Audio Engineering Society

Convention e-Brief 619

Presented at the 149th Convention
Online, 2020 October 27-30

Design and Application of a Native-D Recording Format for Optimal Dolby Atmos Reproduction

Aron Kobayashi Ritch[1]

[1] *New York University, Department of Music and Performing Arts Professions, 35 W. 4th St., New York, NY 10012*

Correspondence should be addressed to Aron Kobayashi Ritch (akr381@nyu.edu)

ABSTRACT

In this paper I will propose a Native-D format immersive microphone system made specifically for mixing and reproduction within the Dolby Atmos framework. While there are a number of immersive recording systems in use today, most have been designed for reproduction within listening environments that use sub 90° angled height channels such as Auro-3D. The microphone system proposed in this paper aims to take advantage of the native 90° height-to-main layer angle found in Atmos speaker systems by combining a non-coincident main layer with a corresponding near-coincident, 90° directional microphone for each height channel. This system excels in its high-fidelity capture of both group and individual sources, and benefits from a high level of decorrelation from channel to channel. An additional advantage of the system is its capture of Native-D format signal, which ensures that no format conversion or complex matrixing must be done, avoiding a loss of fidelity from recording to mixing stages. A case study of this system, consisting of multiple recording sessions, has been done in order to establish the validity of a system of this type, ultimately resulting in a final mix using Dolby Atmos. Both a "native immersive" and a "non-native" recording approach were taken in order to exemplify the system's versatility.

1 Introduction

Dating back to the early 30's and the advent of stereo recording, Alan Blumlein experimented with the use of multi-microphone techniques to capture audio in a way that felt natural to the human ears. His conception of a recorded stereo sound stage — derived from the amplitude-based panning of two directional microphones — led to a plethora of stereo microphone techniques that have dominated modern recordings since [1].

In short, any stereo microphone system — be it an XY, ORTF, AB, NOS, etc... — can be broken down into its usage of either coincidence (the placement of two capsules as close in space as possible) or non-coincidence (the placement of two capsules at a predetermined distance). The usage of these two principals stem from the auditory phenomena of IID (Interaural Intensity Difference) and ITD (Interaural Time Difference). Stereo/spatial information is thereby encoded onto a recording using either 1) intensity differences through the coincidental placement of directional capsules or 2) time delay through the non-coincidental placement of capsules. A third, liminal technique in which both intensity and time differences are utilized, is commonly referred to as a near-coincident system. Near-coincident systems such as ORTF and NOS, typically use directional capsules, but in closer distances (see Figure 1). All microphone techniques

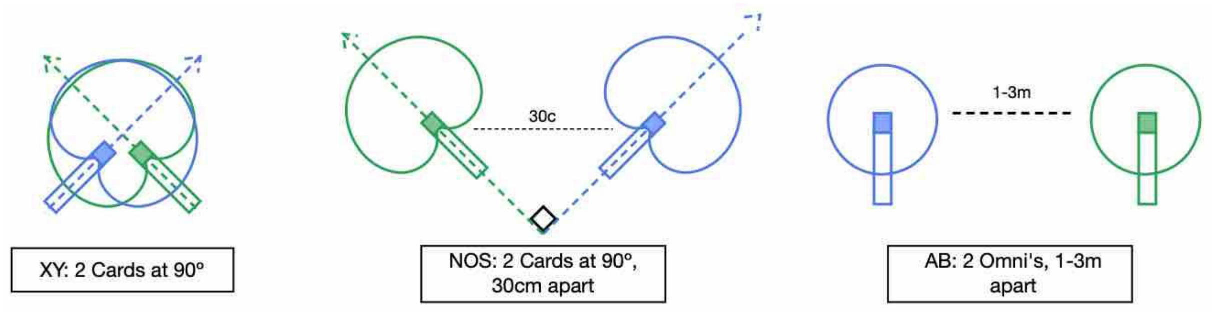

Figure 1. Popular Stereo Microphone Techniques

have their advantages and disadvantages, and a large part of the recording art and practicum is the recognition of which method is most appropriate for a given scenario. Many of these techniques have been extensively researched and thorough explanations can be found in textbooks and manuals such as *Tonmeister Technology: Recording Environments, Sound Sources, and Microphone Techniques*, which is largely considered to be the holy grail of acoustic recording texts [2].

