A New Technique for Capturing True Coincidence in Ambisonic Microphone Arrays

Hello everyone, my name is James Keary, and my thesis project is called … For an introduction, I need to give a brief basics on Ambisonics so that terminology and motivations for my project will be understood.

Ambisonics Basics

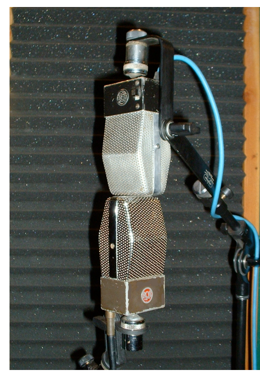
- Surround sound recording, mixing and playback system
- Faithfully recreate a soundfield up to full sphere 3D within a sweetspot
- Fully backwards and forwards compatible
- Not commercially accepted

Ambisonics is a surround sound recording, mixing, and playback system developed in the 1970s by Michael Gerzon and others at the NRDC in the UK. The NRDC's goal was to faithfully recreate the full 3D soundfield within a sweetspot. It was designed to be fully backwards and forwards compatible for any playback configuration from mono to full sphere surrounded by hundreds of speakers! It was designed for use in music and film, but was not widely and commercially accepted for many reasons. It was passed over instead for discrete channel mixing practices such as 5.1, 7.1, and Dolby Atmos which came out recently for 128 discrete channel mixing for movie theatres.

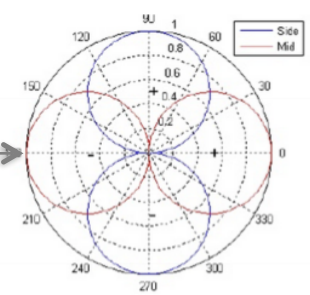
Ambisonics Recording

- Alan Blumlein’s 1931 patent, *Improvements in and Relating to Sound-transmission, Sound-recording and Sound-reproducing Systems*. Establishes the coincident microphone technique.
- 1975, Gerzon’s *The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound* expands the idea to capture full sphere 3D soundfield.

mid side

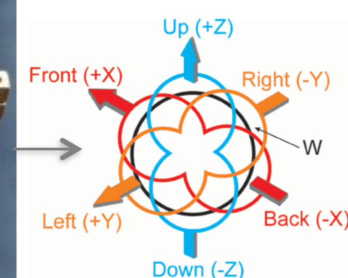


2D



soundfield

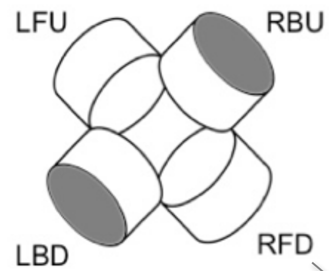
3D



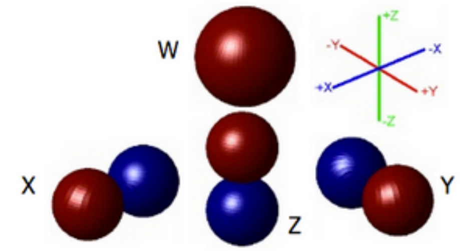
the ideas of ambisonics can be traced all the way back to Alan Blumlein's 1931 patent. In 1931 Alan Blumlein had a patent for many different stereo recording, mixing and playback techniques. In that patent he established what we now refer to as the coincident microphone technique, an idea in which many microphones placed close together and at different orientations act together as 1 to build different parts of the soundfield. So for example, in Blumlein’s mid side pairing microphone technique, 2 mics are positioned 90° of axis from each other, the front facing mic captures the direct sound, and side facing mic captures the ambience. Blumlein wrote about expanding this idea to capture other axes, but this never happened. But in 1975, Gerzon figured out the math and physics to make it possible.

