# Sleuth: An Audio Experience


**Thomas M. Drewes, Elizabeth D. Mynatt**
College of Computing
801 Atlantic Drive
Georgia Institute of Technology
Atlanta, GA, 30332 USA
+1 404 385 1102
{drewes, mynatt}@cc.gatech.edu

**Maribeth Gandy**
Interactive Media Technology Center
250 14th Street NW, Suite M-14
Georgia Institute of Technology
Atlanta, GA, 30332 USA
+1 404 894 7594
maribeth@imtc.gatech.edu



## ABSTRACT

We outline the design, prototyping, and evaluation of an immersive audio game, *Sleuth: An Audio Experience*. We examine related applications based on wearable computing and augmented reality technologies, and discuss the game audio design and user interaction issues. The game environment was prototyped using VRML 2.0 and Java.



### Keywords

Immersive audio game, virtual environment, sound design, user interface


## INTRODUCTION

Envision an immersive game in which you play the role of a detective, gathering evidence in a murder case. The setting is a dinner party at an imposing mansion and the victim is the host and owner of the estate. All of the guests are suspects and have been detained on the premises. Somewhere in the many rooms of the house is the murder weapon. It is up to you, the investigator, to determine the murderer, the weapon, and where it happened in the house. This concept is the basis for the classic board game, *Clue**.

Now, imagine removing all visual aspects of the game. The events still unfold as the detective navigates the rooms of the house, but now the investigator is surrounded by a world of sound. Clues are derived from overheard conversations of the guests and the ambience of the house itself. This concept is the basis for the design of a new audio immersive game, *Sleuth: An Audio Experience*.

The design of this game has implications for both research in audio-centric user interfaces and the design of audio-based applications. From a research perspective, audio display offers a powerful and fluid medium for feedback in user interfaces. Certain aspects of audio, particularly human perceptual characteristics are well understood. Other aspects, such as cohesive design principles, are much less formally defined. A few fundamental characteristics of human auditory perception and selective attention have motivated our design. Spatialized audio perception and display allow the placement of sounds relative to the listener's location [11]. The well-known *cocktail party effect* describes the ability of a listener to selectively attend to sound sources based on differences in spatial location, pitch, intensity, and other characteristics [3]. Additionally, a listener can process background sounds to which he or she is not directly attending, such that changes in these background sounds can draw a listener's attention [4]. However, audio design is more of an art than a science, and sound designers tend to acquire their skills in sound creation, layering, and storytelling over years of experience. We will reflect on our design process in an attempt to better inform the intuition of sound design, and we will discuss the results of a user evaluation of the game.

## RELATED WORK

From an application perspective, new technologies such as wearable computing are creating venues in which audio display may prove useful as an interaction medium. Though the current state of wearable computing display technology precludes visually convincing augmented reality games, it does provide an infrastructure capable of enabling audio-based augmented reality games, such as Guided By Voices [8].

The Nomadic Radio project has explored spatialized audio as an organizational metaphor for the display of audio cues [10].

The use of spatialized audio for tasks in the Nomadic Radio application such as browsing and scanning is complementary to our use of spatialized audio for navigation in a virtual space.

Audio is also a useful medium for a user to maintain awareness of actions taking place in a remote or virtual space. The Sharemon [4] and Arkola [5] work examined awareness of background activities through audio cues in single-user and collaborative workspaces, respectively. Since the action in *Sleuth* takes place in a virtual world with no visual component, sound design for awareness of events in the virtual world is essential.

The work on Audio Aura, describes a system for lightweight audio display of context-dependent ambient information [9]. The audio events in this system were typically triggered based on location. In *Sleuth*, the narrative events and ambient sounds are also location-based, revealing more pieces of the narrative as the user moves from room to room in the virtual environment. Audio Aura also provides a case study in the value of ecological sound design for ambient or background sounds.

## GAME OVERVIEW

A typical interaction with the game begins with the user hearing a short narrative that describes the scenario leading up to the murder and establishes the user's role as detective. Names of the guests, the list of possible weapons, and the names of the rooms in the mansion are provided on paper to act as the detective's notebook. A floor plan of the rooms in the house is also provided, but without details of the contents or names of the rooms.