While stereo recordings have dominated popular culture for the last 50 years, they still do not fully embody the complexity of what the human ears are capable of encoding and decoding. Though Blumlein's original intention of capturing a natural sound stage by mimicking the two-channel nature of our own ears was valid, the specific shape of the human ear (as well as the orientation of our bodies in relation to it) allow us to interpret far more than just stereo width and depth. Humans are constantly receiving and processing auditory information from all 360° of space around us, including height and the entire auditory field behind us (see Figure 2). Compared to stereo recording, the goal of 3D recording is to capture and reproduce a more holistic sound image, whether it be retroactively fabricated or natively recorded. As stated by Kimio Hamasaki, one of the most prominent multichannel researchers, the principal aim of recording multichannel audio is in "reproducing the spatial impression of a reverberant sound field such as a concert hall" [3].

When one intends to create a recording that is made for a 3D or immersive sound stage, all the principals of stereo recording apply, but the added ability to capture height and surround offers more circumstance for innovation.

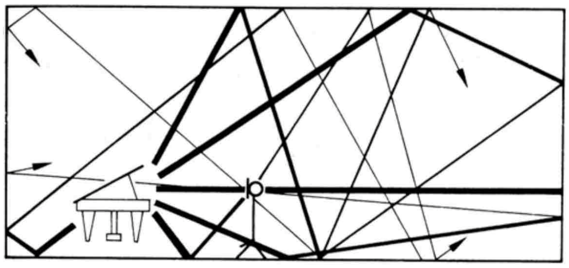

Figure 2. "Sound Propagation in an Enclosed Room" [2]

While there are plenty of previously established immersive microphone techniques, none yet have been designed with the Dolby Atmos mixing environment in mind. In this brief, I hope to fill that gap by designing and testing a new immersive microphone system designed to work in concert with the Dolby Atmos mixing environment.

2 Background

2.1 Desired Attributes for Immersive Recording: The Benefits of Decorrelated Audio Streams and Non-Coincident Systems

While there are many techniques used to create spatial recordings, they are not all created equal, nor do they all attempt to serve the same purpose. Generally, when capturing immersive ambience, one is looking to attain accurate representations of localization, depth, width, envelopment, and spaciousness. A crucial factor in the authenticity of a recorded sound stage is in the proper decorrelation of each recorded signal, meaning each microphone must capture sufficiently different information at each point in space in order to create a sense of spaciousness and externalization [4]. This concept of correlation is an inherent part of the design of microphone systems, and is inextricably linked to the methodical orientation and placement of capsules in space. In comparing different microphone system designs, it is often accepted that non-coincident stereo pairs — which have the highest amount of uncorrelated signal due to the physical distance between the capsules — have a much larger and more enveloping image compared to a coincident system, with near-coincident systems falling somewhere in between the two. While coincident systems excel in their accuracy and ability to capture a higher direct-to-reverberant ratio, when looking to capture ambiences or mixing with spot mics, non-coincident systems are preferred a majority of the time.

Several papers have been authored on the subject of uncorrelated signals and immersive recording scenarios. In the paper cited earlier, Hamasaki states that:

> In order to reproduce the spatial impression of a concert hall, it is necessary to have four uncorrelated feeds to the left, right, rear left and rear right loudspeakers. Therefore, a proper microphone technique that can catch four uncorrelated indirect sounds is needed for reproducing the reverberation of a concert hall [3].

While this notion holds true for any audio presented or captured in the horizontal plane, the vertical one does not work under equivalent pretenses. This distinction is explored by Christopher Gribben and Hyunkook Lee in their paper titled “A Comparison between Horizontal and Vertical Interchannel Decorrelation,” in which they examine the effects of decorrelation on the horizontal and vertical axis in terms of perceived image spread. Their testing concluded that, in general terms, “interchannel decorrelation had a significant effect on auditory image spread both horizontally and vertically, with spread increasing as correlation decreases” [5]. While this was the overarching conclusion, there were some variances when it came to which planes were being decorrelated. Primarily, they found that:

> The effect of vertical decorrelation was less effective than that of horizontal decorrelation. The results also suggest that the decorrelation effect was frequency-dependent; changes in horizontal image spread were more apparent in the high frequency band, whereas those in vertical image spread were in the low band [5].