Ambisonics Mixing and Reproduction

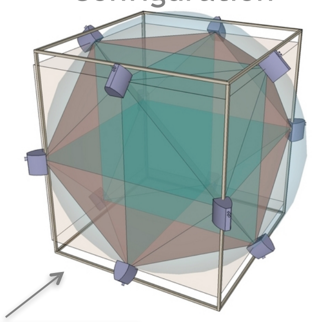
A format



B format



Configuration



W =	LFU + RFD + LBD + RBU *sum of all capsules
X =	LFU + RFD - LBD - RBU
Y =	LFU - RFD + LBD - RBU
Z =	LFU - RFD - LBD + RBU

Matrixing Process

Regular	Irregular
Mono	5.1
Stereo	7.1
Quad	9.1
Cube	Etc.
Octahedron	
Bi-Rectangle	
Full Sphere	

Decoding Process

The soundfield microphone records 4 channels, known as A format. The genius of the ambisonics mixing design is that it separates the transmission channels from the actual speaker feeds. Now, what does this mean, well, these 4 channels in A format are matrixed in a series of sum and difference formulas into what are known as B format, the polar patterns of which we see here. The B format signals are our transmission channels, they can be added at different amounts to whatever speaker configuration is available in a process known as decoding.

Project Motivations

- *"The microphones are effectively coincident up to about 7 kHz and are subjectively well behaved above that"*
  - Gerzon, 1975, **The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound**
- *"this system fails at higher frequencies."*
  - J. S. Bamford and J. Vanderkooy, 1995, **Ambisonic Sound for Us**
- *"Ambisonics has undeniable limitations in terms of imaging stability and size of listening area… Consequently, it suffers from a long history of unimpressive demonstrations [1], and the trade as a whole has pretty much shelved Ambisonics as a historical dead-end."*
  - Nettingsmeier, 2010, **Higher order Ambisonics - a future-proof 3D audio technique**

Now, the problem with this system is that it seems great, but it doesn't work. Gerzon knew at the time that there was an issue with his microphone, i.e. coincidence above 7 kHz. In Ambisonic Sound for Us Bamford and Vanderkooy did a direct comparison between ambisonics and 5.1, and says the same thing, ambisonics fails at high frequencies. Nettingsmeier outlines it perfectly when he says that the issues are with image stability and size of listening area. So what exactly is the problem in the higher frequencies?

Spatial Aliasing

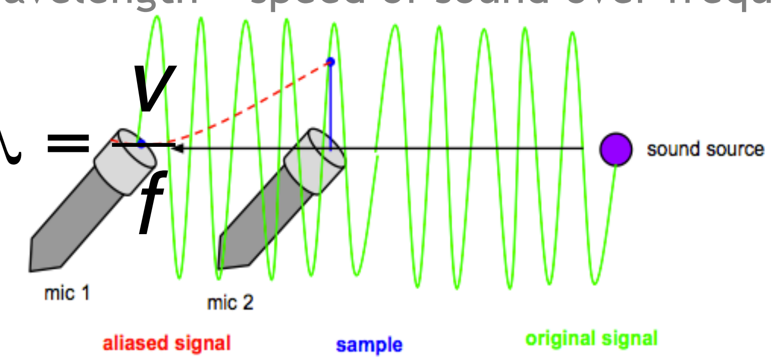
Core Audio’s Tetra Mic

The Spatial Sampling Theorem:

- The wavelength of the Nyquist frequency is 2 times the distance between capsules.
- Spatial Aliasing will occur at wavelengths shorter than the distance between capsules.

Wavelength = speed of sound over frequency

$$\lambda = \frac{v}{f}$$



Alias

20,000 Hz

The problem is what is called spatial aliasing, i.e the space between capsules causes aliasing. So, if we take a look at the drawing here, we see what aliasing actually is: the sound source is emitting a frequency of a particular wavelength, but because the wavelength is shorter than the distance between the two capsules, upon playback, a wavelength of a different frequency is heard. This is exactly what the spatial sampling theorem tells us, that spatial aliasing will occur at wavelengths shorter than the distance between 2 capsules. So, if we look at the Tetra Mic we see that the distance between two capsules is 33 mm, and if we apply that to the wavelength equation than we get a frequency of 9.7 kHz, which means that spatial aliasing is occuring 9.7 kHz and above in Core Audio’s Tetra Mic.

