# Audio Source Localization as an Input to Virtual Reality Games

Agneya A. Kerure
Georgia Institute of Technology

***Abstract*— This paper details an effort towards incorporating audio source localization as an input to virtual reality systems, focusing primarily on games. The goal of this research is to find a novel method to use live audio as an input for level generation or creation of elements and objects in the virtual realty environment. The paper discusses the current state of audio based games,virtual reality and 3-D audio. The paper also details design ideas and requirements of the hardware system consisting of microphone arrays and head tracking systems which will be used to localize the audio. Signal processing techniques which could be used for audio source localization have also been briefly discussed.**

## I. Research Goal

The goal of this year long proposed research is to explore the use of audio source localization as an input for games in the virtual reality environment. The research aims to look at how games, specifically first-person shooter games, and their designs can be augmented with this additional layer of input. The research will try to answer the question of whether this system enhances the experience of the player by increasing the required in-game concentration as the game changes with changes in the player's real auditory environment. The focus of this research, in the first half of the year, would be to create an experimental apparatus for game developers to use audio source localization in their games. In the next semester, the research will focus on specifics of game design which will be able to make the most use of the proposed system by conducting user studies to assess how these enhancements work in practice, collect feedback from players to improve the system and the game feel, and determine if this system indeed improves the game playing experience.

## II. Introduction

In order to describe the proposed system, it is important to understand a few basic definitions. Virtual Reality has seen a new resurgence in the recent years. It is defined as a three dimensional computer generated simulation in which one can interact with an artificial environment [4]. Experiencing virtual reality has four key elements, as described by Sherman et al. - a virtual world which is the content of the given medium, immersion into an alternate reality, sensory feedback through vision, audition or touch and the ability to interact with the virtual environment [11]. Considering all the key elements, a better definition of virtual reality can be given which describes it to be a medium composed of interactive computer simulations which sense the position and actions of the participant and augment the feedback to one or more senses, giving a feeling of immersion in a simulation [11].

3-D Audio usually involves the virtual placement of sound sources in a three-dimensional space. It refers to technologies which are based on binaural hearing of sound. Begault states that it is almost imperative to include distance and environmental effects to maintain a sense of realism in virtual acoustic objects and the most effective physical cue for the perception of environmental context is reverberation [2].

Audio-only games refer to games which can only be played and perceived through sound and acoustics which were initially developed for the visually impaired community but have huge potential for mobile and multi-player gaming [10]. Although these games are limited in the amount of information which can be conveyed to the player, they also have a few advantages over conventional games such as increased degree of spatial freedom, no necessity of screens and other costly equipment depending upon the game, lower computational complexity and hence less latency, making these games perfectly suited for mobile gaming. One important observation by Rber et al. is that the correct rendering of 3-D sound is technologically more complex than the generation of 3-D images [10].

There are many games in the virtual reality category which use music as an input to generate in-game content. Guitar Hero VR, Audioshield and Rockband VR are the most important examples of this genre. These apply signal processing methods to detect properties such as onsets, beats and melodies from the audio to generate specific cues which are presented to the user to act upon. These generally take the form of rhythm based games which are very common in across all platforms - desktop, mobile, console and virtual reality.

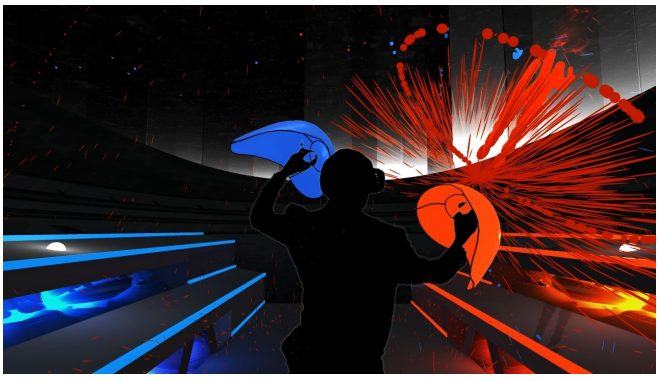

Fig. 1. Audioshield VR User Interface

This paper aims to explore the potential of audio and its localization as an input to games in virtual reality and

describe a system to do the same.

