

*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/user experience. In an ambisonic recording the sound sphere includes the full 360 degrees, which means that it may provide better localization than a traditional recording. A traditional recording would typically place localization upon a linear point where the audio could be placed anywhere from left to right. Additionally, although ambisonic audio is an older phenomenon (started in the 1970s) 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. It will also look at practical applications for using ambisonic audio.

## **Acknowledgments**

This thesis would not have been possible without the help of several advisors and peers. To begin with, my faculty advisor Dr. Christoph Thompson was an invaluable resource while creating this ambisonic exploration. He not only helped me choose my topic, but also provided feedback on microphone placements and where to look to get the best results. I would also like to thank Dr. Willey for guiding me through the steps required to create a thesis project and present it at the end of the semester. I would also like to acknowledge a few of my peers. First, I would like to acknowledge Noah Sweet for providing me with the tools and assistance required to create my own website to showcase the ambisonic audio materials. I would also like to thank Cecelia Germann, Ethan Knox, and Anne Zachodni for helping me brainstorm and being my "guinea pigs" for helping to record the audio throughout this project.

## **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 (through BSU) 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 reason that this project continues the studies that I have completed at BSU is because it continues previous coursework where I got to work with microphones, hardware related to audio, and software related to audio. This ambisonic project takes all these techniques just a step further, and helps to understand ambisonic applications in the real world. Ambisonic audio is very important in the gaming industry currently, so this may even help me to have a more competitive resume when I graduate from BSU. It also continues the research for Netflix in a way because it is looking at how different audio formats/types alter the overall sound quality. It also required me to think a step further, as I have not had significant exposure to things like software development or web design prior to this project.

My research from this project came from a couple of peer-reviewed books on ambisonic audio and a few online resources from others that have worked with ambisonic audio before.

It was also very important to me to think of a topic that I may even get to use outside of college. Immersive/ambisonic audio is very popular in settings like gaming, museum displays, television, apps, and computers - and because of this I think that knowing the skills involved with it will only benefit me in the long run as I work on kick starting my career.

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.

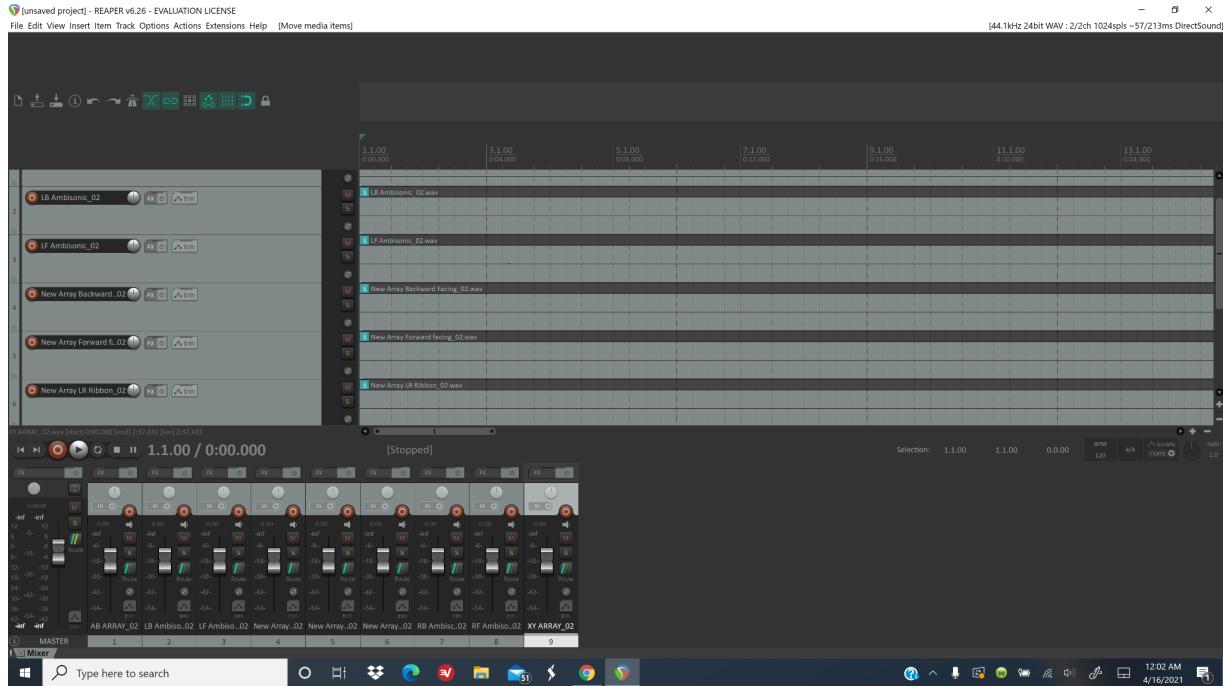
The next step of the project was the formatting/editing (encoding + decoding) of the ambisonic recordings to ensure that they would be able to be listened to on a wide variety of speaker arrays. For this project, I focused on the basic ambisonic formats such as the B format FUMA or the B format Ambix,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

