Echo : a VR-based Approach to Promote Ear Training

Rahul Agnihotri
Stanford University
ragni@stanford.edu

Deepak Chandran
Stanford University
cdeepak@stanford.edu

Abstract

This paper introduces a VR based environment that 'gamifies' the process of practicing music theory topics like ear training and short term pitch memory for non-musicians and people with some previous musical background. We discuss the design behind creating such an environment that requires the player to use audio cues to work towards a final objective. Implementation includes the use of real time spatial audio signal processing using Chunity, a Unity wrapper for 'Chuck', a strongly-timed audio programming language.

1. Introduction

The use of VR scenes for the purpose of teaching has been a widely researched field since it provides learners with simulated environments where they could practice their skills and get real time feedback that is incredibly beneficial during the process of learning. Such environments also provide an opportunity to 'gamify' hard-to-grasp concepts, greatly reducing the learning curve required for the same. This holds particularly true in the field of music theory, where fundamental topics like 'ear training' and 'sight reading' which are important tools in the arsenal of any good musician.

In our implementation the player is locked in a morgue-like environment from which they are required to escape by locating the correct key using echolocation. By using this concept of an escape room, we are able to provide immediate feedback that feels rewarding to the learner at the end of the game. As we elaborate later on in this paper, the environment and the associated background music was designed to instill a sense of urgency in the player which adds to the more 'fun' aspect of the game, while at the same time creating a timed learning environment that greatly increases the efficacy of learning. In addition to this, the environment weve created seeks to improve the spatial awareness of the player by requiring them to locate the relative position of the keys around them based off the locations of the sounds they hear.

2. Related Work

During the initial stages of designing our virtual environment, we tried to seek inspiration from the wide gamut of games that use echolocation as the means of moving around. By placing the player in complete darkness and having them discover their environment using sound waves and reverberation, *Blind VR* was able to create a highly immersive experience with barely any visual feedback. This game mechanic was also explored in *Stifled*, where the player has to carefully choose the right frequencies while echo-locating or risk getting caught. It is important to note that this game mechanic has been extensively used in non-VR games as well.

There exists a wide assortment of ear training tools in the market, ranging from free/cheap software consisting of simple interfaces requiring the learner to play back simple melodies to full fledged packages that analyze learner performance and adapt its' content while provide detailed statistics showcasing the learner's strengths and weaknesses. Such tools have price points that are often much too high for the amateur musician who would like to improve their ear training skills. Also, having had first hand experience with these sorts of tools ourselves, they can be quite hard to understand for beginner musicians due to the large amounts of musical jargon involved which in some cases demotivates the learner.

The gamification of music theory has proven to be extremely effective especially for kids, where attention spans are quite low. Technical exercises for practicing are vital for musicians, but are often frustrating and viewed negatively by students. By including game mechanics like a storyline, social context, control, etc., Birch [3] was able to gamify the technical exercises required for a piano student. The game design elements which were implemented in this study to test for an effect on student motivation included awarding points and virtual trophies for beating levels, the use of avatars, and the sharing of student progress online. It was anticipated that gamification could, in fact, motivate students to practice technical exercises, which would result in increased practice time indicated by improved student achievement.

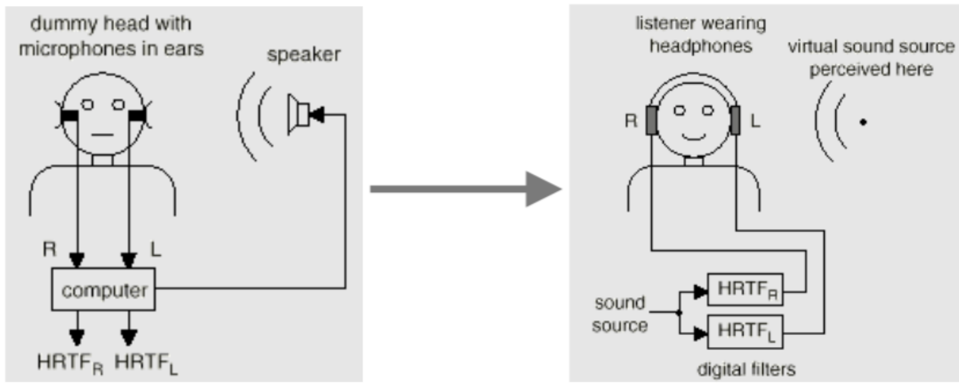

Figure 1. HRTF Generation and its usage

3. Music Theory Concepts

3.1. Ear Training

There are many characteristics of music including the pitch, the intervals between musical tones and the types of musical triads. Ear training serves the purpose of training listeners to recognize these characteristics. Ear training enhances the pleasure of music listening and sensitizes a musician's ears for the study, comprehension, performance, and creation of music. It also enables a musician to identify intervals, chord qualities, rhythmic patterns, and to audiate harmonic and melodic phrases necessary for creating and performing music.