The detective then navigates from room to room, pausing to listen when something draws his or her attention, and perhaps moving towards sounds of interest. In this manner, the detective attempts to explore the house and gather enough evidence to make an accusation. Various ambient sounds and conversations between the other guests in the house serve both as evidence in the case as well as providing navigational clues to the detective. For example, if the character Professor Plum is known to have slept soundly through the night, then it is unlikely that he is the murderer. In another situation, the sound of billiard balls clicking against one another and the banter of the players would suggest that the detective had entered the game room.

Once the detective decides that enough information has been acquired to deduce the perpetrator, the detective declares a murderer, a weapon, and a location. If the guess is right, the detective has won the game. If the guess is incorrect, the detective will have to revisit the rooms of the house to gather more evidence.

## DESIGN ISSUES

As we began our design of the system, several key design issues emerged: cue design and identification, avoiding information overload, basic navigation design, navigation feedback, and designing the interactive narrative.

### Cue Design and Identification

A simple, but essential element of sound design involves the ease of identification of a given sound or combination of sounds. Human listeners are adept at creating stories to explain individual sounds or sequences of sounds—even stories not intended by the sound designer [4]. In some cases, a sound by itself may be quite ambiguous in meaning. Put in context with other sounds, including narration or dialog, the meaning of the sound can be made clearer. As we selected the audio elements of our soundtrack, we noted which sounds had the greatest potential for ambiguity and tried to put them in a contextual relationship with other sounds to clarify the meaning.

For example, as the avatar enters one of the rooms, a distant gunshot and voices can be heard. Moving closer to the corner of the room with a window triggers a further sequence of sounds: the scraping sound of a window opening leads to a quick fade up of another gunshot, outdoor ambience, and the dialog of two characters in conversation during practice on the skeet shooting range. The combination of sounds is intended to establish a cause and effect relationship between approaching the corner of the room and opening a window, as well as establishing that the conversation is outside of the house.

### Avoiding Information Overload

A naïve view of audio design might suggest that "more is better." The value of a rich and thoroughly detailed sound environment cannot be overstated—sound design for a Hollywood film frequently makes use of a dense, layered soundtrack to support the accompanying imagery. However, a listener who is also viewing the primary imagery can more easily interpret a dense soundtrack. Human visual dominance over the auditory system helps ensure that we turn to our visual system to disambiguate unclear or complex audio stimuli [12].

Since our system lacks visuals, careful attention was paid to the layering of elements in the soundtrack and the pacing of narration and ambient presentation. For example, when creating room ambience, it was necessary to balance the salience of the background audio with the clarity of the dialog between characters. The ambience of the game room is perceptible enough that anyone familiar with the sound of billiard cues and balls in action will recognize it. At the same time, the general volume and timing of these sounds do not detract from the primary dialog, which also serves to reinforce the context for the billiards ambience. When Miss Scarlet, a character in the game, mentions a "nice shot" or asks whether it is her turn, this offers some redundant information that can help clarify the nature of the room. Ideally, a combination of well-designed ambient cues and understandable character dialog will lead the detective to the correct conclusion.

**Basic Navigation**

The initial version of the system utilized the default avatar navigation techniques present in the Virtual Reality Modeling Language (VRML) browser, CosmoPlayer. These consisted of mouse-based inputs to translate the avatar in two degrees of freedom and rotate it in a third. These techniques allowed the avatar to take on a continuous range of positions and orientations. Though these techniques provided great flexibility in moving the avatar through the environment, they also were intended for a three-dimensional environment in which there would be a large amount of visual feedback. In an audio-only environment, the large range of possible orientations and positions within the environment coupled with the limitations of human spatial audio perception made navigation quite difficult [11]. Typical problems involved users getting lost in the space and some difficulty in navigating to specific objects in the space.