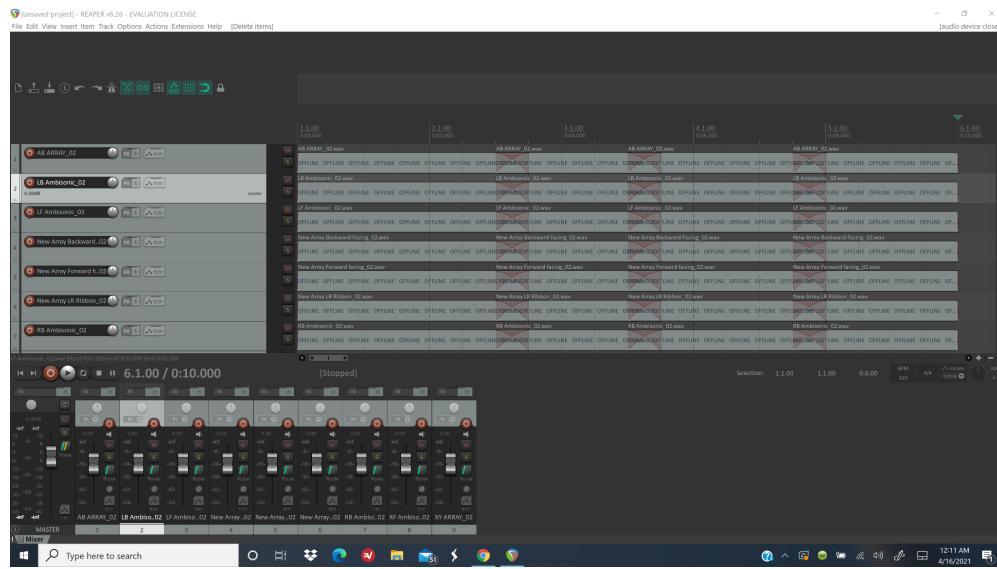
Reaper for this project. I had never worked in Reaper prior to this point, so learning how to manipulate the ambisonic audio through the Reaper DAW proved to be a challenge.