3.2. Short Term Pitch Memory

Musicians often develop a high capacity for keeping musical information in their memories, both short-term and long term. The most basic type of musical memory is called pitch memory. To give a simple example, when you play a tone on a pitch pipe in order to establish the key before singing a song, you are using your short-term pitch memory in order to store the starting tone in your mind before you start singing. The development of a strong memory for pitch regardless of timbre (instrument sound) is an important step in building ones general musical memory.

4. Auditory Spatial Awareness

Spatial Awareness is the awareness of the surrounding space and the position and location of our body within such a space. Spatial ability is an important aspect of human intelligence, and is pervasive in day to day life. We rely on our ears to tell us where sounds – from the chirp of a bird to the call of your name in a crowd – are coming from. Locating and discriminating sound sources is extremely complex because the brain has to process spatial information from many, sometimes conflicting, cues. Spatial awareness resulting from auditory cues is commonly referred to as auditory spatial awareness. This awareness incorporates within itself, sound source detection and acoustic signal recognition [6]. The average person's auditory spatial awareness depends on a multitude of factors ranging from the listeners hearing sense, their auditory experience, familiarity with the surrounding environment, and how involved he/she is in the listening activity (for example, their attention, tiredness, etc.). Using virtual reality and other immersive technologies, researchers are able to use new methods to investigate how we make sense of the word with sound.
Spatial perception in the aural domain is a simple experience to sense but a rather complex one to discriminate, quantify and classify. The complex structure of sound quality being a multidimensional phenomenon has been addressed by several recent works [7] [8] [2].
Acoustic cues are a huge part of this and contribute a lot to establish a persons location in space. The three main cues that determine the source of sounds are as follows:

- Interaural time differences (ITD)
- Interaural level differences (ILD)
- Acoustic filtering produced by the shape of one's ears, and head

Combining these cues together, we can render Head Related Transfer Functions (HRTFs) that provide us with a pretty accurate representation of the acoustic space. While individualized HRTFs based on the features of ones head and torso can be rendered, they can often be unnecessarily time consuming and expensive to implement. Instead, generic

HRTF banks are created that can be used on any device with head tracking capabilities.
It has been shown through research on auditory perception that our representation of acoustic space is fairly plastic. For example by asking participants to block one ear, Kumpik *et al*. [5] were able to show modest improvements in auditory spatial localization over a period of 2-7 days. This plasticity of the human auditory perception plays to our advantage while creating a game that relies mostly on auditory perception.

5. Game Mechanics

The game starts off with placing the player in a pitch dark environment as in Figure 2. The player then hears tones corresponding to different pitches that are emitted from keys that are placed randomly in the room. The job of the player is to locate the correct key by humming back the tone corresponding to that key. If hummed correctly, the player receives feedback by hearing their own voice played back to them.

6. Implementation

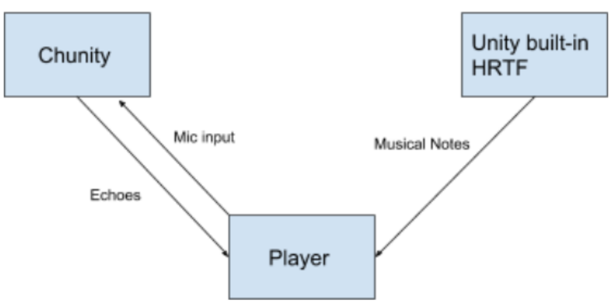

Figure 3. Audio Flow

6.1. Chunity

To implement pitch detection of the players voice and playback of the processed version of the same, we used Chunity, a plugin for ChucK written for the Unity game engine. ChucK is a programming language used widely by musicians in the computer music domain due to its strongly timed nature. One of the primary strengths of ChucK is that the programmer can do analysis on incoming signals with sample-synchronous precision. This allows for the exact scheduling of operations, and also allows any parameter to be changed at any time. For example, for FFT analysis, it's possible to conduct FFT analysis with dynamically changing window, hop and FFT sizes [?].

```
// audio input
adc => FFT fft;
UAna agglomerator => blackhole;
// connecting FFT to feature extractors
fft =^ Centroid c =^ agglomerator;
fft =^ Flux f =^ agglomerator;
fft =^ MFCC m =^ agglomerator;
fft =^ RollOff r =^ agglomerator;

// control loop
while( true )
{
    // upchuck!
    agglomerator.upchuck();
    // HERE: do stuff with results
    // ...
    // advance time
    HOP_SIZE::samp => now;
}
```

Figure 4. Simple Feature Extraction in ChucK

Chunity is a C++ Unity native audio plugin created at CCRMA (Center for Computer Research in Music and Acoustics) [1], that allows programmers to create ChucK objects within Unity scripts. Through a $ChuckInstance$ object within the unity script, a ChucK VM acts as a virtual sound source that can be spatialized within the game world. This method also enables new ChucK VMs to be constructed programmatically with the use of Unity prefabs.