We surmised that limiting the range of possible values for each degree of freedom might help alleviate these problems. Our solution was to limit each degree of freedom in navigation to discrete values. Orientation was limited to ninety-degree rotations, effectively restricting the avatar to face north, south, east, and west in the coordinate system of the world. Forward and backward movements of the avatar were set to a fixed step size. The combination of these techniques created a game board grid, with each grid cell representing a valid position for the avatar. As a result, the avatar can only move to adjacent, orthogonally situated grid cells.

Control of the avatar was accomplished with the arrow keys on a standard keyboard. The up arrow key mapped to a move forward one grid cell, and the down arrow key mapped to a move backward one grid cell. The left arrow key mapped to a left turn of ninety degrees and the right arrow key mapped to a right turn of ninety degrees.

**Navigation Feedback**

We sketched out general floor plans for the environment, including ambience and character sound placement. Auditory icons were used to reflect reality to some degree as well as provide useful navigation feedback [6]. These included sounds made by the avatar, such as footsteps, in addition to other naturalistic sounds in the surroundings. A primary concern was the overall density and position of stationary sound “landmarks”. Given an adequate number of sufficiently distinctive and unique sound landmarks, we hoped the user would be able to use these to discern the relative location of characters and rooms in the house.

In a typical virtual environment including visuals, feedback concerning position and orientation would be indicated by the visuals themselves. In our design, it was necessary to add feedback about the result of an attempt to move the avatar. We included a footstep sound that is played whenever the avatar moves from one grid cell to another, and we included a short scuffing sound to indicate a turn. We also created an “oomph” exclamation sound to indicate that the avatar had tried to walk through a wall. We realized that even with this audio feedback it was difficult for users to remember their orientation from the ambient and landmark cues alone, so we explored a few other methods of feedback describing orientation.

Changes in avatar position and orientation were achieved through brief animated transitions. The motivation behind these transitions was to give the user a better sense of the avatar's spatial changes over time as a result of the user’s navigational input. As the avatar turns, for example, the user hears the audio pan in a manner consistent with the turn, and as the avatar moves forward, sounds in front of the avatar become ramp up to a louder level while sounds behind the avatar become ramp down to a softer level.

**Interactive Narrative**

Designing the interactive narrative component is perhaps the most difficult aspect of game design [7]. Since users may access each set of sound cues in a fairly arbitrary order as they traverse the rooms of the house, care must be taken to ensure that every collection of audio cues in a particular room tells its own simple story. These simple stories must work together to contribute to the larger narrative [2].

*Sleuth* fuses elements of an interactive narrative with concepts drawn from exploratory and puzzle-solving games like *Myst** and *The 7th Guest**. Audio plays a critical role in several of the puzzles in *Myst*, such as providing the combination to open a lock, and serving as a sort of "audio compass" in an underground maze. *The 7th Guest* makes use of audio-visual cinematic cut-scenes to reveal pieces of the story of a haunted house as the user solves various puzzles. *Sleuth* is very exploratory in nature, but retains a cinematic, incremental storytelling quality in its audio vignettes. The ultimate goal of the game, however, is to solve a logic puzzle through the process of deduction. Through exploration of the audio environment of the house, the user acquires the information necessary to solve the puzzle. To motivate this exploration, there must be an engaging and well-crafted story at the heart of the game [7].

## IMPLEMENTATION NOTES

We implemented the environment using VRML 2.0, which supports basic spatialized audio dependent on the capabilities of the underlying hardware. In our prototype system, the output hardware was an Aureal Vortex 1 stereo soundcard that supported the A3D spatialized sound standard [1]. Spatialization with this set up seemed to be best achieved with a pair of closed-ear, high quality headphones. Using a VRML modeling tool, we created the room layouts and positioned the audio source objects. These objects represented room ambience and character dialogs. Proximity sensor objects in the VRML world were paired with the audio source objects to enable triggering of the sounds as the avatar entered different regions of the rooms. The characteristics of the audio source objects included the sound sample to be played, the direction, and the size and shape of the sound field that determines the attenuation of the sound over distance.