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

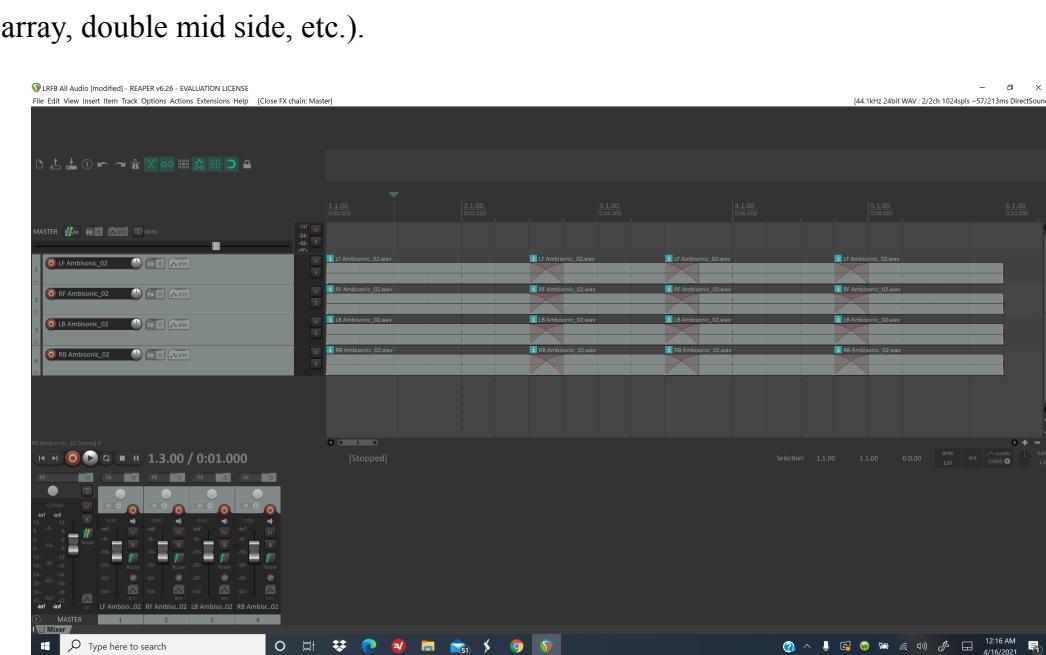
### Step 1: Putting all of the files in the Reaper session



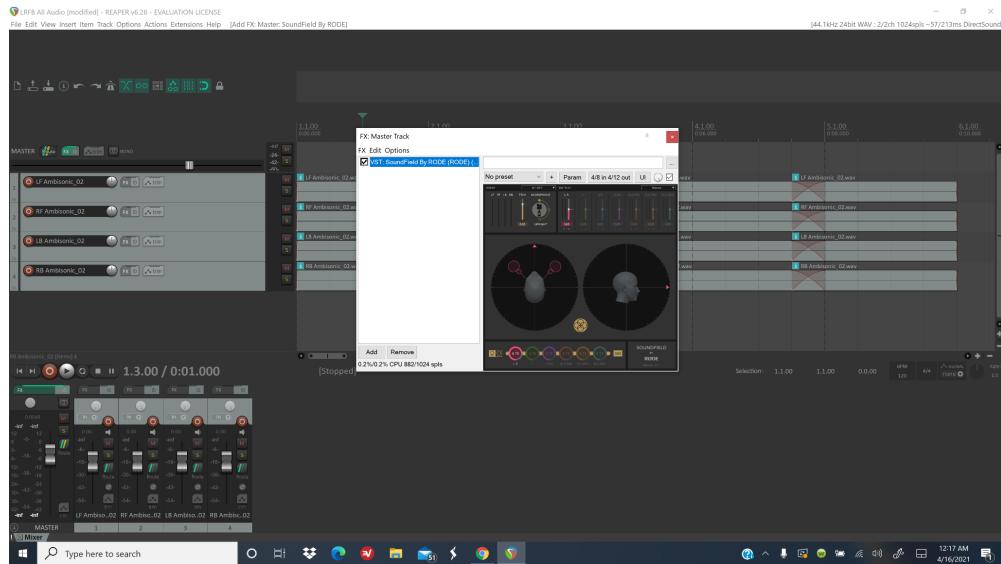
Step 2: The next step is editing all of the audio files to be the same length and feature the same content, which is shown below:



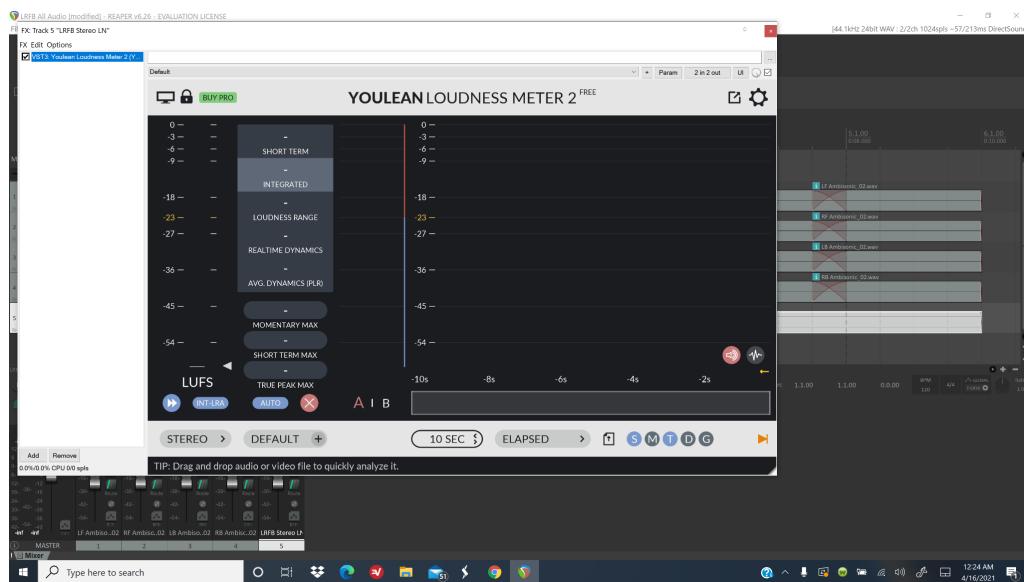
Step 3: The tracks were separated by their respective audio types (ie. Rode NTS-F1 array, stereo array, double mid side, etc.).



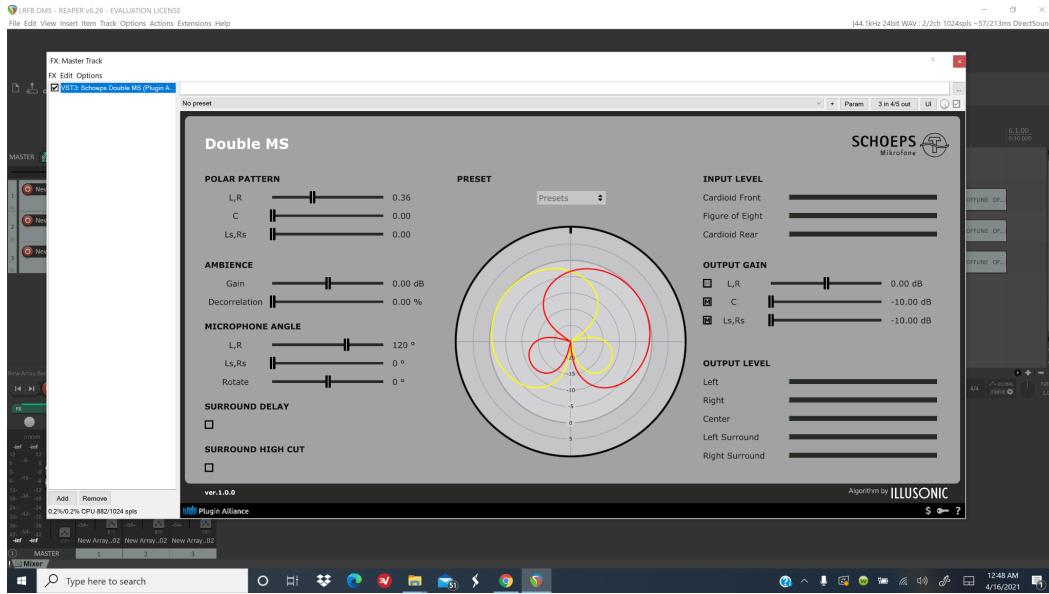
Step 4: This next picture is what the SoundField plug-in by Rode looked like within the Reaper app.