This statement implies that, while decorrelation is incredibly important on a fundamental psychoacoustic level to create any sense of immersive space, its function is not uniformly optimal on every plane, and generally more cues must be present to enhance the sound stage. In another paper by Lee and Gribben titled “Effect of Microphone Layer Spacing for a 3D Microphone Array”, this idea is further explored:

> The primary purpose of the height microphones is to capture ambient sounds for height loudspeakers, whereas that of the main microphones is to localize the sound source image at the height of the main loudspeakers. In this regard, a direct sound component included in the height microphone signal can be regarded as a vertically introduced interchannel crosstalk [4].

Here, Lee and Gribben detail the primary function of height channel microphones and how the exclusion of inter-channel crosstalk — direct signals that are shared between capsules — is integral to their utility. Aggregating the research of both Hamasaki and Lee/Gribben suggests that: the most crucial factors in creating quality immersive recordings using a height layer are the exclusion (or reduction) of both correlated signal and inter-channel crosstalk.

2.2 Ambisonics

The term “ambisonics” was introduced by Michael Gerzon in the 1970’s to describe a method of sound reproduction in which audio could be played back in 360° space, through multiple speakers, while maintaining output agnosticism [6]. In ambisonic recording, sources are typically captured in what is called “A format,” which are audio streams taken directly from the microphone capsules. These are then converted into “B format,” which consists of 4 vector based audio streams in the X, Y, Z, and W planes that make up a complete spherical harmonic representation (see Figures 3 and 4) [7].

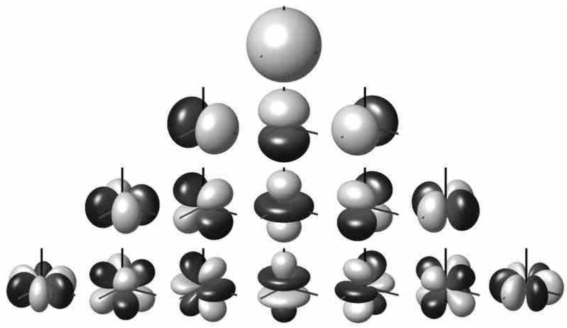

Figure 3. Spherical Harmonics [8]

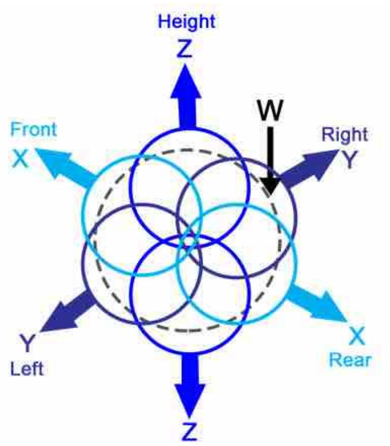

Figure 4. B Format Mapping [9]

Ambisonic recording is advantageous in its flexibility as a medium. Sources recorded using an ambisonic system require a low amount of capsules that can subsequently create a high amount of channels with capable reproduction — to some degree — on virtually every system (2.1, 5.1, 7.2.1, etc...) using the proper matrixing and format conversion techniques. While ambisonics excel in flexibility, their sonic quality can often fall short when subjected to extensive processing. In order to up-mix the original audio streams to span across the high-channel-count systems required for immersive playback, virtual capsules or “beams” must be synthesized. This process produces highly colored audio channels derived from the off-axis recordings of directional microphone capsules [7]. Ambisonic technology is also bound to the physical limitations of the microphones used as the summing and subtraction of signals requires the assumption that the capsules are entirely coincident. Therefore, ambisonic recording systems such as the tetrahedral arrays found in microphones like the Sennheiser AMBEO, can only capture entirely correlated signal.

2.3 Coincident Systems and Native-B Format Recordings

Native-B format recordings are those that utilize capsule orientation in a way that allows the engineer to circumvent the “A-format” completely and record the X, Y, Z, and W information directly. A primary example of such a system is one created by Paul Geluso and Kathleen "Ying- Ying" Zhang, titled the 3DCC microphone system. This array utilizes dual capsule technology popularized by microphones such as the Sennheiser MKH800 Twin to minimize the amount of physical space needed, and maximize the variability of the system. The array uses three of these microphones oriented in such a way that the X, Y, Z, and W information is recorded natively. This system excels in its ability to offer extensive flexibility in terms of scaling, while minimizing the channels needed in the recording stage and circumventing the initial “A format” [10]. However, like Native-A format system’s, its’ Native-B capture still restricts it to entirely coincidental pairings.

