.edu

roden@cs.unt.edu, ian@cs.unt.edu

ABSTRACT
Immersing players in believable and engaging virtual environments is a common goal for many interactive computer games. While PC-based audio only games set in virtual worlds have traditionally been developed for blind players, new technology will soon create a market for audio only games aimed at a mass audience. We propose a framework for authoring interactive narrative-based audio only adventure games set in 3D virtual environments. Our work builds on several years of research into audio only applications for sight impaired users, augmented reality systems and human-computer interaction studies. We argue that a simple user interface enhances both immersion and entertainment value, making audio only games practical for mobile computing. Novel features of our system include real-time gameplay and support for multiple players. We also describe a software architecture for creating audio only games, the current implementation of which uses low-cost existing PC-based hardware and software.

Keywords
Audio, Game Engine, Virtual Environment, Speech

1. INTRODUCTION
Since the introduction of commodity 3D-accelerated graphics cards, the evolution of graphics hardware and software has progressed at an increasing rate. Despite this, entertainment applications using real-time computer graphics are still far from producing images that approach the complexity found in reality. In contrast to graphics, the fidelity of audio found in current commodity audio cards is remarkably true to life.

Another computing trend is the increased use of wireless devices. Cellular phones and other mobile devices have become an important platform for gaming [9].

*Acknowledges support from Creative Technology Ltd., http://www.creativelabs.com

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ACE2005, June 15-17, 2005, Valencia, Spain.
Copyright 2005 ACM 1-59593-110-4/05/0006…$5.00.

There can be little doubt as to the compelling nature of audio only entertainment. Pre-television era radio drama was a widely successful form on entertainment, proven by its commercial success. In the same way that radio made audio only entertainment possible, mobile computing heralds the evolution of audio only entertainment to the level of interactivity.

1.1 Related Work
The software industry has produced audio only games for the blind [1], and research has studied how blind children interact with these games [13]. Other studies have explored how audio enhanced games can be played by both sighted and blind users [11]. Augmented reality games, in which players navigate a virtual world and typically interface with the game using a wearable computer, have been the focus of much research [8, 10]. Similar research has studied how a desktop user interface to an audio only game can be enhanced using hardware devices such as head trackers [7] and software techniques such as auditory 3D pointing devices [12]. At least one study used a simple desktop interface to an audio only game that allowed users to roam a virtual 3D space [5]. Games using narratives have been studied [5, 10] as have non-verbal games using only sound effects [15].

2. OVERVIEW OF OUR APPROACH
We are interested in creating an interactive narrative-based audio only game set in a 3D virtual environment. The game should appeal to a wide audience that includes both sighted and non-sighted players. The game should incorporate all three basic auditory elements: (1) speech for communication and to transmit knowledge, (2) music to express and/or trigger emotions, and (3) sound to depict objects or processes [12]. We choose a narrative approach since narrative is a well understood and effective story-telling technique. Our goal is to develop a narrative-based audio only game that takes advantage of low-cost, current and near future 3D audio technology. At the same time we do not want to overuse the technology in an attempt to make the game more immersive at the expense of making user interaction more difficult. We believe inhibiting efficient user interaction leads ultimately to a less immersive and therefore less entertaining experience.

No special hardware, such as a head tracker [7, 12], should be required to play the game. A sound card capable of producing reverberation effects is recommended for improved realism [14]. Speaker configuration can consist of stereo headphones, stereo

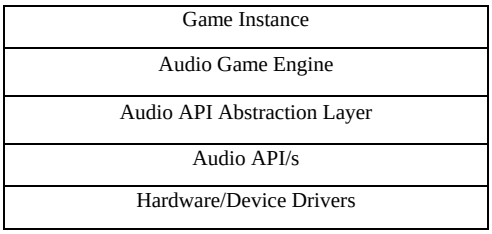

Figure 1. Layered Software Architecture.

desktop speakers or optionally, multi-channel speakers. When using desktop configurations, a sub-woofer for low-frequency (bass) effects is highly recommended as are speakers capable of generating sufficiently high volume. Both an increase in volume and the use of bass have been shown to increase the immersive nature of audio [6]. User interface can be either keyboard and voice or voice alone by utilizing speech recognition software.

