An Exploration in Ambisonic Audio

An Honors Thesis (MMP 495)

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 format 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 around the microphone, which means that it may provide more advanced localization than a traditional recording. A traditional recording would typically place localization on a line where the audio would be placed anywhere from left to right but would not include any height information. Additionally, although ambisonic audio is an older format (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

Process Analysis Statement

   The first step in this project was choosing the research topic 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 current gaming industry, and audio technology is always progressing and moving forward. It also continues the Netflix style research in a way because it is looking at how different audio formats/types alter the overall sound quality. The project required me to think about new and important skills, as I have not had significant exposure to things like software development or web design prior to this point.

While brainstorming for this project it was also important to me to think of a topic that I may be able to pursue 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 becoming familiar with the skills involved with ambisonic audio will only benefit me in the long run – especially as I work on kick starting my career.

My research for this project included several resources such as peer-reviewed books on ambisonic audio and some online resources that detail how the microphone placements worked. I also made use of several tutorials where others have detailed their experiences with ambisonic audio.

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 (each took several hours of set-up, tear-down, and the time spent on creating the recordings themselves), as well as the time needed to learn how to format and manipulate ambisonic audio which included steps like 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.

After selecting the project topic of ambisonic audio the research began. I learned 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. Ambisonic audio arrays can even be made using the "normal" microphones that were regularly available 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. I decided to look at only one other ambisonic audio array called the double mid side technique.

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 to hear the localization on an isolated drum track). This went well, but I learned that I created the double mid side ambisonic array incorrectly, and had to remedy this in the next recordings. However, recording with the Rode NT-SF1 microphone went smoothly and I could hear a difference in localization between those recordings and the normal spaced arrays that I used.

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 relation to the microphone) in the array. Something that we learned during this part of the project is that the localization patterns on the microphone do not necessarily match where your objects are in the room.

I also had a jazz combo record two 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. All of the instruments were easily audible, and I thought that you could in fact localize where the instruments were. This was surprising to me because I thought that the recording would sound muddy and the instruments would blend together. In all of the recordings created for this project 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 attempted using the DAW (digital audio workstation) ProTools but I eventually realized that ProTools does not support ambisonic audio formats. In order to use ProTools for ambisonic audio one must pay for a more advanced ProTools version (like ProTools ultimate – which is out of my budget). 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. Reaper is also considered more of an "open source" DAW, so it has less "bells and whistles" than ProTools, but this can seem more complicated if you are not used to working in Reaper.

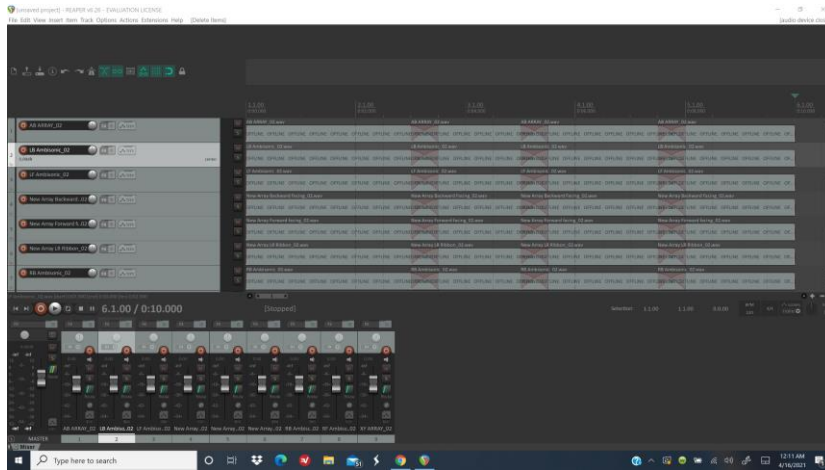I will document the process of editing the audio files in Reaper below (see Figures 1-7):
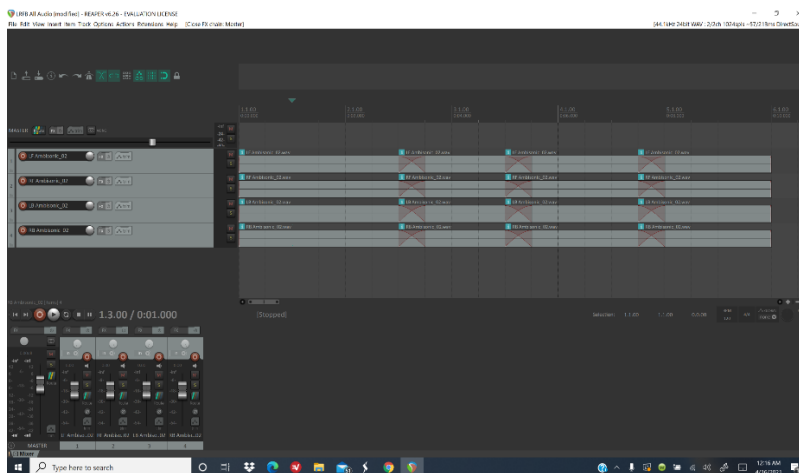
Figure 1