Project Goals

- To fix the high frequency aliasing problem.
- To gain a more accurate 3D soundfield
- To gain a more stable sound image and better localization cues in the high frequencies

Can't put the capsules at the same location in space and time

$\lambda = \frac{343}{40,000} = 8.58\ mm$

The goal of this project is to fix this high frequency spatial aliasing problem. By doing so it was hypothesized that we would gain a more accurate 3D soundfield with a more stable sound image and better localization. So, in order to do so, we first must ask, how close do the capsules need to be in order to avoid aliasing? Well, the highest frequency we want is the highest frequency we can hear, which is 20 kHz, so according to the spatial sampling theorem and the wavelength equation, the distance needs to be 8.58 mm between capsules. if we take into account the dimensions of the microphone fitting all the hardware inside the microphone's diaphragm, the capacitor and transformer, etc, even the smallest ones take up more space than that, which means that in order for this to work we really need to be placing the microphones right on top of each other which is essentially impossible. Unless we have a smaller microphone we cannot place 2 microphones in the same place at the same time.

Goals

A New Technique for Capturing True Coincidence Virtually

We can put them in the same location in space at different times

- Impulse responses to capture the acoustic characteristics of a room one orientation at a time

Benefits

- Convolution reverb plugin unit
- Easy to do IR measurement process
- Avoid gain mismatch
- Fix the spatial aliasing problem.

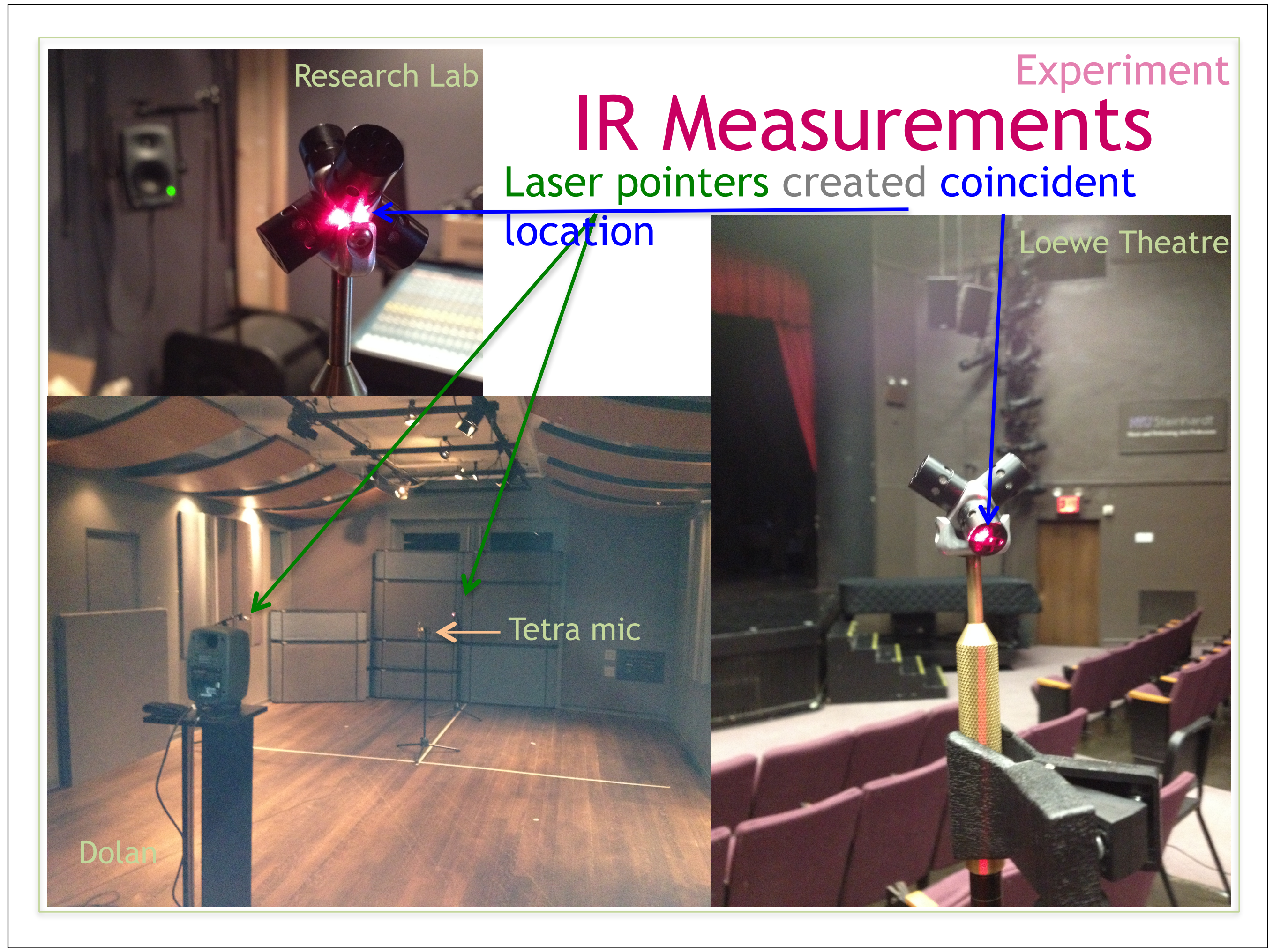
We can however, put them in the same location at different times. And this is the technique proposed in this thesis. By using impulse responses, such as a balloon pop, gun shot, a sine sweep, we can capture the acoustic characteristics of a room, one orientation at a time and then add them up in post.