Though navigation was initially controlled via the built-in interface of the CosmoPlayer VRML browser, we eventually opted to utilize the External Authoring Interface (EAI) in VRML 2.0. The EAI provides an application programming interface (API) that allows a VRML world to be manipulated through method invocations in the Java programming language. This API gave us greater control over the interaction techniques that could be used with the VRML world.

## EVALUATION

Our goal in this evaluation of *Sleuth* was to examine the effectiveness of our design decisions in a qualitative manner. Seven subjects were asked to play *Sleuth*. Subjects were given a checklist of suspects, weapons, and rooms, and were asked to annotate a paper map of the space. The map showed only the relative layout of the rooms and doors and did not name the rooms. Subjects were asked to play until they felt they could make a guess based on their deductions; each game typically lasted one hour. We interviewed the subjects, with a focus on walking through the clues they used to make their guesses.

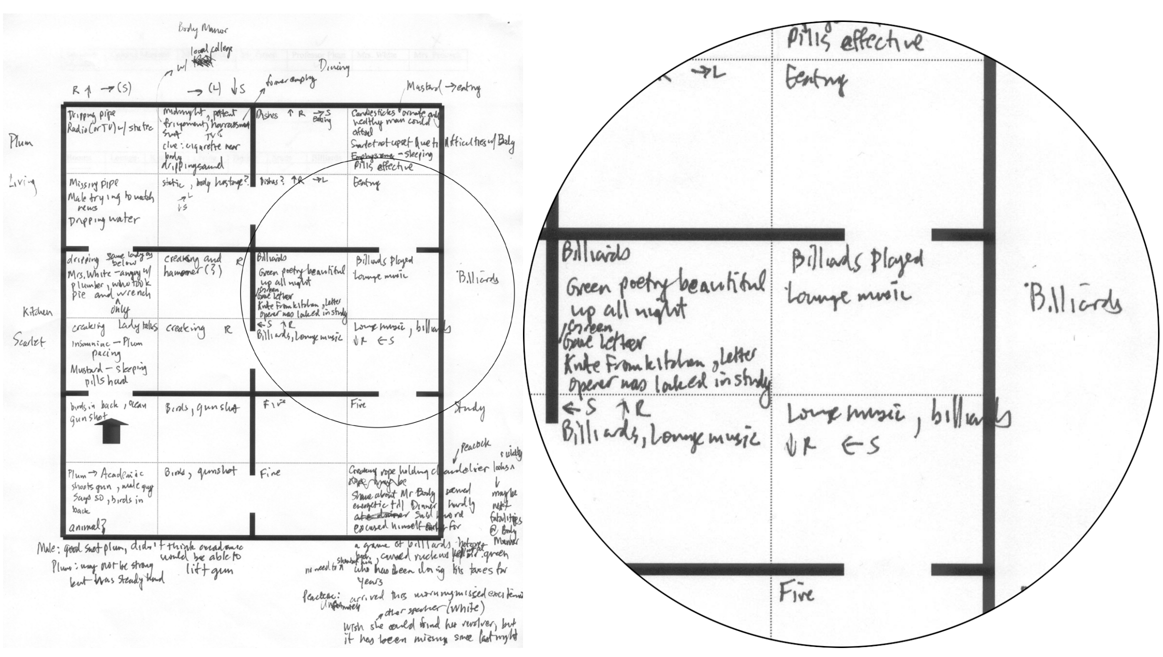


*Figure 1: Annotated map of room layout*

In general, subjects greatly enjoyed the mystery, characters, and the audio-centric concept of the game. In all but one instance, subjects correctly guessed the suspect and weapon. Only two of the subjects correctly guessed the room. Of all the clues, those related to the room were the most subtle and involved indirect connections to the suspect and weapon.

Navigation feedback seemed sufficient, and most subjects felt that they knew where they were in the house for the entire game, though a few reported an initial disorientation. When lost, every subject reported using a technique of "bumping" into walls and comparing the results to the map to regain his or her bearings. We also discussed with the subjects the possible addition of a button to reset the avatar to the start position. This idea appealed to the few who reported initial disorientation, but seemed unnecessary to those who relied heavily on the map. Two subjects' maps contained significantly more annotations than the rest (*Figure 1*), and they also reported using a well-defined search pattern to move through the rooms.