Audio Files In Reaper
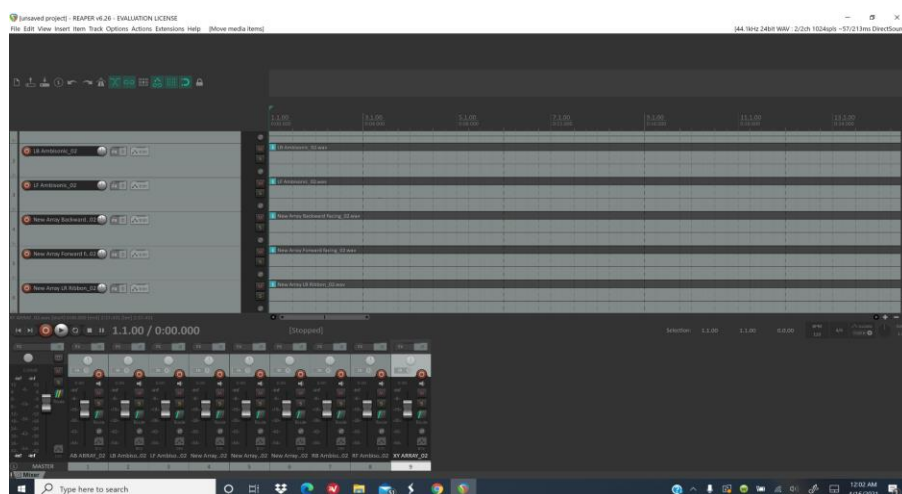
Note: This picture shows what it looked like when I uploaded the audio files for one recorded segment into the DAW Reaper.

Figure 2

Editing of Length and Content

Note: This figure shows what it looked like in the DAW reaper when the audio files were edited

to be the same length and to feature the same content as one another.

Figure 3

Track Separation

Note: Figure 3 shows how the tracks in the DAW Reaper were separated by audio format type.

At this point I would ensure that the routing was correct for each audio type; which was mainly

done with panning.

Figure 4

SoundField Plug-In
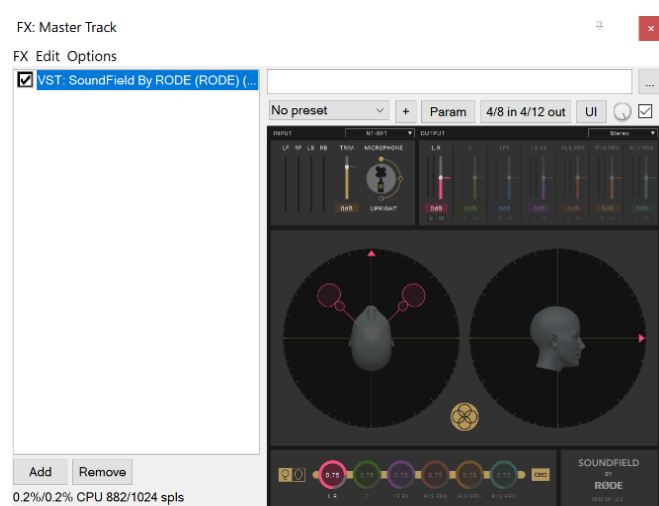
Note: This figure shows an example of the SoundField plug-in. This is the plug-in that I used to manipulate the audio that was recorded from the Rode NT-SF1 ambisonic microphone.