## III. Background

There are many ways in which audio has been made to be the primary means of playing a game or interacting with a virtual or imaginary environment. Anderson describes a game made specifically for the visually impaired to play which uses audio cues to get oriented in the game environment, navigate, communicate with and locate other players and find objects [1]. Roeber et al. describe an interaction paradigm for virtual auditory worlds which can be used in games and other interactive environments which uses sonification of user actions to help represent them such as landmarking and object sounds for navigation, position determination and object analysis [9].

Lyons et al. describe a low cost, lightweight audio-only augmented reality infrastructure using a simple wearable computer and an RF based location system to play digital sounds corresponding to the users location [7]. A game was also implemented using the same wearable system where the player moves around in the real world to trigger actions in the virtual game world. It was found that three layers of audio - ambient sounds, sound effects and narrations - were important in creating an immersible augmented reality game [7].

There are many ways in which a person in a virtual environment interacts with sound. Boyer et al. describe an Audio Virtual Surface as a region of space that triggers a sound when the user touches or moves into it [3]. Gestural patterns which were studied from the interaction of blindfolded participants with Audio Virtual Surfaces show that they were able to learn and complete the task of recognizing concave and convex surfaces quickly. Gestural patterns emerged from the interactions, suggesting the use of auditory representations of virtual objects [3].

There are important human factors which need to be considered while designing audio applications in augmented and virtual reality [5]. It is important to decide the predominant sensory modality of the interactive environment as cross-modal situations can prove to be confusing to users. The nature of the auditory event should also be clearly defined in the system as there is a complex connection between the creation of a signal and a listener's auditory perception of the signal. The means of presentation of the virtual audio should be considered carefully, for example, whether the audio is rendered specifically for each individual, or once for a group. Another important aspect to be considered while designing a system which deals with augmented or virtual reality audio is that if the extent of user tracking - no tracking, GPS based location tracking, or precise head tracking. The availability of the user's head position can afford greater possibilities to the system [5].

### A. Audio Source Localization

The proposed system will have two approaches to sound source localization - virtual sound source localization using binaural source imaging techniques which involve 3-D audio effects, and physical sound source localization using multiple speaker fields and a microphone array. Once the system has been tested using virtual source localization , the hardware implementation using a microphone array will be done over real world audio signals and users.

### B. Why Audio

An important question to consider would be why you audio at all. There are other ways to locate objects and people in a vicinity around a person which include computer vision, infrared tracking and even global positioning systems. The advantage audio offers is the ability to process the received signal and get vast amounts of information from it which can be used in different ways in the game. For example, speaker recognition to produce game objects specific to a speaker's voice, processing background audio to generate a game environment specific to that particular soundscape are just a few possibilities which audio can provide to a game engine.

### C. Algorithm

The most important cues for virtual localization are inter-aural level difference and inter-aural time difference, physical structure of the ear and the environment - reflections, propagation effects and proximity to source. These cues are estimated to determine the head related transfer function which is the frequency response of a listener's ears which is then used to locate audio sources in a 3-D space. Nishiura et al. present a method to localize multiple sound sources based on cross power spectrum phase analysis with a microphone array [8]. The cross power spectrum method localizes a sound source as a crossing point of sound directions estimated using different microphone pairs[8]. Huang et al. present a real-time passive source localization approach which uses linear-correlation least square approximation procedure which yields an efficient source location estimator without assuming prior knowledge of the distribution of sound sources [6]. it is proposed that this estimator performs

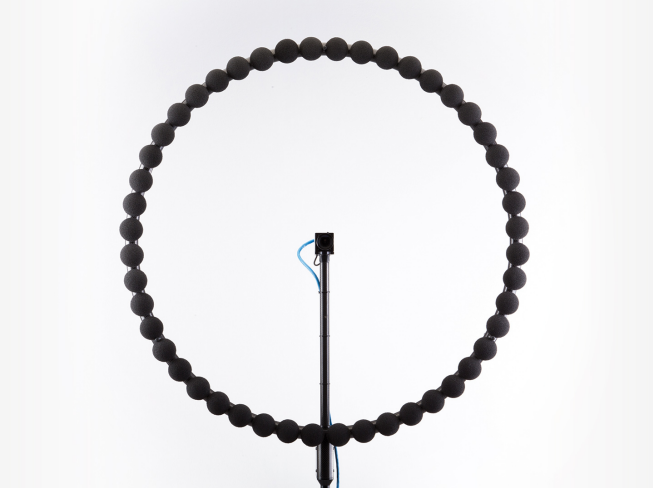

Fig. 2. A Typical Ring Shaped Microphone Array

better under many practical situations than likelihood based spherical intersection and spherical interpolation [6].

