An Exploration in Ambisonic Audio

An Honors Thesis (MMP 495 sub for HONR 499)

By

Emma Boyd-Ryan

Thesis Advisor

Dr. Christoph Thompson

Ball State University
Muncie, Indiana

May 2021

Expected Date of Graduation

May 2021

Abstract

Ambisonic audio is an immersive audio experience that can be used to enhance the audience's perception of sound. Although this is an older phenomenon it has become popular recently with the use of Virtual Reality and 360 degree videos. This thesis focuses on the exploration of the ambisonic audio topic, the recording of ambisonic audio using primarily the NT-SF1 microphone, and how to encode/decode/use these ambisonic recordings in practice.

Acknowledgments

This thesis would not have been possible without the help of my advisor Dr. Christoph Thompson. Dr. Thompson helped with choosing the research topic for my project and also assisted me with the steps of figuring out how to apply the research of ambisonic audio to the ambisonic audio recordings that I completed. The next person I would like to acknowledge is Noah Sweet for providing assistance on creating my own website to host the information I learned in this thesis. Finally, I would like to thank Cecelia Germann, Ethan Knox, and Anne Zachodni for being my "guinea pigs" for recording some of the ambisonic audio.

Table of Contents

Process Analysis Statement

The first step in this project was choosing what I wanted to research to be and who I wanted my advisor to be. At first, I contemplated many different ideas such as writing and recording an album, creating a piece of software for music, having a recital, etc. Although these were all acceptable ideas, I felt as if they would not give me the experience I wanted for my final thesis project here at Ball State University. I had been working with Netflix prior to this point, and I thought that a good place to look might be in continuing some of the research done for them. After a meeting with Dr. Christoph Thompson we decided on researching and using ambisonic audio for my project. Originally, the idea was to create a "survey" to see which audio formats my peers preferred, but I found that extremely difficult to plan in this time of Covid-19, and instead made my project more of an introduction and exploration into ambisonic audio. The music media production department at BSU has access to several ambisonic audio materials such as the NT-SF1 microphone and the software to decode the signal coming from this microphone, and this made this an attractive project for the semester.

The vast majority of the time I spent on this project was broken down into lab time. I participated in four separate recording sessions for this project (which each took several hours of set-up, tear-down, and the time spent on the recordings themselves), as well as spending time on my computer to learn how to format and work with ambisonic audio (which included learning how to process signals that are ambisonic, making sure my DAW was compatible - more on this later, and waiting for my computer to loudness normalize the final audio formats..

Overall, I think that I learned a significant amount of information about ambisonics while conducting this project. However, much of the information that I learned came from "trial and error" and I do think I learned more about what not to do with ambisonics than what I should do with ambisonics.

Poor time management was also something that I struggled with during this project. I was in the MMP 495 class which helped significantly to keep me on track and continue working on this project, but I still think I struggled to get all the aspects of the project done in time (whether due to Covid-19, exhaustion, or being extremely busy). The further the semester progressed the harder it seemed to be to spend the time necessary on the project.

After choosing the project topic of ambisonic audio it was time to research the subject. I found out that the Rode NT-SF1 microphone was only one of many ways that one could record ambisonic audio, and the scope of my project increased. I learned that you could make different ambisonic arrays using the normal microphones that are housed in the MMP department. Although this gave me plenty of research material, it also made it more difficult to focus on the important parts of the project.

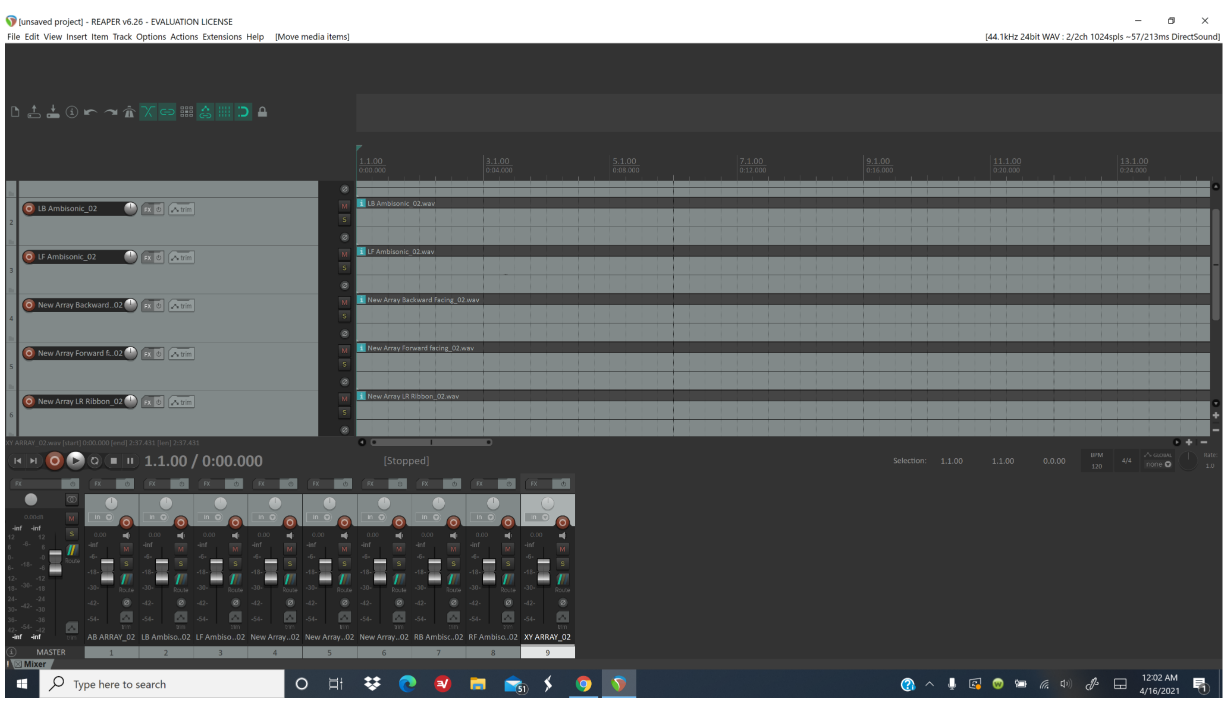
After this, I started thinking about what I wanted to record with the different ambisonic techniques. To begin with I chose to record basic percussion (as a trial run and also just to have isolated drums). This went well, but I learned that I created one of the ambisonic arrays incorrectly, and had to fix this in the next recordings. The other recordings that I used in this project were that of Celia Germann and Ethan Knox walking around the microphone and doing things like clapping, stomping, and saying where they were in the array. I also had a jazz combo record 2 pieces of theirs with the ambisonic microphones only, and I was pleasantly surprised at how well this worked in a jazz combo setting where everyone is in the same room. In all of the recordings I used more than one microphone array, to ensure that there would be a "control" sound to have something to compare the ambisonic recordings to.