Figure 5

Loudness Normalization Plug-In

Note: Figure 5 shows the Youlean Loudness Meter 2 plug-in, which was used for loudness normalizing the audio used in this project. The audio needed to be normalized in stereo formats due to the incompatibility of the meter with ambisonic formats.

Figure 6

Schoeps Double Mid Side Plug-In
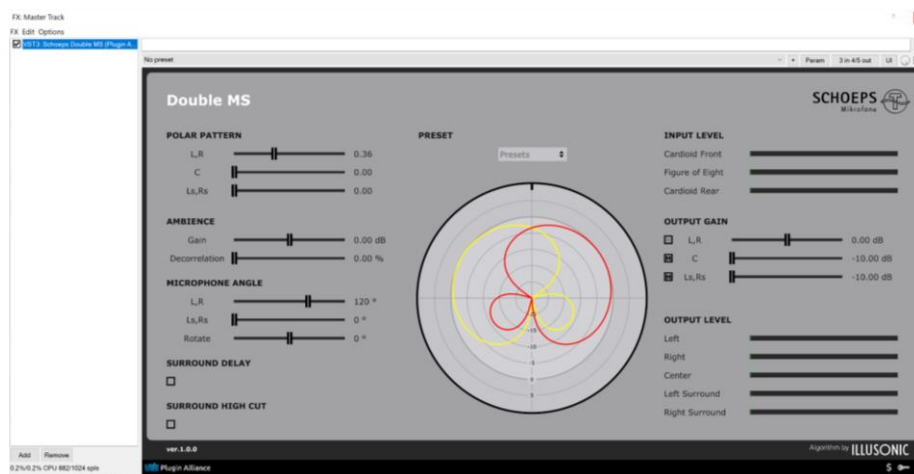
Note: This plug in was used to format the double mid side recordings from this project. It is the equivalent of the SoundField plug-in, but for the double mid side technique.

This process of working on the audio files was followed for each sound selection and each format of sound that was used in the project (using different plug-ins for different arrays). I personally thought that this was the hardest part of the 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.

The typical stereo arrays (such as XY and AB) did not require as much processing as the ambisonic arrays (the Rode NT-SF1 and the double mid side). 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 compacted the audio files so that they were cohesively in a single

audio file rather than separated into several different audio files. The software dialogue for the

Figure 7

The Software Dialogue for the Rendering Process
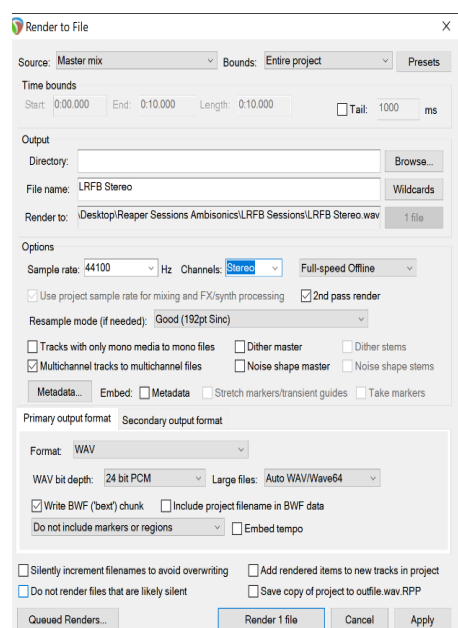
Note: This is the pop-up box that appears when you are rendering your sounds. There are lots of

options at this point like choosing how many audio files you want to render (stereo, quad, etc.),

choosing your sample rate, naming your audio files, etc.

When I first started this project I intended to create a survey after rendering the audio

files to ask participants which audio format 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 (meaning that you did not know which type of audio was which), and I realized that this

end to my project was not feasible due to Covid-19 restrictions.

