Accurately Pinpointing 3D Sound in Virtual Environments

A Senior Project presented to the Faculty of the Liberal Arts and Engineering Studies Department California Polytechnic State University, San Luis Obispo

In Partial Fulfillment of the Requirements for the Degree Bachelor of Arts in Liberal Arts and Engineering Studies

> by Kevin Hongtongsak June 2016

© Kevin Hongtongsak 2016

Table of Contents

Introduction

l

As technology advances and creates increasingly sophisticated applications for hardware and software, consumer technology is also being impacted, most notably in the area of entertainment. It goes without saying that the near ubiquitous usage of computers in entertainment and technology makes using these new technologies a central hub of entertainment and communication (Ip, 2008).

As more and more virtual reality (VR) devices begin to enter the consumer market, many users tend to focus on the projected images from VR devices as being the sole carrier of immersion. While visual cues are important in distinguishing what environment we are currently located in, audio also has an important aspect in immersion¹ (Serafin and Serafin, 2004), as auditory signals allow the brain to perceive where the body is located in space (Begault 12). In the real world, visual and auditory signals orient the body, allowing the individual person to navigate the space around them freely and accurately. Thus, to create an immersive VR experience, a combination of high quality graphics and realistic audio is needed in order to accurately mimic reality in such a way that a user is unable to distinguish a virtual environment from a real one.

Since previous research shows that audio is known to have an important effect on immersion, my research question is: how accurately can sound be pinpointed in a virtual environment? Having accurate sound models in a VR environment would increase the standards for immersion quality in VR environments, which could lead to potential applications

 $¹$ Immersion is defined as the perception of being physically present in a non-physical world. In the context of</sup> virtual reality, it's believing that the virtual environment around a person feels as if it's a real world.

of this technology in areas of entertainment and education. This paper will discuss the effects of sound as it relates to spatial perception and virtual immersion.

Project Deliverables

The deliverable as it pertains to this project is a standalone Unity application, designed to be run on a computer running a Windows operating system. I developed this application by myself using the provided audio plugin from DTS, Inc., formerly known as Digital Theater Systems, Inc. This application features an extremely sparse virtual environment in order to minimize any chances of motion sickness or disorientation for users. Figure 1 displays a screenshot of the virtual environment. The cyan highlight was used only in the development version in order to highlight the sound object. In the release version of the application, every box is shaded grey. This color was chosen due to its strong contrast with the background, which would allow users to easily see the boxes in relation to their current position.

Figure 1. A screenshot from the application. The highlighted box represents the sound object.

While a user is actively engaged with the application, pressing the '1' key spawns 25 boxes at random locations around the user. One of those boxes, shown the in figure above as the cyan box, produces a sound that conforms to the behavior specified by the DTS plugin. For example, if the user is close to the object, they would expect a louder sound from the object and vice versa. The user is also free to move around the environment in order to locate the sound object. A script that records the user's distance from the sound object is executed every frame, which is approximately every 0.016 seconds. When the user feels that they've located the sound object, pressing the 'E' key records their final distance and removes all the boxes in the scene. All distances recorded during the active session are written to a plain text file. Finally, two spreadsheets, one each for a control and an experimental group, will be delivered. These spreadsheets contain data on each participant's run with the Unity application.

Literature and Technology Review

A literature review of virtual surround sound suggests that it is still a relatively new focus of research and development. Effective surround sound is generally created by utilizing a mechanism called a Head-Related Transfer Function (HRTF), which allows a human listener to pinpoint where a sound might be located in three-dimensional space based on the time difference of audio waves entering the left and right ears (Talagala et al., 2014). Talagala et al. also go on to note how HRTFs vary widely in humans, as individual head and ear shapes affect how sound waves interact with those physical characteristics, affecting the time difference of sound waves entering the left and right ears. This ultimately affects how the individual would perceive the incoming sound. Thus, in order to create a believable and immersive surround

sound experience, the HRTF must be personalized for each person in order to produce a truly realistic surround sound experience. There has been some work done on personalizing the HRTF for each person by using their individual ear measurements, but a truly personalized HRTF remains an open question in computer science (Zotkin et al., 2014). Further research is expected in this area of computer science as the capabilities of virtual reality continue to increase.