## IV. System

The project will include a hardware microphone array mounted on the user's head capturing incoming audio and sending the data to the software to process the data and locate the audio source and form graphical game elements and objects.

### A. Hardware

The author proposes a system which includes a hardware extension for the HTC Vive Virtual Reality headset. The HTC Vive is a virtual reality headset developed by HTC and Valve Corporation.It allows users to move in a restricted 3-D space and interact with it using hand-help controllers which are motion tracked. The hardware extension to the HTC

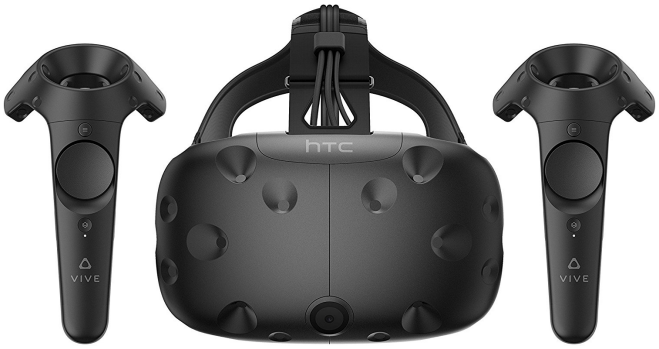


Fig. 3. HTC Vive

Vive would include a microphone array arranged in a ring around the head of the user which would be used to localize audio sources in real time. The exact details, construction and design of the microphone array system is subject to experimentation and analysis which would be carried out in the coming months. The system will use the audio captured by the microphone to locate the source of the sound. The accelerometer and gyroscope included in the Vive headset along with the built-in Lighthouse tracking system will be used to monitor the users position with respect to the audio sources.

If this proves to be too much work for one semester, the author proposes to use binaural audio to simulate audio sources in 3-D space around the player. This can be done in two ways:

- Simple single source localization using two microphones is problem which has been solved and might be easier to implement. This will reduce the accuracy of the system but still have the same working functionality.
- Capturing Real-time binaural audio using binaural recording earphones which are currently available in the market. The captured audio can then be processed to locate sound sources. This method might not be as accurate as using a microphone array.
- Generating binaural audio in the system during gameplay to create virtual objects in the VR environment. This method will make the presence of other players redundant.

### B. Software

The proposed system includes a software implementation of a novel real-time audio source localization algorithm which is also subjected to change with due experimentation and analysis. Unity is a cross platform game engine which is developed by Unity Technologies that supports both 2-D and 3-D graphics. The development and analysis of the algorithm for audio source localization will be done in Matlab and then ported to Unity to build a virtual reality environment using the obtained data.

### C. Game Design

The main aim of the project is to provide a system which would be able to use the location of an audio source in a virtual reality application. The nature of the application can be very varied - from games to test simulations, from training environments to social interactions. This research will explore the avenue of game design using the provided source localization data. First person shooter games, which involves the user in a weapon based combat environment, would make the best use cases for the proposed system as it would intensify the feeling of surprise and would increase required concentration from the user, thereby increasing their involvement in the game. Apart from the graphical environment, the auditory environment will be very important for this system as it plays a significant part in setting the stage for the game's scene and play. The games and interactive environments created based on this input system will be very sensitive to sounds and must make good use of the capabilities provided by the system. The data of the localized audio source could also be used to create virtual sound objects which can be interacted with by the user and other sound objects - giving rise to interesting paradigms and game design avenues. Another important aspect to consider would be the correct synchronization of sound and moving 3-D objects, the absence of which can falsely represent action times.