The next step of the project was the formatting/editing (encoding + decoding) of the ambisonic recordings to ensure that they would be able to be listened to on a wide variety of speaker arrays. For this project, I focused on the basic ambisonic formats such as the B format FUMA or the B format Ambix, as well as formats that would be compatible with traditional stereo arrays or headphones such as XY or AB.

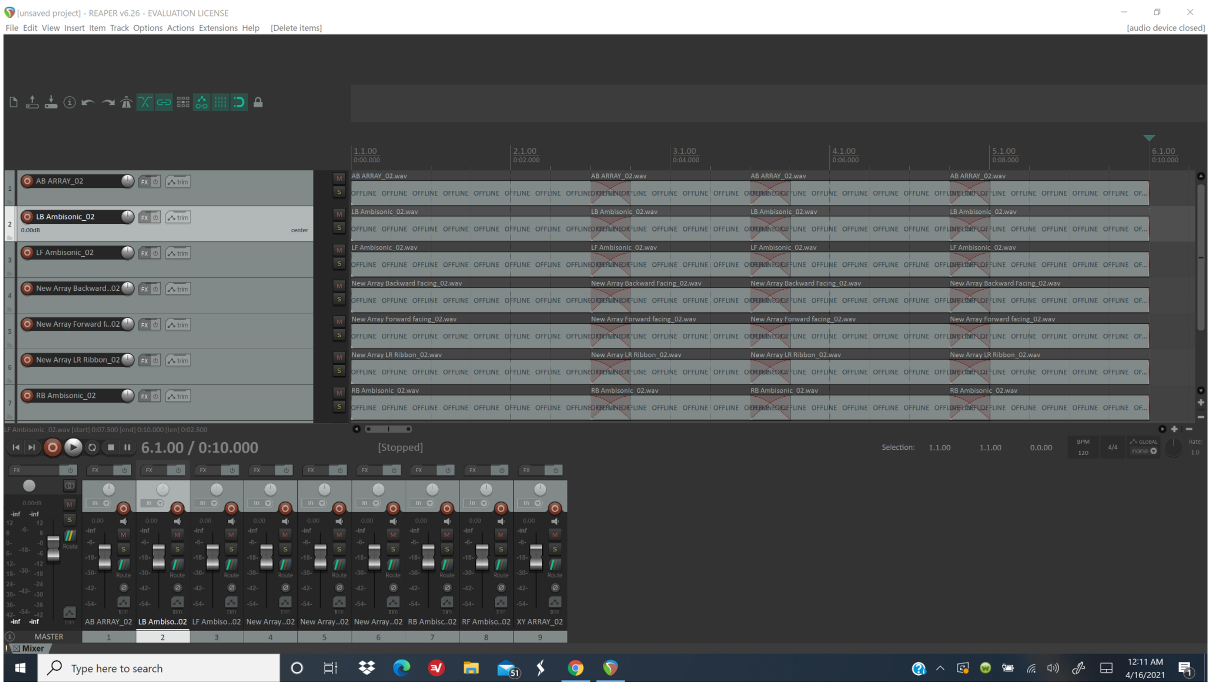
When I first tried to manipulate the ambisonic audio I was using the ProTools and after a while I realized that ProTools does not support ambisonic audio formats. You have to pay for a more advanced ProTools version (like ProTools ultimate). I then did some more research and found out that I could use the DAW Reaper for free for a trial period, and so I ended up using Reaper for this project. I had never worked in Reaper prior to this point, so learning how to manipulate the ambisonic audio through the Reaper DAW proved to be a challenge.