2.4 Non-Coincident Systems and Native-D Format Recordings

Native-D format recordings are “one-to-one” systems, meaning they require a capsule for every speaker and tie a direct line between the microphone's input and the speaker's output. While there is capacity for post-processing such as VBAP (Vector Based Amplitude Panning), this format does not require any format conversion or extensive matrixing to function properly. A primary example of such a system would be the "Hamasaki Square," a multichannel surround system developed by Kimio Hamasaki. This system consists of four bi-directional microphones, spaced in a square with 1-3m of distance on each side, and the positive lobes of the microphones facing out (see Figure 5). The microphones used actively attempt to null front and rear sources in order to accurately capture a wide representation of room ambience. The system borrows principals of an AB system with its 1-3m capsule distancing, but with the more directional nature of a figure-8 microphone. This system can also be easily adapted with the addition of a height layer to produce an immersive 3D image [3].

Other systems such as the Twin Square and Twin Cube, introduced by Gregor Zielinsky, use a similar non-coincident underpinning but with an added height layer, thereby creating a cube of microphones for immersive capture (see Figure 6) [11]. Because these systems are "one-to- one" and circumvent both "A" and "B Formats" entirely, they benefit from their ability to use non-coincident capsule placements that take advantage of large-scale decorrelation. This is compared to all non-Native-D format systems, which must utilize coincident capsule placements. These attributes also allow for the purest possible transferal of audio signal from microphone to speaker, mitigating the effects of spatial processing and format conversion on the quality of the sound as a whole. However, the systems one-mic-per-speaker approach means that they require a high channel count and posses a lack of scalability when compared to "A Format" and "B Format" systems.

2.5 Dolby Atmos

Dolby Atmos' mixing environment uses a vast array of speaker setups. However, its usage of a 90° height array is unique compared to other immersive speaker setups, like Auro-3D, which use a 45° array. AES convention briefs such as "The Design of ATMOS Post Production Theatres," [13] as well as the info provided by Dolby Atmos (see Figure 7) are critical for understanding the design and layout of Atmos mixing environments. Systems like Atmos allow for the flexible intertwinement of both audio beds —