The game is authored and played using a graph paradigm. Each node in the graph represents a unique location in the virtual world. The size of the world is not limited by spatial constraints unlike augmented reality games in which physical restrictions on player movement limit the size of the world [8, 10]. An advantage of audio only games is they enable perception of an environment beyond the visual range [12].

We do not allow the player to explore the interior space represented by a node since research has shown that too much freedom of navigation can result in players becoming disoriented [5]. Also, players can have difficulty locating where they are using only 3D audio [14]. We do not want players to feel like they are blind explorers in the virtual world. Instead, we want players to have a clear mental model of where they are at all times. Our primary goal is to immerse players in the narrative.

Figure 1 shows our overall software architecture. At the lowest level, hardware is made available to the Application Programming Interface (API) layer either directly or indirectly using drivers supplied by the hardware manufacturer. The API layer consists of one or more industry standard interfaces. The Audio API Abstraction Layer (AAAL) insulates the game engine from changes to the API layer, and in our current implementation, interfaces with several existing industry APIs. The Game Engine is a data-driven core set of routines that enable different instances of the game to be executed. With this basic architecture we can design and test audio games using PC hardware and later transfer our work to a mobile platform by rewriting the necessary code in the AAAL.

3. WORLD REPRESENTATION

The virtual game world is authored by editing a 'world' data file. Currently, this is a plain text file. We encode the world in memory using a graph. Each node in the graph represents a unique location in the world. There are several general categories of nodes: outside, indoor and underground. Figure 2 shows a portion of a world graph with non-player characters (NPCs) and items contained at various location nodes. The basic information contained in a node is a text description and a set of sounds to be played. Each node also contains environmental properties that enable sounds to be customized according to the particular acoustics at each location in the virtual world. We leverage the capabilities of current PC sound cards to enable dynamic playback of sounds using custom environment settings.

At each location in the world, individual sounds have several attributes including type, time-of-day, volume, rate, and 3D position. There are currently four types of sounds: ambient, random, entry, and exit. Ambient sounds play continuously in a loop and are used both to transmit knowledge about the virtual environment and to express or induce emotion. Ambient sounds include background noise and/or music. Music is primarily used for emotional impact but can also be used to sonify a change in light intensity [12]. Care must be taken to ensure ambient sounds transition smoothly from the end of the sound track back to the beginning. We recommend ambient sounds be of sufficient length so as not be overly repetitive. Dynamic techniques can also be used to mask repetition such as varying the pitch, volume and position over time [3]. Ambient sounds should also be played at a lower volume relative to other sounds to both mask repetition and since a variety of other sounds can play concurrently, including the narrator.

Random sounds play at intervals randomly determined according to their rate. Entry sounds are played once when the player enters a location. Exit sounds play when the player leaves a location. Transition sounds should be used as much as possible and are particularly effective to alert the player to dramatic changes in the narrative [12]. A time-of-day attribute is used to vary the way sounds are played at locations depending on the time of day. Currently, we use only three categories: day, night, and all. Each sound has an associated volume ranging from inaudible to full volume. Finally, sounds can be positioned in a variety of ways. Surround sounds play either in stereo, if using headphones or a dual speaker setup, or in 5.1 surround sound if using a suitably equipped speaker system. Single channel (non-stereo) sounds can be placed in 3D using fixed positions such as front, left, right, rear. Random placement is also possible.