I will document the process of editing the audio files in Reaper below:

Step 1: Putting all of the files in the Reaper session

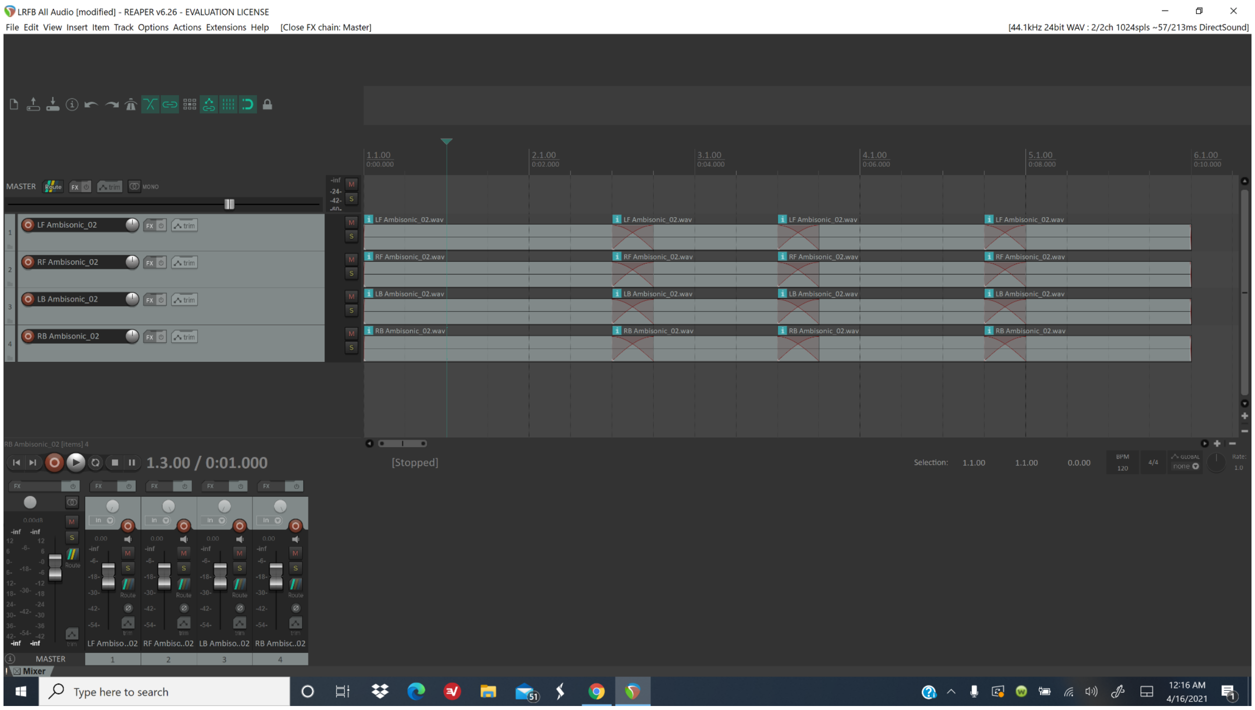


Step 2: The next step is editing all of the audio files to be the same length and feature the same content.



Step 3: I separated the tracks by respective audio types (ie. Rode NTS-F1 array, stereo array, double mid side, etc.). In the next picture below you will see an example of what only the

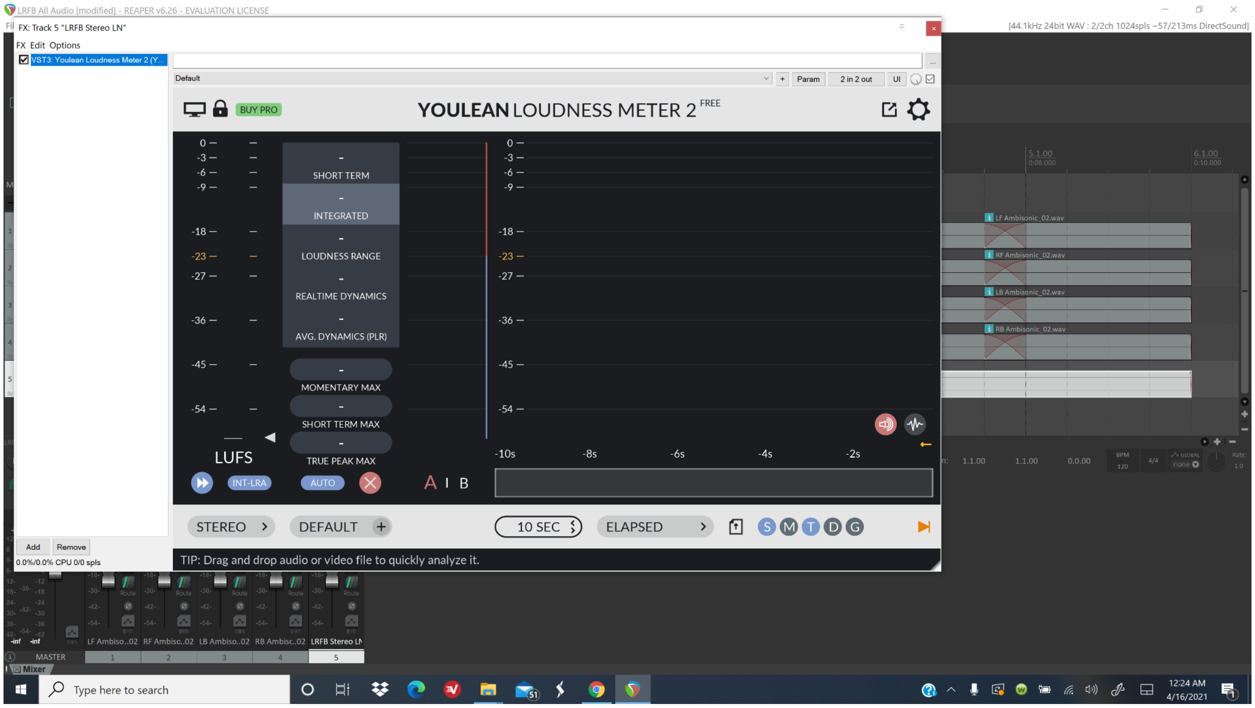
Rode NTS-F1 ambisonic array looks like in reaper.