So what would the benefits to such as system be? It is true that we cannot record live music this way, however, Ambisonic microphones are not used in the studio anyway. An adage that audio engineers generally stick to when considering their budget and requirements, is that they will go with the product that fulfills their needs in the most cost-effective manner. Simply put quality over quantity. Producers would rather by 1 high quality cardioid, than 1 ambisonic mic with 4 medium quality capsules for the same amount. So with 1 high quality cardioid microphone, and a high quality impulse response signal, and a room with good acoustics, we can create a professional grade convolution reverb with it.

Also, with this technique, we could skip the matrixing from A format to B format, and do just B format angles, which are much easier to find than A format angles.

By using the same capsule repeatedly instead of 4 different ones, we can avoid gain mismatch caused by the different hardware in the individual capsules of the ambisonic microphone.

Oh and yes, we can fix the spatial aliasing problem.

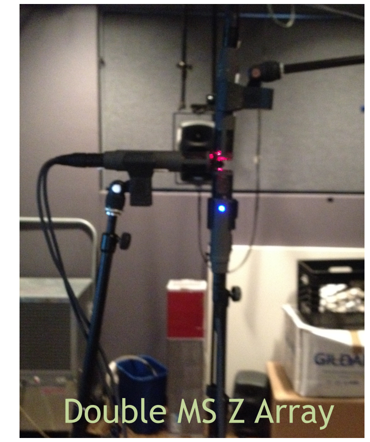


The IR measurement sessions were based on the most recent work done in the field of surround sound IR measurement, Farina and Ben Hardor primarily. Multiple 15 second sine sweeps were captured and averaged in a technique established by Farina for capturing high quality multichannel impulse responses. 3 rooms were measured, a small semi-anechoic listening room, a medium sized live recording room, and a large concert hall. The coincident point was created using laser pointers.

Mics and Configurations

Experiment

- Tetra Mic (A format)
- Tetra Mic Capsule at Coincident Location (A format)
- DPA 4011A Cardioid at Tetra Orientations at Coincident Location (A format)
- Double MS Z Array (B format)
- Sennheiser dual capsule Twin at Double MS Z Orientations at Coincident Location (B format)