Two types of difficulties related to the audio content arose: *differentiating speakers* in conversations where names were not spoken and *identifying rooms* with somewhat ambiguous ambiences. Differentiating speakers was an issue for the conversation in the dining room involving two male characters that also appear in the lounge. In the lounge they address each other by name, but in the dining room they do not. Though a few subjects claimed to enjoy the challenges of speaker recognition, many were simply confused by the lack of differentiation.

The difficulties in identifying rooms stemmed from the distinctiveness of the associated ambient cues. For example, though the kitchen contained a "dripping" leaky pipe, the presence of a television set confounded many subjects' attempts to identify the room. A distinctive, salient ambience was also essential for identification of a given room. Subjects demonstrated this concern by consistently identifying the living, dining and billiard rooms, but they had more difficulty identifying the subtle or non-characteristic ambiences of the lounge, study, and kitchen.

Interestingly, a number of subjects considered the motives of possible suspects. Though not one of the three explicit checklist items, we feel this activity illustrates the value placed on the narrative by subjects. One stated that he "ruled out" possible alternatives to his choice of suspect because "they didn't have convincing motives."

## CONCLUSIONS

We have presented a prototype for an immersive audio game called *Sleuth* and investigated solutions to the inherent challenges of navigating an audio virtual environment. We have also described our primary design issues, including cue design and identification, avoiding information overload, basic navigation techniques, navigation feedback, and the design of an interactive narrative. We feel that these key principles in the design of *Sleuth* will also be useful for the design of other types of audio games, particularly those with strong narrative components. Audio augmented reality gaming is already possible through a range of wearable computing and augmented reality technology [8], and we hope to apply our experiences with *Sleuth* to the design of such future games. Though our user population did not include any visually impaired individuals, our work might also be extended to complement the existing body of work in support and user interfaces for visually impaired persons. Further explorations will hopefully reveal more information useful to the design of audio-based user interfaces, especially for applications in the domains of wearable computing and augmented reality.

*Clue *is a registered trademark of Hasbro, Inc.,* Myst *is a registered trademark of Cyan, Inc.,* The 7th Guest *is a registered trademark of Virgin Games, Inc.*


## ACKNOWLEDGMENTS

We are extremely grateful for the help of Irwin Coleman III, Amy Mitchell, Steve Voida, the evaluation subjects, the Everyday Computing Lab and the GVU AudioLab.


## REFERENCES


1. A3D White Paper, Aureal, Inc., 1999. Available as: <http://www.aureal.com/newaureal/technology/whitepaper.asp>.

2. Back, M. and Des, D. Micro-narratives in sound design: Context, character, and caricature. *Proceedings of ICAD '96*, November 4-6, 1996.

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

4. Cohen, J. Monitoring background activities. In Kramer G. (ed), *Auditory Display*, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, 1994.

5. Gaver, W.W., Smith, R.B., O'Shea, T. Effective sounds in complex systems: The Arkola simulation. *Proceedings of CHI '91*, April 27-May 2, 1991.

6. Gaver, W.W. Using and creating auditory icons. In Kramer G. (ed), *Auditory Display*, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, 1994.

7. Glassner, A. Some thoughts on game design. Essay, 1997. Available as: <http://www.research.microsoft.com/glassner/work/talks/games.htm>.

8. Lyons, K., Gandy, M., and Starner, T. Guided By Voices: An Audio Augmented Reality System. *Proceedings of ICAD 2000,* April 2000.

9. Mynatt, E., Back, M., Want, R., Baer, M., and Ellis, J.B. Designing Audio Aura, *Proceedings of CHI '98*, April 1998.

10. Sawhney N. and Schmandt, C. Design of spatialized audio in nomadic environments. *Proceedings of ICAD '97*, November, 1997.

11. Wenzel, E. M. Spatial sound and sonification. In Kramer G. (ed), *Auditory Display*, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, 1994.

12. Wickens, C.D. Engineering Psychology and Human Performance. HarperCollins, 1992.