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. I would not have been able to use this audio library as a means to conduct the survey because it was not a blind study, and the audio files would be clearly labeled. Instead, to make this audio library more accessible to more people I decided to create a website. 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 could have used a resource like Wix to create a website, but I really had no prior experience with building a website and thought that this might be important information for me to learn. This way I would learn more about how to work with technology and software- which is a skill category that is growing in the audio workforce.

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. There were also several edits that we did to the website – like changing how the audio plays – that took a longer time to figure out than I would have anticipated.

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. 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 audio technicians.

Much of the information that I learned while conducting this project was also on a "trial and error" basis where I would try something and then have to re-trace my steps to figure out why it did not work. This was a valuable lesson for me, as previously to this I had not had much of an opportunity to make my own mistakes and learn how to fix them in an academic setting.

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. Despite these obstacles, I still completed my project – and that is something that I am proud of.

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 by providing more enhanced localization than a traditional recording – which can make the sound sphere sound more realistic 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 can allow for more accurate 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). It is important to note that this difference does not make ambisonic formats better, just more suited to certain applications like virtual reality. Because of the placement 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 this project I focused on the Rode NT-SF1 microphone (I also included a home-made 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 cardiod 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. The double mid side technique that I used is also an ambisonic array, but it does not have a microphone that is pointed upwards. This means that it does not have the same height information that the Rode NT-SF1 microphone would provide.

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. One of these arrays is 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:

Figure 8

"Native 2D Ambisonic recording (the Double-MS recording)"

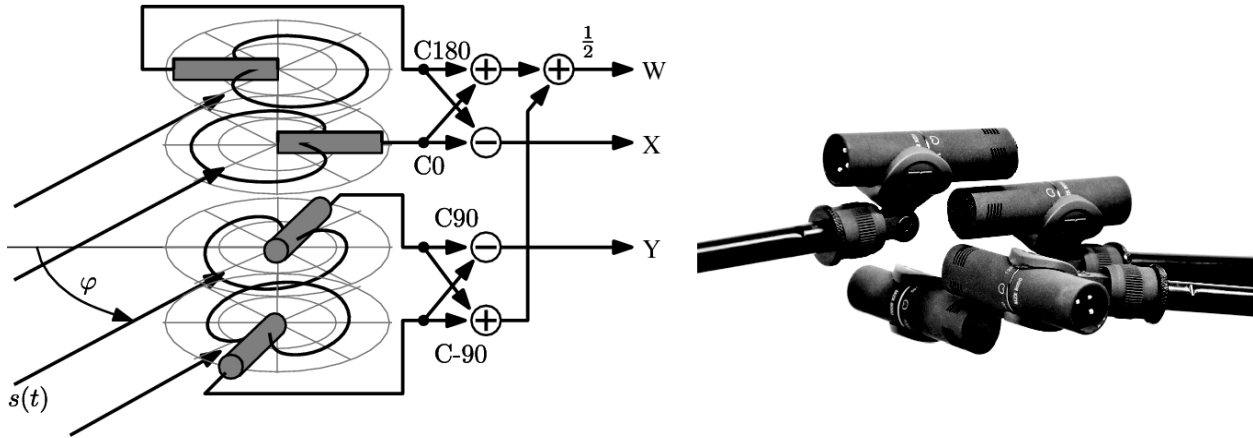(a) Native 2D FOA recording   (b) Picture of recording setup
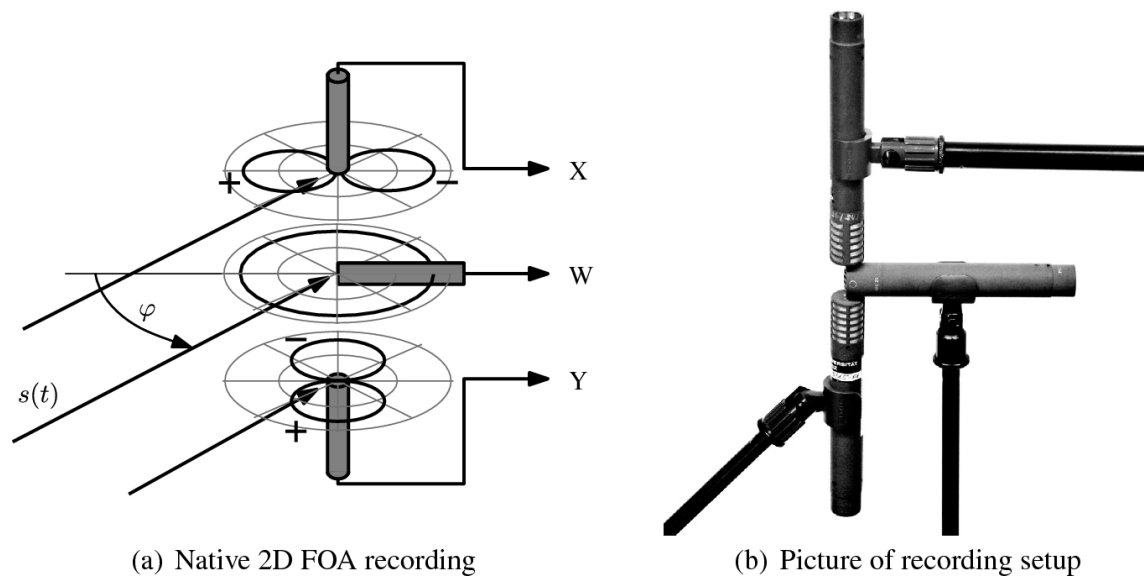
Note: This is an example of how to set up the first-order ambisonic array mentioned above. From Zotter, F., and Frank, M. (2019). *Ambisonics. A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Topics in Signal Processing 19.