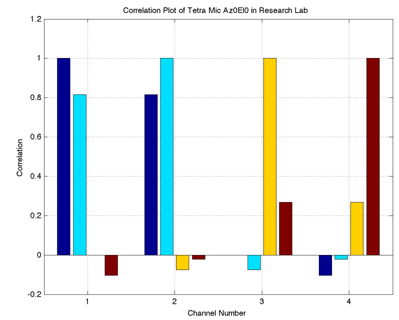


My goal is to see my process is better than or just as good as theirs. So it was important to do a 1 to 1 comparison against the actual spaced microphone array itself. So for A format, I captured impulse responses for the Tetra Mic, a tetra mic capsule at a coincident location, and for a high quality A format, the DPA 4011A cardioid microphone. For the B format angles Paul Geluso's Double MS Z array was used. As we see here the Double MS Z array uses 3 dual cardioid sennheiser twin microphones. The dual cardioids can be added together for an omni signal or the back channel can be phase inverted for a figure of 8, thus giving us the B format transmission channels easily. So for the B format, I used the Double MS Z array and then the senheisser twins at a coincident location as we see here.

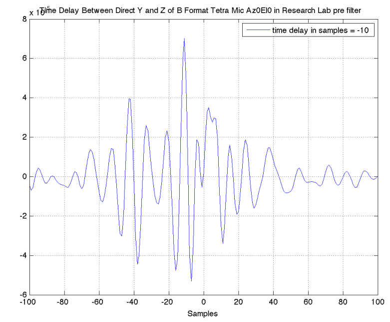
Analysis Methodology

- Signals were all matrixed to B format and analyzed on 3 factors; correlation, cross correlation and spectral content.

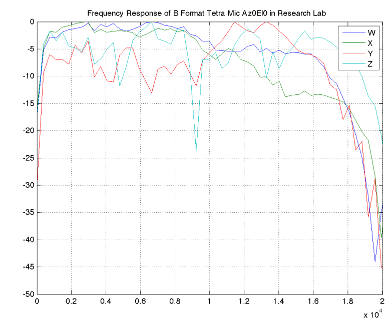
Correlation



Cross Correlation



Frequency Spectra



Definitions:

- Correlation: the degree to which signals vary over time.
- Cross Correlation: a measurement of correlation as a function of a time lag.
- Frequency Spectral Content: microphone’s response at all audible frequencies.

Goals:

- Keep correlation low in the reflections and reverberation of the signal.
- Remove the time lag due to capsule distance
- Have a flat frequency response throughout the 360 degree soundfield

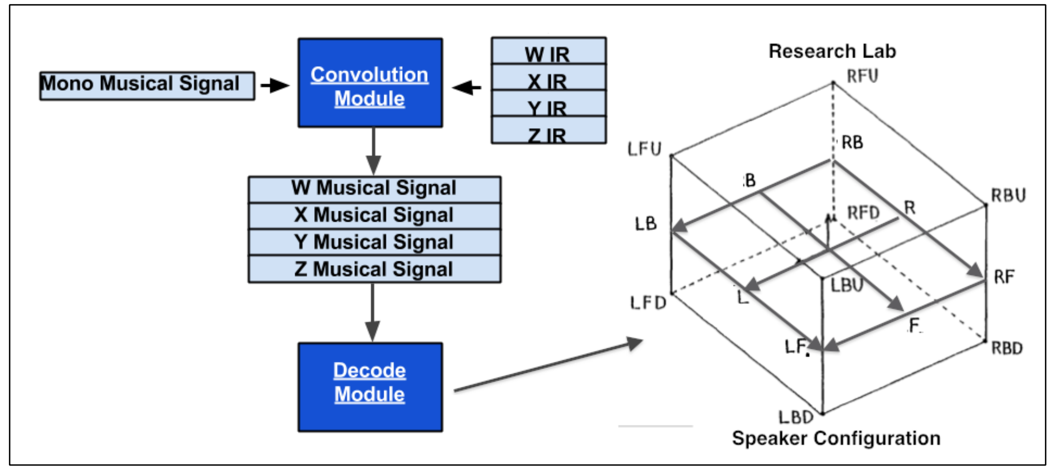
The signals were all matrixed to B format, and then analyzed on 3 factors correlation, cross correlation, and frequency spectral content. Correlation is the degree to which signals vary over time, it is content dependent. In these cases, the content is the impulse responses reflections off of different parts of the room, so the correlation will be low. Correlation will be high in 2 signals that are very similar, complete correlation being 2 of the exact same signals. Cross correlation is the measurement of correlation as a function of time, it will tell you how many samples apart the greatest correlation is. And the frequency spectral content will tell us how the microphone responds at all the frequencies it was built to capture, all audible frequencies. The goal in each case are as follows, for this new technique, I hoped to keep correlation low in the signal’s reflections and reverberations, while removing the time lag in the cross correlation plots, caused by the distance between capsules. And I wanted to see a uniform frequency response throughout the 4 channels, to create a smooth full sphere soundfield.