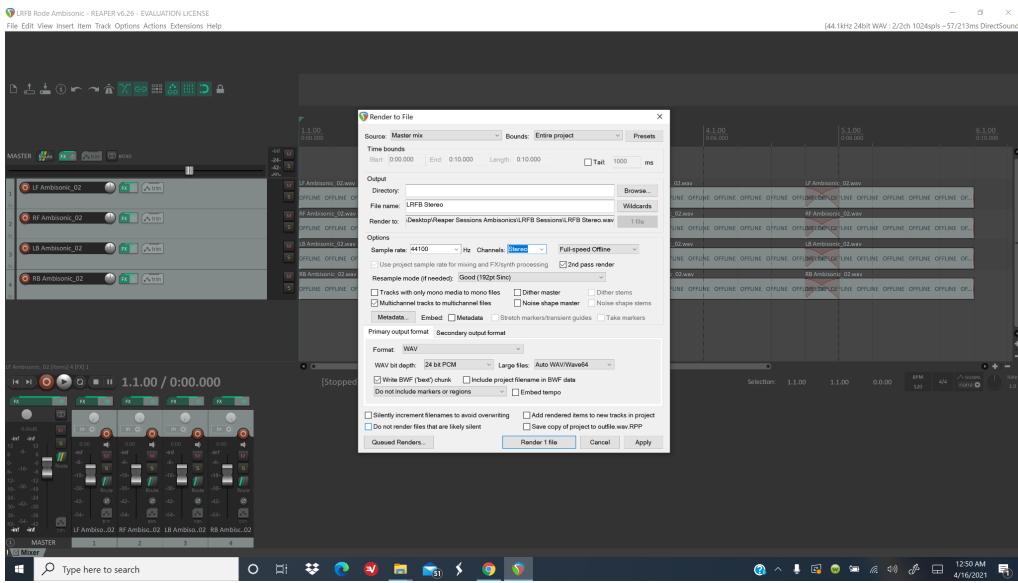
Step 4: After ensuring that all of the routing was correct (on each individual track, the masters, the panning, and fed into the SoundField plug-in I would then loudness normalize the selections. For this loudness normalization I used the Youlean plug-in, which is shown below:



Step 5: This process was completed for each sound and type of sound that is used in the project (using different plug-ins for different arrays). I thought personally that this was the hardest part of my ambisonic project because it took a significant amount of time to figure out how to make the DAW, the plug-ins, and the audio itself interact. Another plug-in that I used (specifically for the Double Mid Side technique) is shown below:



The typical stereo arrays (such as XY and AB) did not require as much processing as the other ambisonic arrays. Instead, I just had to make sure they were panned correctly and were loudness normalized. For the sake of consistency, I made the loudness normalization specs for this project around -24 LUFS, and -1 true peak. After ensuring that the correct plug-ins were used and the loudness was normalized I had to “render” or bounce my audio files. This ensured that they were cohesively in a single audio file rather than separated into several different audio files. The box for the rendering process will be shown below:



When I first started this project I intended to create a survey to ask participants which audio they favored (between stereo XY or AB, double mid side, or the Rode NT-SF1 microphone). This would have been done in a similar way to the Netflix class I was in the previous semester where our job was to look for artifacts and inconsistencies in the audio accompanying movies and shows. This test was done blind with Netflix, and I soon realized that this end to my project was not feasible due to the need to create my own app for this, and also because of Covid-19 restrictions (and people's fear about participating during the Covid-19 pandemic).

Instead of doing this "survey" I instead decided to create a sound library with all of the audio that I collected. This seemed like a better way for me to get other people to listen to the ambisonic recordings and compare them. However, this study would have been biased if conducted as a survey because it was not a blind study, and the audio files would be clearly labeled. To make this more accessible to more people I decided to instead create a website that could showcase all of the audio that has been collected throughout the semester. This way, the

audience could directly compare how the different arrays had captured the sound. If they have keen enough ears, they might even be able to identify some artifacts.