Each sound can also be enhanced with one acoustic filter chosen from; obstruction, occlusion and exclusion. Obstruction is used to simulate an object partially intervening between the sound source and listener. Occlusion simulates a sound source that is completely obstructed with no direct air pathway between the sound and the listener, such as a sound placed in an adjacent room. Exclusion is similar to occlusion except a direct air

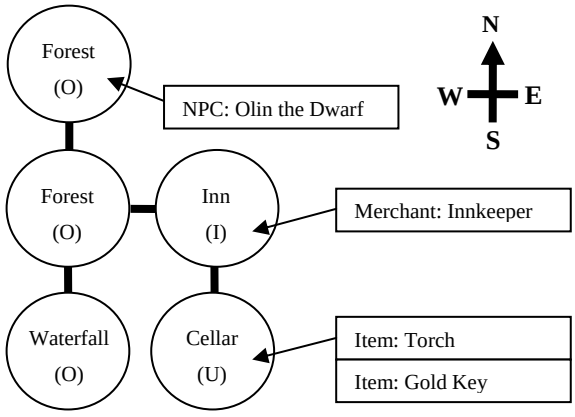

Figure 2. World Graph.

pathway exists, such as a sound in an adjacent room where a door connecting the rooms is open. Filters can be further customized by choosing the type of material the sound transmits through. Example materials are wood, stone and cloth.

Interactive adventure games are much more entertaining when the player has the ability to affect the environment. In particular, the player should be able to pick up, drop and transport a variety of items encountered during the course of the game [10]. To enable this, the game maintains an inventory of items carried by the player. In addition to the player's inventory, one or more items may exist at each location in the game. Items can be static or dynamic. Static items cannot be taken by the player. Dynamic items may be taken, carried and dropped - possibly in a different location. Items are further grouped into containers and items. Containers are always static and may hold one or more items. For example, a container might be a chest that contains items of treasure.

Items can have associated sound effects. These sounds can further be subdivided into sounds that are audible when the item is present at a location in the game and sounds that are audible only while the item is being carried by the player. In general, however, items should not make sounds when carried by the player. Since many items will not make any sounds at all, the narrator is primarily responsible for describing both the presence and characteristics of items to the player at the player's request. Dynamic items exist primarily to be used in some manner during the game. For example, a key may be used to open a door. Items can have a limited or unlimited number of uses. For example, a flask of water might only be used a few times before being exhausted and removed automatically from the player's inventory. When authoring a game, items are placed in the same data file with world location information. A separate 'items' file details specific attributes for both common and unique items.

The user can interact with a variety of NPCs in the game. Like items, NPCs are initially placed in the world via inclusion at the proper locations in the world data file. A separate 'npc' file contains specific attributes and data for NPCs. NPCs can be static or dynamic. Static NPCs remain in the location where they are placed while dynamic NPCs may move throughout the virtual world. Dynamic NPCs can roam freely throughout the world or be restricted to certain types of locations. NPCs are further divided into merchants and non-merchants. Merchants offer a variety of items for sale. While individual characteristics of each merchant may vary, such as voice and personality, it is important the trading sequence be the same for all merchants. In this way the player can intuitively interact with merchants that are newly encountered in the game. Most interaction with a merchant is in the form of yes or no answers on the part of the player. This offers the potential of a speech only interface using suitable speech recognition software.

Each location in the world can contain, at most, one merchant or NPC. NPCs represent important personae in the virtual world. While they do not offer items for sale, they can provide the player with useful information in addition to adding a variety of dramatic elements to the narrative. When the player enters a location containing an NPC, the NPC approaches the player automatically if the NPC has information to convey to the player. Optionally, if an NPC does not have new information to provide the player, the NPC can approach the player with a random greeting. Information provided by an NPC can be triggered using a variety of in-game data. Examples include location, time-of-day, player inventory and player exploration history. Exploiting player history can be a particularly effective technique to advance a narrative and has been used in other similar systems [10]. Both merchants and NPCs can optionally be flagged to only appear during specific times of the day.