Analysis Results

- Lowest correlation: Sennheiser dual capsule Twin at Double MS Z Orientations at Coincident Location (B format)
- Smallest lag and highest correlation: DPA 4011A Cardioid at Tetra Orientations at Coincident Location (A format).
- Smoothest surround frequency response: DPA 4011A Cardioid at Tetra Orientations at Coincident Location (A format).

In analyzing the results, it was found that the configuration with the lowest correlation in the reverb and reflections is the Sennheiser dual capsule Twin at Double MS Z Orientations at Coincident Location (B format). The configuration with the smallest lag and highest correlation is the DPA 4011A Cardioid at Tetra Orientations at Coincident Location (A format). The configuration with the smoothest surround frequency response is the DPA 4011A Cardioid at Tetra Orientations at Coincident Location (A format). So at this point, we can do a short listening session of these. ….

Professional Listening Sessions Setup



AB Comparisons of 4 scenarios:

- Tetra Mic vs Coincident Tetra Capsule
- Tetra Mic with software filtering vs Coincident Tetra Capsules
- Double MS Z Array vs Coincident Sennheiser Twin (B format)
- Coincident Sennheiser Twin (B format) vs Coincident DPA Cardioids (A format)

Ratings:

- Sense of space
- Emersion
- Digital distortions
- Timbre quality
- Overall preference

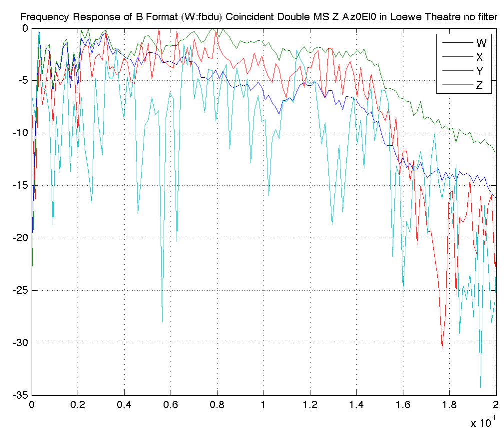
Next, listening tests were done to see if these differences translated to a better listening experience. 4 professionals in the field of music production and audio participated. It needs to be professionals because the differences are slight, they were chosen for their experience in noticing tiny timbre differences. The signal used was a solo piano, a broadband music signal with both tonal and transient qualities. The mono music signal was convolved with the impulse responses from the Dolan measurement sessions, and then the convolved signals were decoded to the following speaker array in the Research Lab. Height quad, low quad, ear level octagon. This was a bit of an unconventional ambisonic array, but it was done in order to get all types of signals, height, sides, mid, etc.

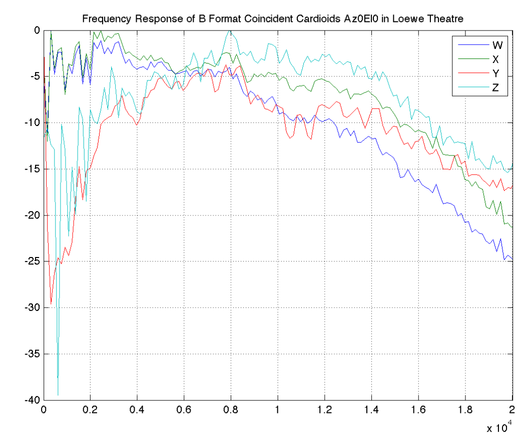
The actual test itself consisted of 4 AB comparison. List….