For this part of my project I teamed up with Noah Sweet (a computer scientist going to school at IUPUI) to work on a website to publish my audio files. I thought that this would be a better way to showcase the information I have collected because I would learn more about how to work with technology and software- which is a skill category that I really need to enhance. I think that creating my own website really helped me to understand how much work goes into creating content like websites. We stuck to a fairly simple design, but it still took a long time to get the website up and running.

Overall, I think that I learned a significant amount while conducting this project. It helped me with some of the simpler aspects of a project such as time management, and not creating too wide of a topic for my project. It also helped me to learn more about the technical side of ambisonic audio and how to manipulate and work with ambisonic audio. Although the project ended up more as an exploration than a survey I still think that I accomplished what I wanted to accomplish with this project. I had several “aha” moments when working with Reaper, as there were aspects of signal flow or using plug-ins that I didn’t have to worry about when using ProTools, but learned how to use them more efficiently in Reaper. I also think this project is beneficial to the audio world as a whole, because it serves to make more exploratory audio formats like ambisonic audio more commonplace and accessible to more audio technicians.

## 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 but specifically 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 (or localized) at any location in this sound sphere, which allows for better localization than is possible using traditional types of recordings (like XY, or AB, or using just one microphone) and panning. 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 spherical localization 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 degree sphere 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 is thought to give more control over how to “mix” the recording, and also allows for the sonic aspects (similar to objects) 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 a technique 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 to trick them into perceiving spatially (Paris, 2020). HRTFs (Head Related Transfer Functions) are commonly 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 video games/experiences (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 like a real-life soundscape. Furthermore, digital processing has improved considerably in the past few years to become faster, which resulted in it becoming more widely available and also at a lower cost. This makes it easier and cheaper to produce more ambisonic content than ever before (Paris, 2020).

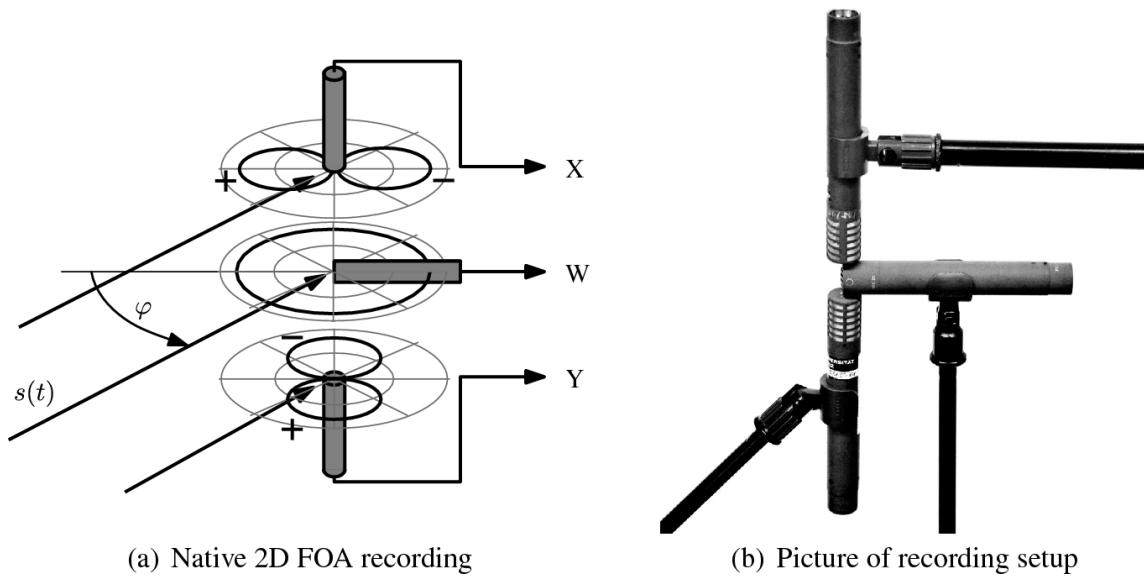
Ambisonics are often used as the preferred method for spatial audio for several reasons. To begin with, ambisonics are based only on the physical characteristics of the acoustic sound field (Arteaga, 2018). This is important because it ensures that the sound is as realistic as possible (and not altered with synthesized sound). 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 (instead of just being placed somewhere on the continuum from left to right) (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 one only needs 4-x audio files rather than an audio file for each object that is in the sound field) (Arteaga, 2018).