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 will be shown on the next page:

Figure 9

*"2D FOA with 4 cardioid microphones"*

| (a) 2D FOA with 4 cardioid microphones | (b) Picture of recording setup |

Note: Figure 9 shows how one would set up the ambisonic array that requires 4 cardioid

microphones. From 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.

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:

Figure 10

Microphone Polar Patterns
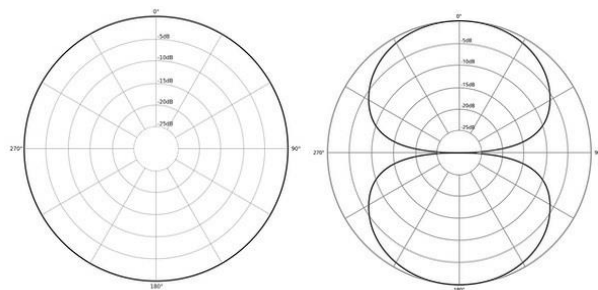
Note: These are examples of omni-directional and figure-8 polar patterns. From Arthur. "The

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).

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-

formats 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 decodified as stereo (Arteaga, 2018). One can also use what is called a UHJ

decoder, which can decode into ambisonic formats (Arteaga, 2018).

5. Conclusion

As is evident throughout this paper, ambisonic audio is an immersive audio format that is

extremely flexible and useful in a large variety of situations. The 360 degree "sound sphere"

created from the directional information used in ambisonic audio creates an immersive audio

experience that is commonly used for applications like virtual reality, mobile apps, Facebook

360, etc. There are several ways to record ambisonic audio including but not limited to the Rode

NT-SF1 microphone and home-made arrays like the double mid side array, as well as many ways

to apply the recordings in formats like Ambix, FuMa, stereo, and 5.1. Ambisonic formats can be

played back on a large variety of speaker arrays, and there are also several ways to encode and

decode this audio. This flexibility makes ambisonic audio an extremely attractive spatial and

immersive audio candidate. Although many ideas connected to ambisonic audio have been

discussed in this paper this is merely the beginning of an exploration. Simply put, the

possibilities of ambisonic audio are nearly endless. There is much more to learn – and more

importantly – much more to create.

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/.

Ambisonic Audio Fundamentals (2020) | Simon Hutchinson. *YouTube*. 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).

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 below is the link to the website that I coded with Noah Sweet with the ambisonic sound library created throughout the semester:

       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.