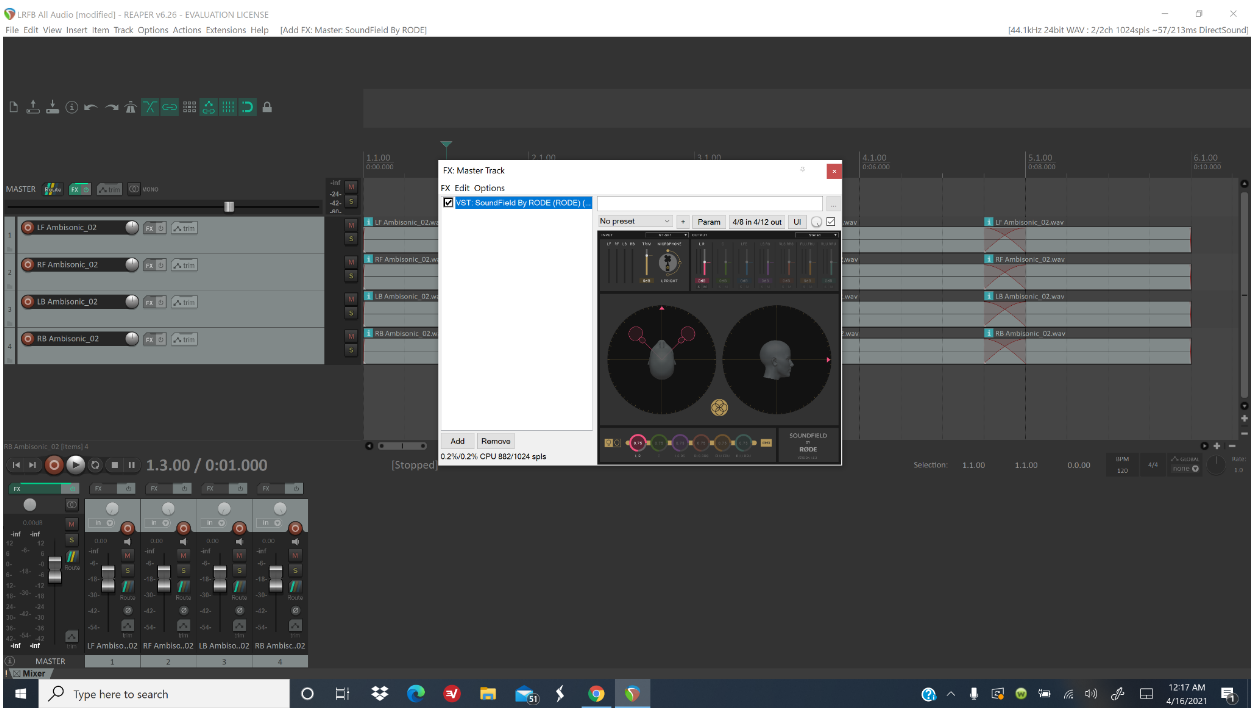


This next picture is what the “plug-in” to work with the ambisonic audio looked like. It was called the SoundField by Rode plug-in.