Currently, the two major platforms for VR development are Unity and Unreal Engine 4, which are two popular game engines used by game developers. Both of these game engines are very powerful since they include features that streamline the game development process and a variety of computer graphics processing techniques, one such example being physically-based shading². Both Unity and Unreal Engine 4 have been used for all kinds of game development, including both 2D and 3D games, though now both are pioneering support for virtual reality development. Besides these two engines, I was not able to find any other engines that supported VR development. Neither one seems to more popular than the other, but as is the case with most software developers, there is plenty of bias for each engine depending on what the developer is familiar with. Both engines seem more than capable of supporting high-end VR development and do not seem to have anything that makes one particularly better than the other.

As VR is still a newly developing area of technology, the types of VR devices available are limited, though, more and more companies have been releasing their own devices as the

 $\overline{}$

² Physically based shading is a computer graphics technique for displaying images that simulate the interaction between light and materials in the scene in real-time, which deliver more consistent lighting results. (Pranckevičius, 2014)

technology to support VR has improved. Currently, the most well-known brand seems to be the Oculus Rift, due to it receiving large amounts of media attention after raising over 2,000,000 USD and gaining 9,522 backers since its launch on Kickstarter on August 1, 2012 (Oculus, 2012). Other notable VR devices include Samsung's GearVR, Google Cardboard, and the HTC Vive.

Technology Overview

The technology being used to develop this senior project is the Unity game engine that has the DTS audio plugin and the Oculus Rift Development Kit 2, abbreviated DK2 (Figure 2).

Figure 2. The Oculus Rift DK2.

The game was developed in Unity3D in order to support usage of the Oculus Rift. Unity was chosen as the game engine since it DTS had developed a plugin for it, in addition to being both powerful and versatile. Additionally, the programming languages supported by Unity, C# and JavaScript, are

both languages I had been interested in learning, so it was an excellent way to simultaneously learn about audio engineering and a new programming language at the same time. The Oculus Rift was chosen due to Unity featuring built-in support for it, and because it is relatively lower priced for the performance it provides compared to other VR devices.

For testing, I utilized Google Forms and Microsoft Excel to collect and analyze data after each participant used the application. I chose Google Forms because it has an extremely streamlined data collection system and has the benefit of being able to be accessed anywhere with an internet connection. I used Excel collate my data in order to produce graphs and perform statistical analysis.

Implementation and Timeline

Originally, this project was developed as a result of a collaboration between DTS and California Polytechnic State University, San Luis Obispo. The original goal was to provide us, a select group of students from the Computer Science and Liberal Arts and Engineering Studies departments, a way to gain experience in audio programming and research by using the provided DTS plugin. It has then evolved into being a project on audio and virtual reality immersion. As it relates to the Liberal Arts and Engineering Studies senior project, the timeline of this project began in January 2016. The starting goal was to finish and submit a demo application of a virtual movie theater using the plugin. We developed the virtual movie theater over the course of several months prior to January 2016, slowly gaining assets and an understanding of what kind of audio setups are typically found at most commercial movie theaters. During this time period, we learned more about the Oculus Rift and its capabilities with Unity, which included automated head-tracking input. Once we were able to successfully complete the demo application and hand it off to DTS, each member of the group took on particular sections of the project to pursue further research. I chose to focus my research on identifying locations of 3D sound in virtual environments by using the provided DTS plugin.

During February 2016, I began to develop the Unity application. I spent time researching how to produce an environment that would minimize motion sickness and allow for a broad range of sound distances to be covered. The sounds to be used in the simulation are in the range of 6 kHz to 10 kHz. This range was chosen because most humans past the age of 50 tend to have reduced hearing facilities past 8 kHz (Arvin et al, 2011). Ultimately, I chose a built-in Unity sound, which was the "laser blast" sound. The sound was very distinct and easy to identify, which would aid participants in locating the location of the sound object.

From March to April 2016, I refined the application further, ensuring that the expected movement behavior matched what I intended. For example, the user of the application should be able to turn their head while still moving forward. In addition to turning their head, the sound should also rotate with their head, so that the directionality of the sound is maintained. Once I was satisfied with the application, I began testing in May 2016.