Ambisonic arrays are unique because they 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). This ability to “shapeshift” to fit many different formats and applications is part of what makes ambisonic recordings so interesting and attractive to those wanting to improve the directionality of their audio. In an ambisonic recording one only requires one ambisonic microphone or an ambisonic array (like the Rode NT-SF1 or the double mid side technique) to do the work of what could have been many other microphone capsules and other microphones, which could ultimately make ambisonic recording cheaper than traditional recording because less gear is required.

### 3. How do Ambisonic Recordings Work?

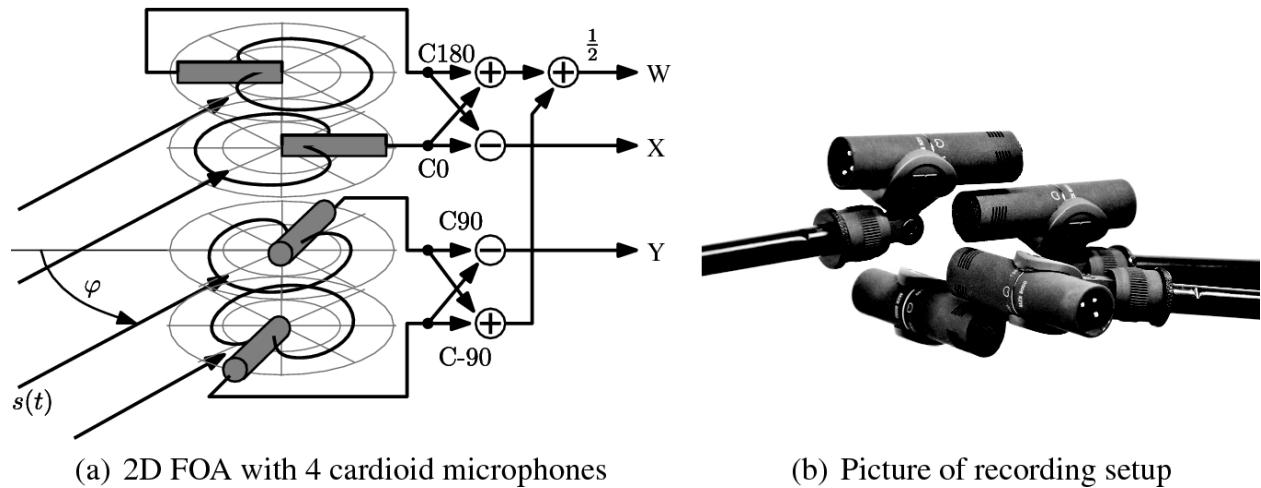
To begin with, one needs an ambisonic array (or microphone) to make an ambisonic recording. For the sake of this project I focused on the Rode NT-SF1 microphone (as well as a double mid side array), as well as a couple of more traditional microphone arrays for comparisons (like XY and AB). The Rode NT-SF1 uses what is called a “tetrahedral array,” which means that four cardioid 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 for an ambisonic recording.

This “tetrahedral array” (meaning the Rode NT-SF1 microphone) is also known as a first-order ambisonic array (Zotter and Frank, 2019). A first-order ambisonic recording has at least four channels of audio but less than nine. 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).