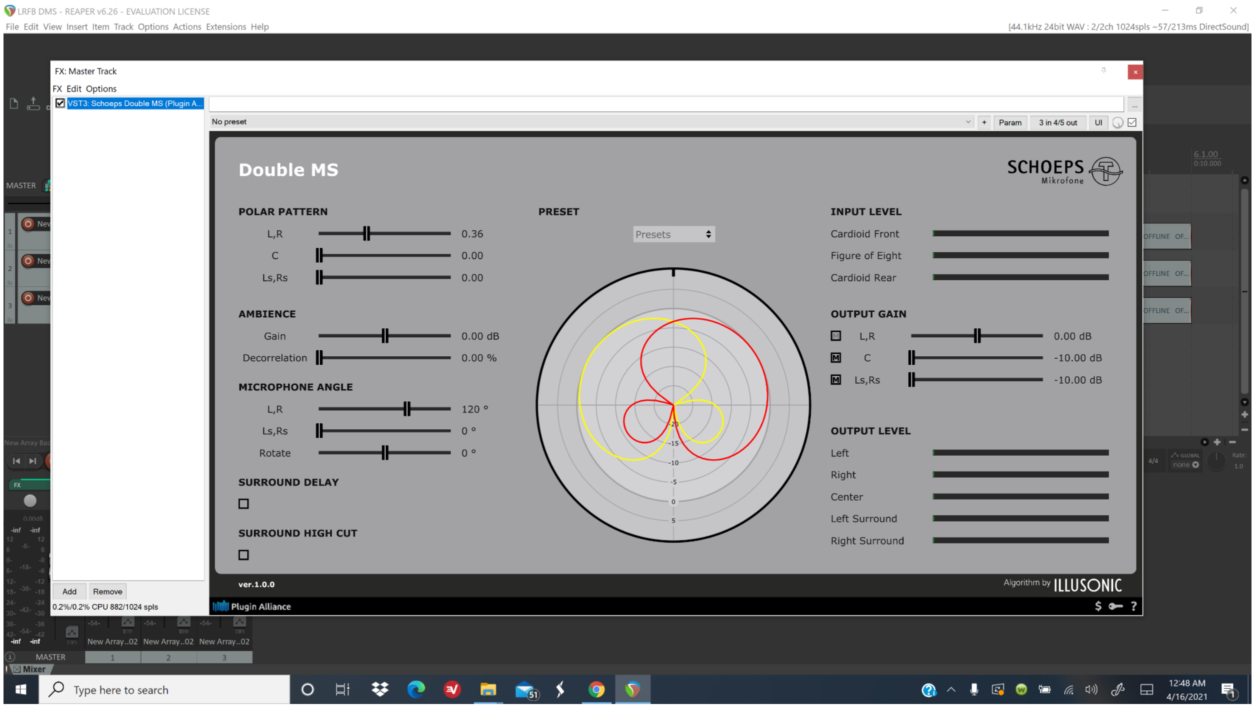
I would loudness normalize each sound with the youlean plug-in, shown below

Here is an example of the ambisonic screen. At this step it was important to make sure the routing was correct (this was done through panning

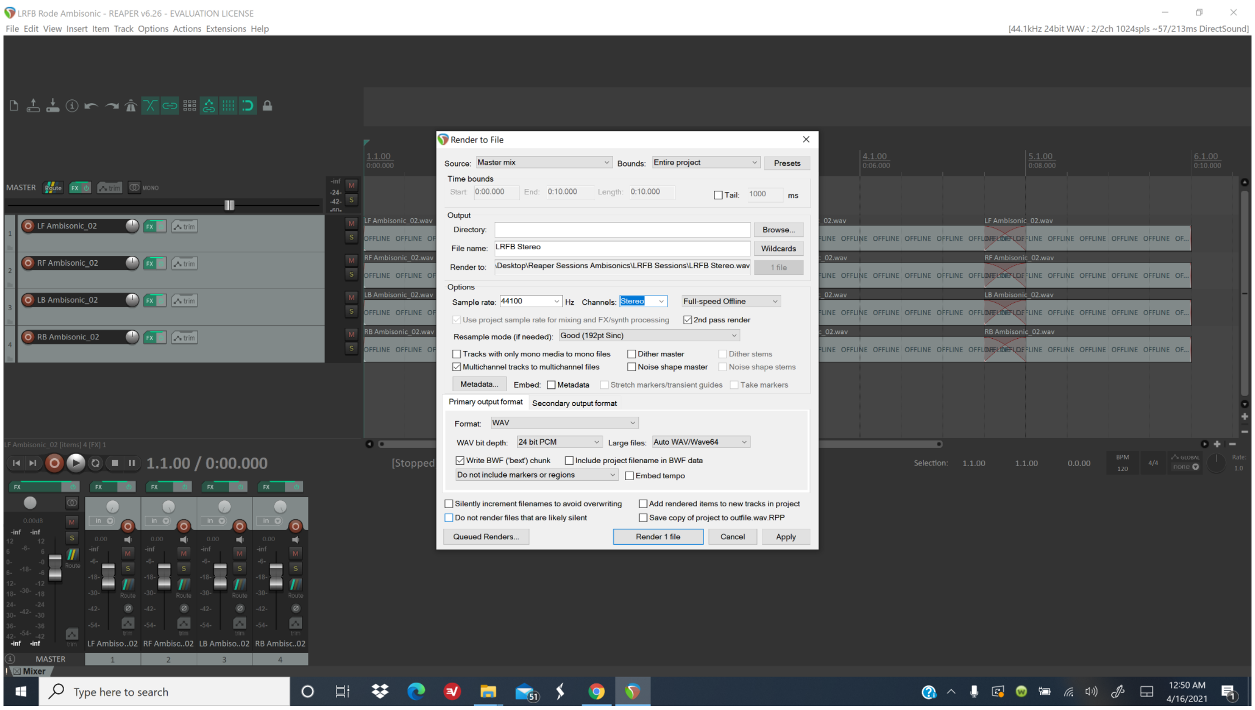




I did this for each “type” of sound. For example, here is the Double Mid Side array.



I did not have to do as much processing on the purely stereo arrays, so I will leave those pictures out. I had to repeat this process with each sound recording group and render them ensuring I was rendering for the correct amount of audio files each time. Here is what the “rendering” box looks like.



This was the biggest learning curve of the project, as I was not used to working in Reaper and was also trying to figure out how to use the ambisonic recordings.