Testing took place over two and a half weeks, from May 12, 2016 to May 30, 2016. A variety of participants were collected from the California Polytechnic State University, San Luis Obispo, campus. About five participants were collected from within the San Luis Obispo community. In order to facilitate the test, I asked participants to come to relatively quiet areas on campus, specifically within Kennedy Library and the Liberal Arts and Engineering Studies laboratory. The headphones used in the experiment were a pair of Sennheiser HD239 headphones. If a participant's head was too large or wide for the Sennheiser headphones, I used a pair of Panasonic ErgoFit In-Ear EarBud headphone, specifically using the RP-HJE120-G model.

Before conducting the experiment, I would calibrate each participant's head to the Oculus Rift head tracker camera using the Oculus Rift Configuration Utility provided by the Oculus software development kit (SDK) version 0.8. This would ensure that each participant's head would be accurately represented in space while they were engaging with virtual environments. Following the calibration, I asked users to interact with the Oculus World demo provided by the Oculus SDK, in order to allow participants to familiarize themselves with moving around in a virtual environment. This environment is primarily used to demonstrate the capabilities of the Oculus Rift, which served as a primer for those unfamiliar with virtual reality. It also had the added benefit of informing me whether or not participants would experience motion sickness while using the Oculus Rift.

Finally, I conducted the experiment itself. Upon launching the Unity application, the participant would be placed in my virtual environment. I gave them instructions on how to interact with the world: the keyboard arrow keys for movement, and the 'E' key to record their final distance and eliminate the currently spawned boxes. Pressing '1' once the boxes were eliminated would randomly spawn a new set of boxes. I would repeat the process of having participants find the sound object five times, so that I could get a clear sense of how each participant was moving and reacting within the environment. Additionally, I conducted a control group using the standard Unity 3D sound component in order to determine if there were any differences in sound perception in using the DTS plugin versus the built-in Unity sound components since both sound models propagated sound differently through virtual space.

Analysis and Verification

In total, I had 42 volunteers who participated in my senior project: 28 in the experimental group and 14 in the control. The age distribution of the experimental group are as follows: 60.7% were 18-21, 25% were 21-24, 10.7% were 24-30, and only one person, the remaining 3.6%, was over 30 years old. This gave me a broad range of ages in the experiment, which allowed me to see if a variety of age groups could accurately determine the sound's location using the DTS plugin. 75% of the volunteers did not have any virtual reality experience before, which meant that they would be experiencing virtual audio for the first time. The age distribution in the control group are as follows: 78.6% were 18-21 and 21.4% were 21-24. With the exception of one person, the entire control group had never had experienced using virtual reality.

Qualitatively, for the experimental group, none of the participants indicated on the Google survey that they felt the sound was unrealistic, although 14.3% of the participants reported that they felt unsure whether or not the sound was realistic. In their responses, they mentioned having difficulties hearing the sound or being unsure of what to compare the experiment's sound to. These responses did not make any mention of how the DTS plugin propagated sound, so they can be attributed to underlying personal preferences in sound quality and volume. The remaining 85.7% of participants reported that the sound felt realistic. These participants felt it was realistic for a variety of factors, though it appears that direction and sound magnitude were the key qualities that gave a sense of realism. This is also reflected when participants were asked to rate, on a scale of 1 to 5, with 1 being very difficult and 5

being very easy, how easy it was to locate the sound. The survey revealed that 75% of participants rated the easiness being at least 4, meaning easy; 14.3% rated it a 3, meaning neutral; and the remaining 10.7% rated it 2 or below, meaning difficult.

For the control group, all the participants indicated that the easiness of finding the sound was at least a 3, with 58.3% of the participants rating it a 4. Interestingly, most users from the control group also felt that the sound propagation felt realistic through the sound magnitude and directionality. However, the responses seem to only indicate sound magnitude as the deciding factor in locating the sound.