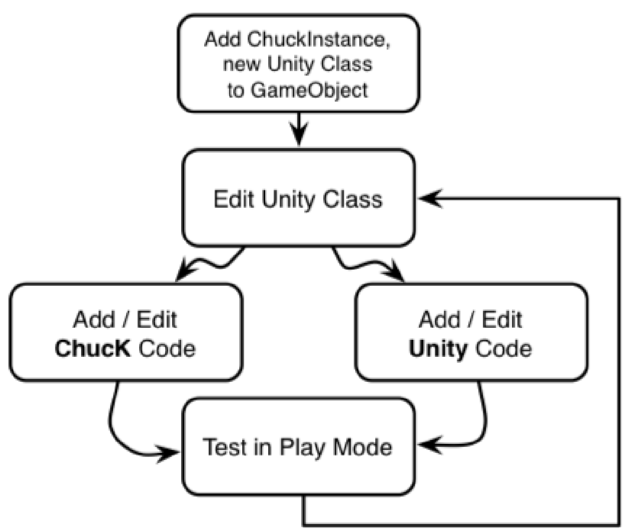

Figure 5. Chunity Workflow

Figure 2. Our Virtual Environment

7. Evaluation

Due to the time constraints we were unable to conduct a thorough evaluation of how effective such a system would be in learning music theory and improving spatial awareness. However, we would like to put forth a couple of hypotheses, backed by our literature review, that we could test out as part of our future work.

- As mentioned earlier, gamifying music theory concepts provides learners added incentive to improve their skills, greatly increasing student motivation. Additionally, it's been shown that ear training personalized for the learner is often much more affective [4]. These two hypotheses can be tested out by using standardized ear training tests.

- As noted by Berger *et al*. [2], there were no improvements to spatial awareness solely using auditory cues in the absence of visual cues. This would be interesting to evaluate in our environment, due to the nature of the visuals that we've used (almost pitch darkness with the ability to teleport to areas with slight more light with the additional helping hand of a flashlight attached to the).

8. Discussion

This project has provided us with the opportunity to explore the gamification of music theory through the VR medium. After receiving feedback from people who tried out the the game during the demo poster session, weve come up with a few scenarios that can be further explored in our future work:

- Currently, the player is asked to recreate the exact pitch of the tones that are played in the background. While this wasnt too hard to accomplish for people with some sort of musical background, we did notice that some of the players getting completely stuck since they were humming the pitches in a different octave. This could be fixed by introducing a beginner mode that allows the player to recreate the same note in any octave they want.

- It was noted that there wasnt adequate visual feedback for the user to indicate the pitch that they were humming. This would be helpful since this could contribute to the learning aspect of the game (which this environment was originally created for).

- We received a few suggestions to increase the brightness of the environment, however this went against our design ethos of creating an eerie environment where the player was to solely rely on their auditory sense. This could be, in part, solved by allowing the player to emit a 1 second long pulse of light that extends radially outwards from the player.

9. Acknowledgements

We would like to thank the entire teaching team from EE267 to provide us with the motivation to bring this conceived idea to fruition as well as the everyone at Stanford CCRMA to provide us with support and the equipment to make this project happen.

References

[1] J. Atherton and G. Wang. Chunity: Integrated audiovisual programming in unity. *NIME.*

[2] J. Berg and F. Rumsey. Systematic evaluation of perceived spatial quality. *Proceedings of the AES 24th International Conference on Multichannel Audio,.*

[3] H. Birch. Motivational effects of gamification of piano instruction and practice.

[4] S. Kim. An assessment of individualized technical ear training for audio production. *The Journal of the Acoustical Society of America.*

[5] K. A. J. Kumpik D. P., Kacelnik O. Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. *J. Neurosci.*

[6] T. R. Letowski and S. T. Letowski. Auditory spatial perception: Auditory localization. *Army Research Laboratory*.

[7] T. Neher. Unidimensional simulation of the spatial attribute ensemble depth for training purposes, part 1: pilot study into early reflection pattern characteristics. *Proceedings of the AES 24th International Conference on Multichannel Audio,*.

[8] N. Zacharov and K. Koivuniemi. Unravelling the perception of spatial sound reproduction: techniques and experimental design. *Proceedings of the AES 24th International Conference on Multichannel Audio,*.