4. SOUND STAGE

We use the metaphor of a stage to position sounds for playback. While we want to stimulate the auditory senses into believing in the virtual world, attempts at true 3D spatialization have proven to reduce recognition [4]. Instead we divide all audio into two categories for the purpose of spatialization. Background audio is generally non-speech audio and includes ambient sounds, music and sound effects. Most speech is defined as foreground audio and includes the narrator, merchants, NPCs, and other human players in a multiplayer game.

Background audio is allowed to play as naturally as possible. Ambient sounds are played in stereo for headphones and dual speaker configurations or in 5.1 surround sound. Other sound effects can be placed randomly in 3D space or in fixed positions such as front, rear, right and left, all relative to the user. Because of its importance in advancing the narrative, foreground speech is always played at fixed positions in the sound stage, all within a 120 degree field. The narrator is heard directly in front of the user. In a dual speaker configuration this results in both speakers playing the narrator. As shown in Figure 3, merchants and NPCs are positioned at a 60 degree angle to the right of the user. In a multiplayer game, a second human player voice is positioned at a 60 degree angle to the left of the user. In addition, we adopt the method proposed in [4] of placing speech in a ring topology on the horizontal plane formed by the speakers. That is to say, speech is placed neither above nor below the listener in 3D.

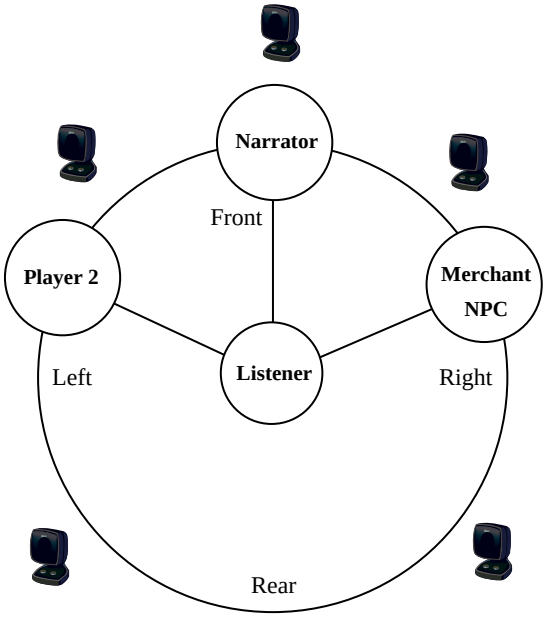

Figure 3. Sound Stage with 5-speaker surround sound configuration shown.

5. THE NARRATOR

In a narrative audio space there are several key questions to be answered such as: Where am I? What is going on? What happens next? [2]. We place three technical requirements on the narrator: 1) The narrator must be clearly audible at all times; 2) the narrator must be easily distinguished from the other characters and personae in the game, and; 3) the narrator's voice should not convey emotion since the narrator's responsibility is simply to provide information. We accomplish these goals in the following ways. First, we place the narrator's voice directly in front of the user. Also, we borrow the technique of enhancing the volume of the narrator relative to other sounds and we do not subject the narrator's voice to the acoustics of the virtual world [10]. We use a synthetic text-to-speech voice that is distinctive from other speech in the game. The synthetic voice contrasts sharply with the pre-recorded human voices of merchants and NPCs. Another benefit of using a synthetic voice for the narrator is the synthetic voice is generally unemotional. Unlike previous generations of speech synthesis software that have a characteristic mechanical sound, current software is more life-like and supports multiple voices, both male and female.

6. USER INTERFACE

The game's user interface is speech only or a combination of speech and keyboard depending on several factors including; availability of a keyboard, installed speech recognition software, and personal taste. The user can interact with both the game and a second human player. Interaction with a second human player is done using voice only. If both players are present in the same location we allow them to converse naturally. Interfacing with the game can be done either with a speech interface or a keyboard.