In Excel, I plotted each participant's distance from the sound object as a function of time. Ultimately, the goal of the project was to see whether or not there was a significant difference in the total distance covered using DTS plugin compared to the Unity sound component. Since I had the participants do five runs in the virtual environment, each graph of their test had five lines; a few examples can be found in Appendix C. I then approximated the area under each of the lines by numerical integration using the trapezoid rule. Since the trapezoid rule specifies that the area under a line is found by dividing the area underneath into trapezoids, and then taking the sum of all the trapezoid areas, we arrive at the following mathematical equation:

$$
\int_{k}^{b} f(x) dx \approx \frac{1}{2} \sum_{k=1}^{N} (x_{k+1} - x_k) (f(x_{k+1} + x_k))
$$

The area under each line indicates how much distance a participant covered in order to find locate the sound object. Statistically, the lower the area under each line, the faster each participant found the sound object, which indicated how well the DTS or Unity sound

component assisted them in finding the sound object. To prove that the DTS plugin mimicked sound realistically, the average area under the respective graphs should be smaller from those who experienced the Unity sound component.

I also performed my statistical analysis entirely within Excel. I first determined whether or not the data I collected had a normal distribution by using the Shapiro-Wilks test for normality. The null hypothesis for the test is that the data collected follow a normal distribution, and the alternative hypothesis is that the data do not follow a normal distribution. We can calculate the test statistic using the following equation, where *xⁱ* are the ordered random sample values, and *a^I* are the constants generated from the covariance, variance, and mean of the sample size *n* from a normally distributed sample.:

$$
W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
$$

I selected an alpha level of 0.05 when deciding to reject the null hypothesis. After calculating the test statistic, the control group was found to have a p-value of 0.346 while the experimental group had a p-value of 0.611. Thus, both values are greater than 0.05, so we fail to reject the null hypothesis that the data are normally distributed. Knowing that the data is normally distributed makes it possible to perform a t-test to find any statistically significant difference between using the DTS plugin and the Unity sound component.

Since I was only interested in whether or not the DTS plugin portrayed 3D sound more accurately than the Unity sound component, I used a one-tailed t-test for my statistical analysis. Altogether, I had 14 samples from the control and 22 samples from the experimental group

used in my analysis. I had six outliers in my experimental group, so I did not include them. Both groups did not have similar variances, so I used an unequal variances t-test, also known as Welch's t-test, with an alpha value of 0.05. The test statistic is calculated as follows, where \bar{X} , s_1^2 and N_1 are the first sample mean, sample variance, and sample size, respectively:

$$
t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}
$$

Since both sample groups have unequal variances, we use the Welch-Satterthwaite equation to approximate the degrees of freedom v , where $v_1 = N_1 - 1$ and $v_2 = N_2 - 1$:

$$
\nu \approx \frac{\left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_1}\right)^2}{\frac{S_1^4}{N_1^2 \nu_1} + \frac{S_2^4}{N_1^2 \nu_2}}
$$

After calculating the test statistic and the degrees of freedom, we find that they are 1.872 and 11.151, respectively. Looking up the test statistic in a standard t-table puts the calculated p-value between 0.05 and 0.025. After linearly interpolating the calculated test statistic between the provided test statistics for a p-value of 0.05 and 0.025, we find that our pvalue is 0.0454, which is less than our alpha value of 0.05. Thus, we have statistically significant evidence that the DTS plugin is better at portraying 3D sound in virtual environments than the Unity sound component.

Societal Impact

The impact a highly immersive and stimulating VR environment can have is huge. The most practical applications of VR are in the development of training simulations. There are training simulations being developed for all kinds of professions, which have been shown to help improve job performance, especially if the job is technically demanding (Seymour et al, 2002). Having a VR training environment can be incredibly beneficial for the trainee because it allows them to make mistakes in a controlled environment, yet gain valuable experience based on a real-life situation they may have to encounter. This can be extended to military uses as well – a soldier can train themselves for combat by experiencing a VR combat scenario. With high resolution visual graphics and an accurate, stimulating auditory environment, the solider can react accordingly to what they see or hear on the virtual battlefield, gaining combat experience that can keep them and their fellow soldiers safe.

Other practical uses of VR can be found in the treatment of phobia. Virtual reality exposure therapy (VRET) can help patients confront their fears in a controlled environment. VRET can be found to be effective in carefully selected patients (Parsons, Rizzo 2007). Having even a more immersive environment through the addition of well-developed sound in those VR environments could prove to be even more effective in helping patients overcome their fears by allowing them to come face-to-face with a nearly life-like scenario.