After I figured out how to format the audio I decided to make a sound library with the audio. Previously, I thought that I would be able to have other students “evaluate” the sound and decide if they like ambisonic or traditional arrays better, but found this to not be feasible due to Covid-19 guidelines and not having enough time to create the app that would be necessary to blindly compare audio.

To spice things up I decided to put the audio on my own website (that I coded). I learned a ton from this process as it was still relatively new to me - and I think that this part of the project really helped to expand my views on what career opportunities are open to me post graduation. I can also show this webpage to potential future employees (or other colleges) to showcase some of the work that I did here at Ball State University.

Written Thesis

1. What is an Ambisonic Recording?

An ambisonic recording places its listener in a 360-degree “sound sphere” (Paris, 2020). This sound sphere is intended to create a more convincing listening experience, and perhaps even persuade the listener that they have transported to the sound environment created by the ambisonic recording. Rode (the NT-SF1 manufacturer) mentions that an ambisonic recording represents the "sound field at a point or in space" suggesting that an ambisonic recording is a way of capturing the entire sound environment at one point in time (*The Beginner's Guide To Ambisonics*, 2020).

A unique attribute of ambisonic recordings is that the material recorded using this technique is able to be placed at any location in this sound sphere, which allows for better localization than is possible using traditional types of recordings and pan pots. Traditional recording techniques are not able to record the whole "sound sphere" and instead record directionally and only capture part of this "sound sphere". This means that most of the sound in a traditional recording is sonically placed on the horizontal axis (meaning left or right) rather than anywhere in a sound sphere (meaning up, down, left, right, and combinations of these) (Paris, 2020). Because of the placing of the capsules in an ambisonic microphone, they capture all of the directional information they need to do this when the material is recorded (*The Beginner's Guide To Ambisonics*, 2020). Directional information includes not only the vertical information (meaning height), but also the entire 360 degrees around the microphone (*The Beginner's Guide To Ambisonics,* 2020). According to Zotter and Frank this works because ambisonics provide "spatially undistorted omnidirectional and figure-of-eight" patterns that are created and preserved in the directional mapping from the material recorded with the ambisonic microphone (Zotter & Frank, 2019). This style of recording gives more control over how to "mix" the recording, and also allows for the sonic aspects in the recorded material to be placed more accurately to how they would sound in real life in a real three-dimensional space (Paris, 2020).

2. When and why were ambisonic recordings created?

Ambisonic recordings were originally developed in the 1970s, although they began to gain more popularity recently for use in the immersive audio field (Ambisonic Audio Fundamentals, 2020). The practice of ambisonics was developed by Michael Gerzon, Peter Felgett, and Geoffrey Barton at the University of Surrey and the University of Oxford (*The Beginner's Guide To Ambisonics*, 2020). Michael Gerzon and Professor Peter Craven developed the prototype of the microphone that is now being used for this ambisonics project in 1975 (a SoundField microphone) *(The Beginner's Guide to Ambisonics*, 2020). Head Related Transfer Functions were also something that these scientists were working on in the 1970s, with the goal to filter the sound to reproduce the three-dimensional sounds that a person's ears will naturally hear (Paris, 2020). HRTFs are typically used in binaural applications.

Ambisonic recordings are widely used in many different industries for many different applications. For example, a common use for ambisonic recordings is in the gaming industry- specifically with Virtual Reality (Paris, 2020). They are also used for cinemas, home theatre systems, mobile apps, Dolby Atmos, Auro 3D, Facebook 360, Google, etc. (Paris, 2020). Ambisonics are thought to improve the user experience by making it more realistic. Additionally, digital processing has improved considerably in the past few years to become faster, and it has also become more widely available and at a lower cost. This makes it easier and cheaper to produce more ambisonic content than ever before (Paris, 2020).

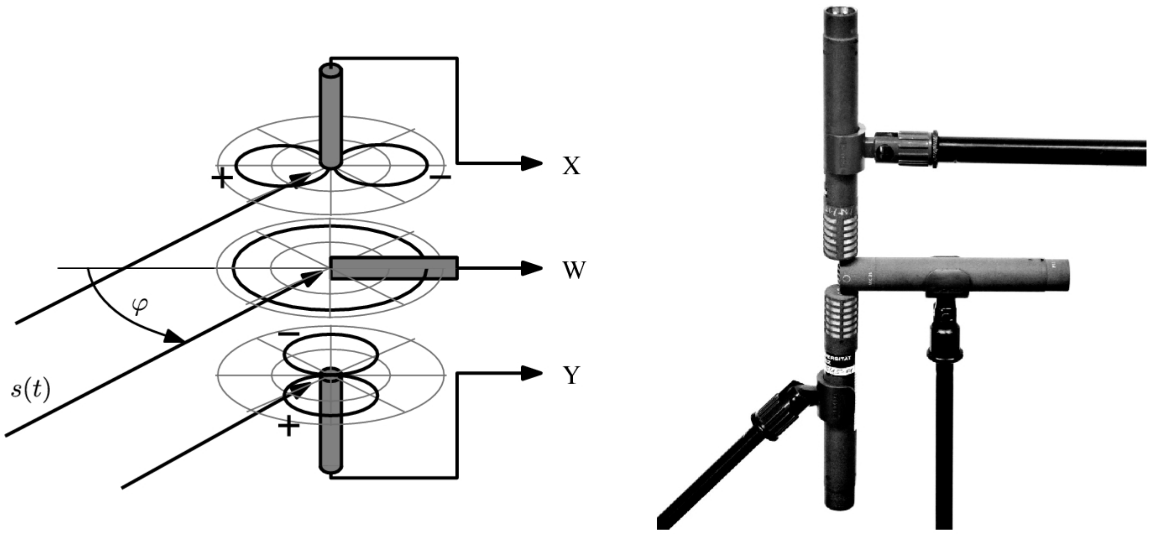
Ambisonics are used as the preferred spatial audio method for several reasons. To begin with, ambisonics are based only on the physical characteristics of the acoustic sound field (Arteaga, 2018). This is important because it ensures that the sound is as realistic as possible. Additionally, ambisonics are "not restricted to single plane waves" which means that it is more flexible and can take into account the entire sound field (Arteaga, 2018). They are also layout-independent (more on this in the next paragraph), and not object-based (which ends up saving memory data because you only need 4-x audio files rather than an audio file for each object that is in the sound field) (Arteaga, 2018).