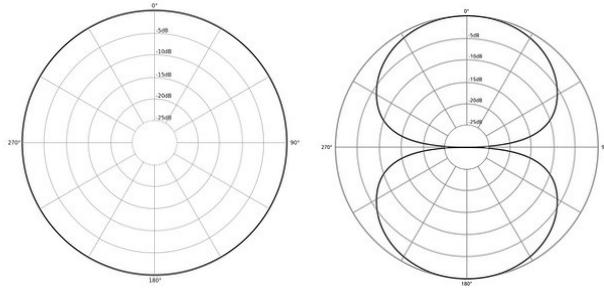


However, for this project, only first-order ambisonic recordings will be used. There are many different arrays one can use to create a first-order ambisonic recording. Some of these arrays are called the “Native 2D Ambisonic recording (the Double-MS recording)” which at its base level is created with 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. An example 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 (or directional information) 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 decoded as stereo (Arteaga, 2018). One can also use what is called a UHJ decoder, which can decode into ambisonic formats (Arteaga, 2018).

Regardless of which decoding method one chooses to use, it is evident that ambisonic material in general is extremely flexible and useful in a large variety of situations.

## References

- Altman, Randi, and Randi Altman. "VR Audio - Differences between A Format and B Format." postPerspective, June 15, 2017.  
[https://postperspective.com/vr-audio-differences-format-b-format/.](https://postperspective.com/vr-audio-differences-format-b-format/)
- Ambisonic Audio Fundamentals (2020) | Simon Hutchinson. YouTube.*  
[https://www.youtube.com/watch?v=9m4VxFSYSEA.](https://www.youtube.com/watch?v=9m4VxFSYSEA)
- Arthur. "The Complete Guide To Microphone Polar Patterns." My New Microphone, April 25, 2021.  
[https://mynewmicrophone.com/the-complete-guide-to-microphone-polar-patterns/#Bidirectional-\(Figure-8\).](https://mynewmicrophone.com/the-complete-guide-to-microphone-polar-patterns/#Bidirectional-(Figure-8).)
- 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.>
- 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>.
- 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.

## **Digital Supplements**

The link attached to this page is the link to the website that I coded with Noah Sweet with the ambisonic sound library created throughout the semester.

Here is the link:

<https://ebr8599.github.io/>

## Photos



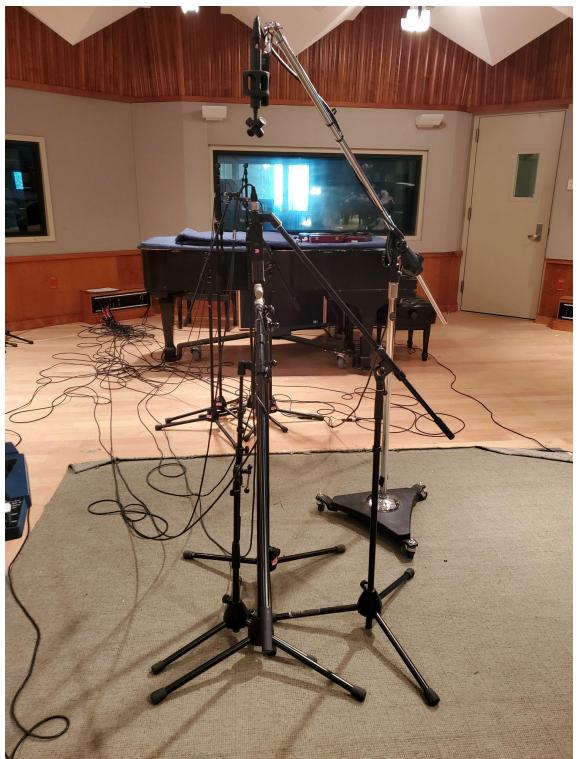
This photo was taken during the Jazz Combo recording session. You can see the Rode NT-SF1 microphone as well as a double mid side array.

This photo was taken during the directional session. It shows an XY and an AB stereo array.



This is an example of a double mid side array.

This is the Rode NT-SF1 microphone and a double mid side array.



This is an example of the entire ambisonic setup.



This is a side view of the entire ambisonic setup.