Furthermore, VR can be utilized to provide entertainment for the consumer. It has the potential to be a common household device for entertainment. A possible example of this could be its use for relaxation and de-stressing by allowing the wearer to virtually transport

themselves into a soothing environment. Another potential example could be using VR to watch television shows or movies in an immersive theater environment.

Future Work

If had more time, I would have liked to investigate how a user might perceive a visually rich environment combined with detailed surround sound, for example, a view from a hill that overlooks an area. Would the visuals affect what the user expects to hear? If user saw that they were on top of the mountain, would those users expect a certain quality of sound for the wind howling by, or would there be no perceived difference between arbitrary wind sounds? My senior project only focuses on the auditory aspect of VR, but I believe that it would be an interesting topic to pursue further research, especially now that VR is becoming more widespread as technology advances. As better immersion is sought after with more advances in VR technology, I think learning all the implications and consequences of VR will be necessary to promote ethical and safe usage of a VR environment.

Conclusion

The lag between graphics and audio in virtual reality is slowly closing as more and more research is put into producing high quality, 3D sound for virtual environments. Existing 3D sound models, such as the Unity one, portray it effectively, but not enough to the degree where it feels realistic. This simply isn't good enough for convincing immersion. The DTS plugin has demonstrated its ability to portray 3D sound even more realistically than the Unity one through this research experiment. The implications for immersive sound in the virtual reality market are

absolutely massive; more engaging applications in areas such as entertainment can be created. Further work in this area could lead to realistic sound integration within game engines such as Unity or Unreal Engine, which could widely spread the usage of realistic 3D sound in commercial applications. A plugin like the DTS one is just a small stepping stone towards more sophisticated and realistic sound models in portraying sound realistically in virtual environments.

In the future, as the software and hardware that powers these virtual reality devices becomes more and more advanced, we can expect to see immersion being a key part in delivering high quality experiences that make the virtual world feel physical. The path to total immersion is becoming more and more defined as technology is able to integrate each of the five human senses into the virtual world.

References

- Arvin, B., N. Prepageran, and R. Raman. ""High Frequency Presbycusis" - Is There an Earlier Onset?" *Indian Journal of Otolaryngology and Head & Neck Surgery* 65.S3 (2013): 480-84. 30 Nov. 2011. Web.
- Ashihara, Kaoru. "Hearing Thresholds for Pure Tones above 16 KHz." *The Journal of the Acoustical Society of America* 122.3 (2007): n. pag. 21 Aug. 2007.
- Begault, Durand R. Introduction. *3-D Sound for Virtual Reality and Multimedia*. Boston: AP Professional, 1994. N. pag. Print.
- Ip, B. "Technological, Content, and Market Convergence in the Games Industry." *Games and Culture* 3.2 (2008): 199-224. Web.
- Oculus. "Oculus Rift: Step into the Game." *Kickstarter*. N.p., 1 Aug. 2012. Web.
- Parsons, Thomas D., and Albert A. Rizzo. "Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-analysis." *Journal of Behavior Therapy and Experimental Psychiatry* 39.3 (2008): 250-61. Web.
- Pranckevičius, Aras. "Physically Based Shading in Unity." Game Developer's Conference 2014. San Francisco. Web.
- Serafin, Stefania, and Giovanni Serafin. "Sound Design to Enhance Presence in Photorealistic Virtual Reality." *International Conference on Auditory Display* (2004): n. pag. Web.
- Seymour, Neal E., MD, Anthony G. Gallagher, PhD, Sanziana A. Roman, MD, Michael K. O'Brien, MD, Vipin K. Bansal, MD, Dana K. Anderson, MD, and Richard M. Satava, MD. "Virtual Reality Training Improves Operating Room Performance." *Annals of Surgery* 236.4 (2002): 458-64. Web.
- Talagala, DS, W Zhang, TD Abhayapala, and A Kamineni. "Binaural Sound Source Localization Using the Frequency Diversity of the Head-related Transfer Function." *Journal of the Acoustical Society of America*, 135.3 (2014): 1207-1217.
- Zotkin, D.N, R Duraiswami, and L.S Davis. "Rendering Localized Spatial Audio in a Virtual Auditory Space." *IEEE Transactions on Multimedia*, 6.4 (2004): 553-564.