Ambisonic arrays are not regulated to certain speaker arrays. This suggests that they can be used for a varying number of speakers, different arrays and layouts of these speakers, or with headphones (as if for a binaural experience) (Paris, 2020). The ability to "shapeshift" to fit many different formats and applications is part of what makes ambisonic recordings so attractive. In an ambisonic recording one only needs one ambisonic microphone (like the Rode NT-SF1) to do the work of what could have been many other microphones, which will ultimately make ambisonic recording cheaper than traditional recording.

3. How do Ambisonic Recordings Work?

To begin with, you need an ambisonic array (or microphone) to make an ambisonic recording. For the sake of this project I mainly used a Rode NT-SF1 microphone, as well as a couple of more traditional microphone arrays for comparisons. The Rode NT-SF1 uses what is called a "tetrahedral array," which means that four microphone capsules are used in the microphone (*The Beginner's Guide To Ambisonics*, 2020). However, you can use any array that uses four or more cardioid capsules.

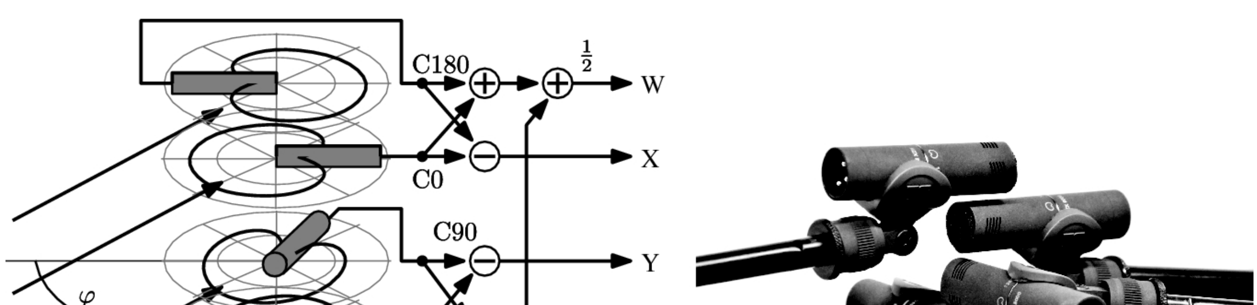
This "tetrahedral array" is also known as a first-order ambisonic (Zotter and Frank, 2019). A first-order ambisonic recording has four channels of audio. The more channels of audio you have the higher the order goes (a second order ambisonic recording has nine channels of audio) (Zotter and Frank, 2019). In general, though, the ambisonic recording becomes more accurate the higher the order is. A zeroth order ambisonic recording (less than 4 capsules/microphones) is thought to contain information about the pressure field at the origin (which is called channel W - more on this later), while acoustic velocity is added with a first order ambisonic recording (these channels are X, Y, Z - more on this later) (Arteaga, 2018). On orders higher than second order information will be added consisting of higher order derivatives of the pressure field (Arteaga, 2018). However, for this project, only first-order ambisonic



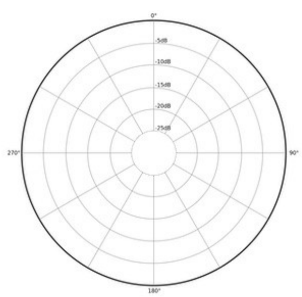
(a) Native 2D FOA recording (b) Picture of recording setup

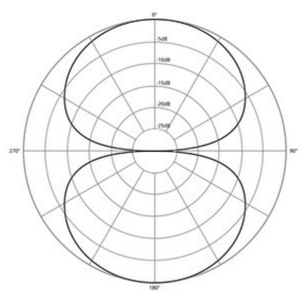
recordings will be used. There are many different arrays one can use to create a first-order ambisonic recording. Some of these arrays are called the "Native 2D Ambisonic recording (the Double-MS recording)" which at its base level is created with two figure-of-eight microphones and one omnidirectional microphone (Zotter and Frank, 2019). A picture taken from Zotter and Frank will be shown below:

Another way to make a first-order ambisonic recording is to create a setup consisting of four 90 degree angled cardioid microphones. A picture of this setup, taken from Zotter and Frank, will be shown below:



The format that naturally comes out of these ambisonic arrays is called the A-Format (also known as the raw format) (*The Beginner's Guide To Ambisonics*, 2020). This A-Format changes depending on which ambisonic array you use, and as thus is not transferable between the different arrays. The four channels in the A-Format are then processed into a matrix, and they are named W, X, Y, and Z (Paris, 2020). Each of these channels represents a specific part of the ambisonic microphone. The W channel is traditionally an omni-directional polar pattern (which means that it will pick up sound equally from all directions) (Paris, 2020). A picture of an omni-directional polar pattern will be provided below:





It is important to note that there are no nulls (or areas where sound is not picked up) in this polar pattern. Then there is the X channel. The X channel creates a figure-8 (or bi-directional) polar pattern that is facing forward and backward (Paris, 2020). Unlike an omnidirectional microphone a bi-directional polar pattern is only equally sensitive to sounds at the front and back (respectively 0 and 180 degrees). It has nulls at 90 and 270 degrees where it will not pick up sound. A picture of a figure-8 polar pattern will be provided above. The Y channel is also a figure-8 polar pattern, but it is instead pointing to the left and the right (Paris, 2020). The Z channel is also a figure-8 polar pattern, but it is instead pointing up and down (Paris, 2020).

The combination of these channels (W, X, Y, and Z) creates the different B Formats. B-format can also be created out of non-ambisonic audio recordings (such as mono, stereo, or multi-channel) (Paris, 2020). These formats are standardized (unlike the A formats) and are thus able to be used for ambisonic processing. The two formats are called the Ambix format and the Furse-Malham format (also called FuMa) (*Ambisonic Audio Fundamentals*, 2020). The Ambix format follows the order of WYZX (and is known as the typical standard for distribution platforms) (Paris, 2020). The Furse-Malham format is more commonly used for audio plug-ins and other processing tools for ambisonics (Paris, 2020). Different services will use different types of the standardized B formats.

4. Encoding, Decoding, and Computer Software

The first step in the ambisonics chain is to encode the audio files. This encoding process is the equivalent of creating a sound field by either creating this sound field out of mono, stereo, or multi-channel recordings, or recording it using a microphone (or a microphone array) that is already optimized for ambisonic recordings (Arteaga, 2018). In the encoding process this typically means taking the A-format recordings and changing them to B-format recordings. When these channels are encoded the pressure and velocity at the origin is also encoded, which is what creates the spatial aspect of ambisonic audio (Arteaga, 2018).

Each soundfield microphone (this is a microphone that records the entire soundfield) will record in a tetrahedral array with nearly-coincident microphones (Arteaga, 2018). In practice this is nearly impossible because the four microphones cannot actually occupy the same physical space. Because of this, there is filtering that needs to be done (especially with the high frequencies) when using the A format of an ambisonic soundfield microphone (Arteaga, 2018).

For the transmission and manipulation of an ambisonic recording one could use an object-based approach. This means that they will send one mono audio track, as well as the metadata that describes spatial characteristics of an ambisonic recording (Arteaga, 2018). However, there are alternatives to this. One of them, called the C-format codifies ambisonic recordings to make them compatible with normal stereo decoders (which are not compatible with raw ambisonic recordings) (Arteaga, 2018). C-format codifiers use the four channels LRTQ instead of the WXYZ channels that are used in the B-format (Arteaga, 2018). In a normal mono decoder, the L and R channels are summed and become mono, and in a stereo decoder these two channels are decodified as stereo (Arteaga, 2018). One can also use what is called a UHJ decoder, which can decode into ambisonic formats (Arteaga, 2018).

There are several strategies to decode ambisonic audio. These decoding methods are separated into physical decoding and psychoacoustic decoding (Arteaga, 2018).

References

Ambisonic Audio Fundamentals (2020) | Simon Hutchinson. *YouTube*. https://www.youtube.com/watch?v=9m4VxFSYSEA.

https://mynewmicrophone.com/the-complete-guide-to-microphone-polar-patterns/#Bidirectional-(Figure-8)

(microphone pictures for polar patterns)

https://postperspective.com/vr-audio-differences-format-b-format/

(differences between A format and B format)

https://vanesaparis.wixsite.com/vanesaparis/post/immersive-audio-ambisonics-and-daw-plugins

Paris, V. (2020, January 30). *IMMERSIVE AUDIO: AMBISONICS AND DAW PLUGINS*. Vanesa Paris.

https://vanesaparis.wixsite.com/vanesaparis/post/immersive-audio-ambisonics-and-daw-plugins.

(immersive audio: ambisonics)

https://www.rode.com/blog/all/what-is-ambisonics#:~:text=Ambisonics%20is%20a%20full%2Dsphere,a%20point%20or%20in%20space.

The Beginner's Guide To Ambisonics. RØDE Microphones. (2020, April 21). https://www.rode.com/blog/all/what-is-ambisonics#:~:text=Ambisonics%20is%20a%20full%2Dsphere,a%20point%20or%20in%20space.

(the beginner's guide to ambisonics)

Zotter, F., and Frank, M. (2019). *Ambisonics. A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Topics in Signal Processing 19.

(ambisonics book)

Digital Supplements (this is where you will post your website)

Photos and other images

Editing Notes:

- You are loudness checking in Youlean on stereo!

- -24 plus or minus 1, -1 plus or minus 1
- Fuma, Ambix, Stereo, 5.1, mono for ALL recordings