Using a speech only interface, a situation can occur when the user is in the same location as another human player and wishes to talk with the other player and also speak one or more commands to the game. We handle this in two ways. First, we do not use voice activated input. Instead we require the user to manually hold down a key while speaking. A separate key is used to speak to the game and another key is used to speak to the second human player. In addition, we do not allow the second player to hear commands given by the user or the responses from the game resulting from the commands. Similar to [8] we consider user commands local to the user in contrast to other activity such as dropping an item which we consider a global sound effect which would be heard by all players at the same location.

7. EXPLORING THE VIRTUAL WORLD

Movement in the virtual world is accomplished using a compass as shown in Figure 2. At each location there are potentially six directions the user can move: north, south, east, west, up and down. When describing adjacent locations, the narrator makes clear the direction of each adjacent location. Using a six-direction compass simplifies the layout of the world and allows the user to maintain a clear mental model of the surroundings. A limited degree of movement also scales well to input devices with fewer keys such as a cellular phone keypad. When moving between locations we use footsteps sounds that are matched to the environment. Sounds in the two locations are cross faded for the duration of the movement which is a few seconds [2, 5]. Acoustic environmental reverberation properties can also be morphed if appropriate audio hardware is installed. Time passes in the game regardless of user actions. The environment reacts to changes in time by using different sets of sounds. In addition, the availability and location of merchants and NPCs can be affected by the time of day.

8. CONCLUSION

The Interactive Audio Special Interest Group (IASIG) has recently formed the IASIG Mobile Audio Working Group to discuss issues involving interactive audio production for cell phones and other mobile devices including mobile audio games. We believe a key to the success of mobile audio only games will be incorporating the wealth of past research, much of which is chronicled in this paper.

9. REFERENCES

[1] http://www.audiogames.net. Website, 2004.

[2] Back, M., and Des, D. Micro-narratives in Sound Design: Context, Character, and Caricature in Waveform Manipulation. *Proceedings of ICAD 1996*, Nov. 1996.

[3] Boer, J. "Dynamic Variables and Audio Programming," in Game Programming Gems 4, Charles River Media, 2004.

[4] Crispien, K., and Fellbaum, K. A 3D-Auditory Environment for Hierarchical Navigation in Non-visual Interaction. *Proceedings of ICAD 1996*, November 1996.

[5] Drewes, T., E. Mynatt and M. Gandy. Sleuth: An Audio Experience. *Proceedings of ICAD 2000*, April 2000.

[6] Freeman, J., and Lessiter, J. Hear, There and Everywhere: The Effects of Multichannel Audio on Presence. *Proceedings of ICAD 2001*, July 2001.

[7] Gaye, Lalya. A Flexible 3D Sound System for Interactive Applications. *Proceedings of ACM CHI 2002*, April 2002.

[8] Hadrup, R., et al. Designing an Auditory W-LAN based Game. *Proceedings of Mobile Entertainment: User-Centered Perspectives*, 2004, pp. 207-214.

[9] Lasky, M., "The Business of Mobile Games," in *Game Developer*, November 2001, pp. 36-37.

[10] Lyons, K., Gandy, M., and Starner, T. Guided by Voices: An Audio Augmented Reality System. *Proceedings of ICAD 2000*, April 2000.

[11] McCrindle, R. and Symons, D. Audio Space Invaders. *Proceedings of the Third International Conference on Disability, Virtual Reality and Associated Technologies, 2000*, pp. 59-65.

[12] Rober, N., and Masuch, M. Interacting with Sound: An Interaction Paradigm for Virtual Auditory Worlds. *Proceedings of ICAD 2004*, July 2004.

[13] Sanchez, J., et al. AudioBattleship: Blind Learners Collaboration through Sound. *Proceedings of ACM CHI 2003*, pp. 798-799.

[14] Shinn, B., and Ram, S. Identifying Where You Are in a Room: Sensitivity to Room Acoustics. *Proceedings of ICAD 2003*, July 2003.

[15] Targett, S., and Fernstrom, M. Audio Games: Fun For All? All For Fun? *Proceedings of ICAD 2003*, July 2003.