A few game ideas which could be implemented using this technology would be:

- A simple shooting game where the opponent can create targets around the main user using their voice. The main player can shoot at the targets using the hand held controllers. Another possibility would be to create graphical sound objects which would have the players audio recorded into it. These objects would have physical properties associated with them and would be affected by the game physics engine. These can be used as weapons against the targets or enemies produced by the other players.
- A Harry Potter-esque game where the player battles other players by casting spells at them. The hand-help controllers act as a wand in one hand and a shield in the other. The opponent player will move around the player with the headset on and try to defeat him by casting

spells which hit the player or by casting stronger spells which counter shields. This game would use speech recognition to identify spells and individuals with specific powers. Although this might end up becoming a shouting contest, constraints could be put up in the game to counter this

## V. Challenges and Milestones

There are a few main challenges which will be address in the different stages of development of the proposed system:

- Making the source localization algorithm available to Unity or developing it in C# for direct use by the Unity Audio Engine.
- Real-time multiple audio source localization
- Designing games which would be able to make the best use of the proposed system
- Latency and processing time could be a big issue and can hinder game play. This also depends on the system which is being used.

The author proposes the following time line to tackle the proposed research:

- September - October, 2017: Audio source localization algorithm development, verification, data acquisition and testing
- October - November, 2017: Hardware implementation - microphone array and head mount development
- November - December, 2017: Algorithm implementation in Unity with prototype game
- January - February, 2018: Hardware Development, Improvement and Verification
- February - March, 2018: Game Re-Design and Development
- March - May, 2018: User Testing and fine tuning the system

## VI. Conclusion

This paper proposes an innovative system which involves audio source localization and signal processing to provide input to a virtual reality environment. Milestones and development strategies were also discussed and the hardware ad software requirements of the system were mentioned. The author believes that this is a one-of-a-kind system which can take current state of virtual reality to newer heights.

## VII. Acknowledgments

The author would like to convey his gratitude to the Georgia Tech Center for Music Technology for giving me the opportunity to work on this research topic. I would also like to thank Dr. Jason Freeman and Dr. Alexander Lerch for their constant guidance and support throughout the stages of ideating for the project. The research would have been drastically different, and probably less successful, without their help. Of course, a big thank you to everyone who provided their valuable feedback and suggestions.

## References

[1] Andresen, G. Playing by ear: Using audio to create blind-accessible games. *Games Developer, October* (2001), 28–32.

[2] Begault, D. R. *3DD Sound for Virtual Reality and Multimedia*. Academic Press Professional, Inc., San Diego, CA, USA, 1994.

[3] Boyer, E. O., Vandervoorde, L., Bevilacqua, F., and Hanneton, S. Touching sounds: audio virtual surfaces. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)* (March 2015), pp. 1–5.

[4] Deniz, L. L. The development and implementatin of virtual reality for training in the future. *Commission on Peace Officer Standards and Training* (2000).

[5] Huang, W., Alem, L., and Livingston, M. A. *Human factors in augmented reality environments*. Springer Science & Business Media, 2012.

[6] Huang, Y., Benesty, J., Elko, G. W., and Mersereati, R. M. Real-time passive source localization: a practical linear-correction least-squares approach. *IEEE Transactions on Speech and Audio Processing 9*, 8 (Nov 2001), 943–956.

[7] Lyons, K., Gandy, M., and Starner, T. Guided by voices: An audio augmented reality system. In *Proceedings of International Conference on Auditory Display (ICAD) 2000* (2000), Citeseer.

[8] Nishiura, T., Yamada, T., Nakamura, S., and Shikano, K. Localization of multiple sound sources based on a csp analysis with a microphone array. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)* (2000), vol. 2, pp. II1053–II1056 vol.2.

[9] Röber, N., and Masuch, M. Interacting with sound: An interaction paradigm for virtual auditory worlds. *Proceedings of DiGRA 2005 Conference: Changing Views - Worlds in Play* (01 2005).

[10] Röber, N., and Masuch, M. Playing audio-only games. *A compendium of interacting with virtual, auditory worlds (Article)* (2005).

[11] Sherman, W. R., and Craig, A. B. *Understanding virtual reality: Interface, application, and design*. Elsevier, 2002.