Appendix

- A. Charts from Google Forms Survey
	- a. DTS plugin group
	- b. Unity sound component group
- B. Compilation of Survey Responses from Google Forms
	- a. DTS plugin group
	- b. Unity sound component group
- C. Chosen Graphs of Player Distance from Sound Object over Time
	- a. DTS plugin group
	- b. Unity sound component group

A. Charts from Google Form Survey

a) DTS plugin

Total responses: 28

Have you ever used a virtual reality-enabled device before? (28 responses)

Have you, or do you still, play video games regularly? (28 responses)

How old are you? (28 responses)

Did you encounter any motion sickness during the simulation? (28 responses)

How easy was it for you to determine where the sounds were coming from? (28 responses)

Did you feel like the sound's direction was realistic? (28 responses)

b) Unity sound component

Have you ever used a virtual reality-enabled device before? (14 responses)

Have you, or do you still, play video games regularly? (14 responses)

Did you encounter any motion sickness during the simulation? (14 responses)

How easy was it for you to determine where the sounds were coming from? (14 responses)

Did you feel like the sound's direction was realistic? (14 responses)

B. Compilation of Survey Responses from Google Forms

a) DTS Plugin

Positive Responses

it changed depending on where the box is

Very directional

The direction was somewhat easy to locate when the sound was faint, and pretty easy once I got closer.

It got louder the closer I got to it.

The sound progressively grew louder as I approached the box.

The way you could hear it in one ear before the other made it feel similar to how you tell where a sound is coming from in real life, and didn't have a feeling of being off like you sometimes get from a surround-sound system

The direction and magnitude of the sound

the direction of the sound was clear

Louder as you approached the box

the sound moved through space very effectively. additionally the visuals were immersive.

Closeness, and being able to turn to change the balance between the ears

It got louder/more pronounced when I got closer, would change sound in R/L ear if on that side,

poundingly loud when I was in front of it

It changes when I turned my head

The sound.. and the shading of the boxes when I stood next to them

Left and right panning, and distance

When I got closer, the sound was louder. I could also tell which side it was coming from, so if I approached it from a certain side or another side the sound direction changed as well.

Not sure

Though I turned my head slightly, I could sense the slight directional change of the sound. I couldn't see anything else but the screen, even when I looked around

As I got closer it got louder and the directionality was tellable

direction of sound and how head movement changed the reaction of sound

I could close my eyes and figure out what direction I could hear it coming from. I could use senses other than my sight.

The more I moved around, the more the sound changed in terms of getting louder or softer which made it feel like real life and also made it easier to distinguish where the noise was coming from. The sound actually getting louder as you got closer and the direction it was coming from

Unsure Responses

I had a hard time hearing the sound

The perception did not seem completely realistic but there was some depth and sense of direction to lead me. The sound seemed to come in like increments in change instead of like a smooth gradual change. But then again the sound came every few seconds

didn't sound too clear

I've never tried to analyze the source of sounds like I did during the simulation; I would need to compare the sounds with something more familiar to see how realistic it was.

b) Unity Sound Component

Positive Responses

As I moved my head, so did the screen

visuals, sound, distance

I was able to turn left and hear the sound directly in the left channel. It was pretty easy to determine sound between nearby boxes. The sound distance felt a little linear.

the closer i got to the right box, the louder the sound became

Sounds got louder as you approached the box that maybe was correct

Proximity = loud or quiet. Direction of sound.

Louder as it got closer

That as you turned and got closer, the noise got louder. Conversely, when you were getting farther away, it got quieter.

noise is balanced from each headphone

The degree in which the sound came out of each earbud was realistic

it got louder the closer i got

the textures and environment

When I would turn my head the sound grew closer as I approached each box

Negative Responses

Sounds wise, the left side took over most of the time; telling the left or right direction of the sound was difficult. Depth wise, it was realistic.

C. Chosen Graphs of Player Distance from Sound Object over Time

b. Unity Sound Object