The listeners were asked to discuss and rate the spaciousness, envelopment, test for distortions, test for timbre quality, test for overall preference.

Results from Listening Sessions

- On Emersion: naturalness was noted in the coincident location configurations, however, phasing issues were noticed more regularly.
- On Timbre Quality: Sennheiser Twin at coincident location (B format) was preferred over the DPA 4011As at a coincident location (A format).
- On Spaciousness: the space between the capsules provided a better sense of space and a stronger spatial image of a sound source in a room.





Even though the goals I set out to achieve were achieved, this did not translate into a universal better listening experience for music. Across the board in the 5 ratings categories no consensus was reached. This does however mean that the new technique in a general sense preformed as well as the spaced microphone array. However, there are some points that we can take away from the listening sessions. On Emersion, read point 1, that was unforeseen. It was thought that the coincident point would not cause phase issues, however, because they were unique impulse responses causing different air pressure fluctuations it is thought that this caused phasing issues in playback. On Timbre Quality, read point 2, that was also unforeseen, it was thought that because the frequency responses of the DPAs were more consistent, that this would translate to a better listening experience. But the direct B format angles were more preferable. This makes sense since the A format microphone captures off axis sound by its very nature, since no capsule are supposed to be facing directly toward the sound source. On the third point, on sense of space, read point 3 and click, this can be seen directly in the cross correlation plots, where the distance between the capsules directly translates to the time lag in the correlation plot, which causes an early medial reflection in the signal causing a greater sense of space. The same phenomenon was also recognized to a lesser extent in the A format Tetra spaced array. Although in the A format spaced array it did not create a strong early lateral or medial reflection, but a bunch of weak ones seen as side lobes around the main maximum index of cross correlation.

Conclusions

- The new technique succeeds at creating a natural 3D soundfield.
- Space between capsules causes a greater sense of space
- Spatial Aliasing is unnoticeable in broadband musical signal

The new technique succeeds at creating a natural 3D soundfield. However, preference on spaciousness usually went to the spaced arrays because of the early lateral and medial reflections causing a better sense of being in the room with the instrument where the instrument is located somewhere. This is a point that is not discussed or written about in Ambisonics but proves to be very important. And finally, a more unsatisfying result is that spatial aliasing although discussed at length as being a huge problem with ambisonics is practically unnoticeable in a broadband signal.

Conclusions

Final Thoughts

- Improving the Technique -> Inverse filtering to fix the phasing issues.
- Modifying the Technique -> Capsule distance creates sense of space. Optimal distance instead of coincident point.
- Contributions to the field -> potential convolution reverb plugin. Most natural 3D soundfield.

On Improving the technique, inverse filtering can be used to fix the phasing issues. No filtering was done on purpose because it was thought that there would be no phase issues since the capsules are at a coincident point. However, it seems that the slight variations in the air pressure waves caused by using separate impulse responses did indeed cause phase mismatch. However, filtering is used all the time in ambisonics. The ambisonic decoding process should use shelf filters for instance. The Tetra Mic comes with a software called VVTetra to fix the phase issues it has with an inverse filter.

On Modifying the technique, the whole purpose of this thesis was to hear true coincidence. But, what was found is that capsule distance is actually important. If this is the case, then perhaps we should keep them apart. Just as you can choose different sampling rates to suite your needs, you can choose different capsule distances to suite your needs. There is an optimal distance for every frequency that is the wavelength of that frequency. We learned the optimal distance for 20 kHz early on, 8.58 mm, which at a 96 kHz sampling rate translates to 4.8 samples, or 2.205 samples apart at 44.1 kHz.

The value of work this field. -> I can see this being a convolution reverb plugin for high quality ambisonics once the phasing issues are worked out. But I think the real contribution is establishing the most natural 3D soundfield. I think this project went too far, in choosing actual coincidence over the optimal distance. However, there is a balance that can be struck and it is in that balance that we will find the most natural 3D soundfield.