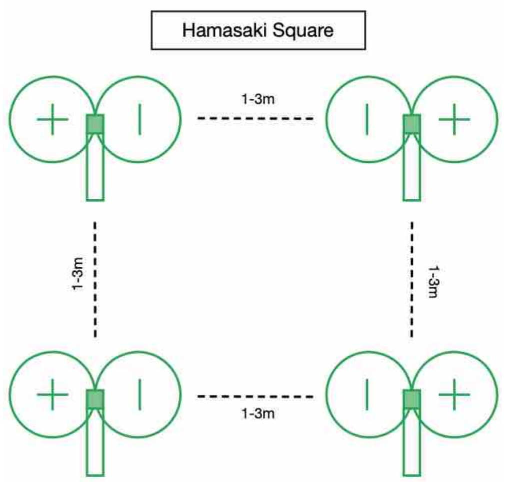

Figure 5. Diagram rendering of Hamasaki Square Microphone Technique

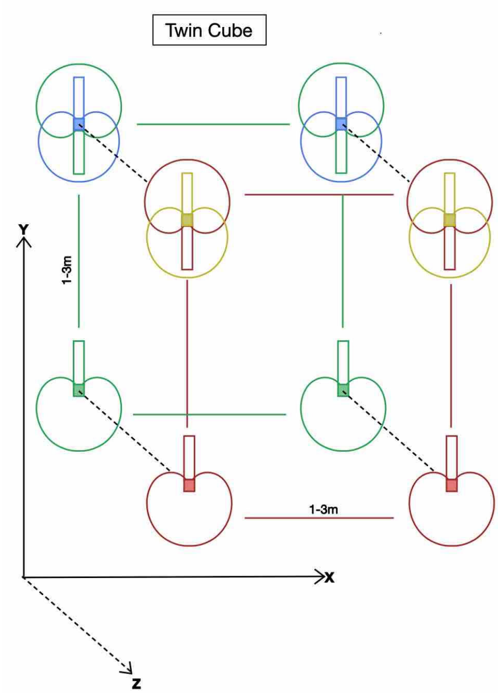

Figure 6. 3D bird's eye diagram rendering of the Twin Cube microphone technique

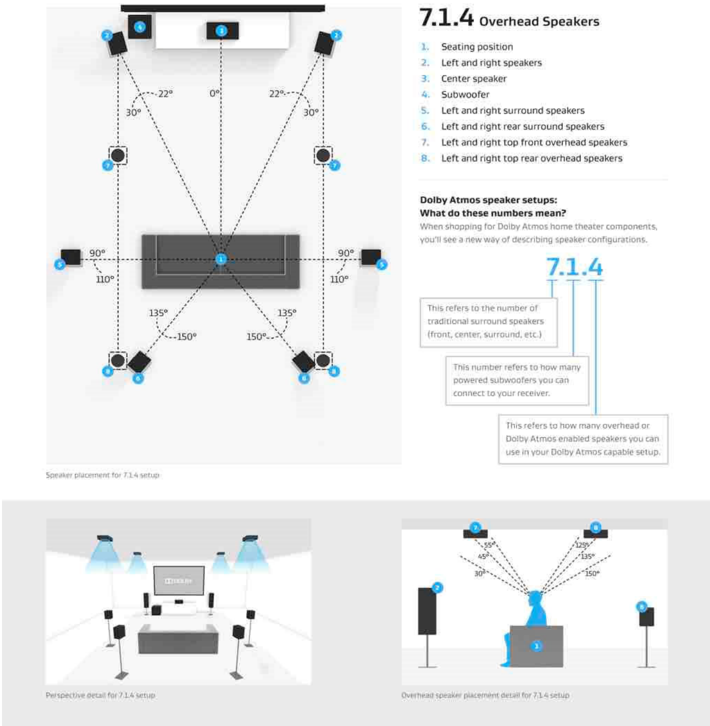

Figure 7. Rendering of a 7.1.4 Dolby Atmos speaker layout [12]

static audio streams akin to channels in stereo environments — and objects — audio streams whose placement in the 3D environment is determined on the user end by the predetermined vector metadata applied by the mixer [14]. In the realm of object based audio streams, mixes are scalable and highly intelligent, allowing for the high-quality transmission of immersive audio across many platforms.

3 Proposed Microphone System

Pooling knowledge from the research noted above as well as personal experience with immersive recording, I knew that I wanted to design a Native-D format system in order to avoid any complex matrixing and format conversion that could degrade the audios signal path from microphone to speaker. While this does highly complicate the recording setup, for the purposes of trying to create an *optimal* system for Dolby Atmos, this seemed to be the best approach. I also knew that I wanted the system to have maximum decorrelation for proper room ambience capture, while maintaining minimal inter-channel crosstalk for good height layer capture. Lastly, I knew that in order to optimize the system for the Atmos environment, I would need to mirror its' 90° height channels in my height microphone system, thereby reducing the valley between capture and reproduction as much as possible.

3.1 System Design

The actual design of the microphone system ultimately came down to the primary establishment of the listening environment. I decided to work within a standard 7.1.4 system consisting of L, C, R, Lss, Rss, Ls, Rs, LH, RH, LHs, and RHs channels (see Figure 7). The main layer of the proposed system mirrors an AB format, consisting of six capsules spaced 1m from one another. While traditional AB systems utilize omni-directional capsules, the null points on directional capsules can be incredibly useful in minimizing inter-channel crosstalk. Therefore, four cardioid capsules were used for the L, R, Ls, and Rs channels, with the front system's null points facing the rear of the room, and the surround system's null points facing the front of the room. Two bi-directional capsules were used for the Lss and Rss channels, with the positive lobes of the capsules facing out and the null points facing the front and rear of the room (see Figure 8).

While the employment of directional capsules helps to minimize inter-channel crosstalk, distancing the capsules maximizes decorrelation. Since these two methods can often be present in opposing circumstances, I chose a near-coincident system for my height layer that allowed me to take advantage of both. The height layer of my system took advantage of the 90° height angle by utilizing a pre-existing NOS microphone array. NOS arrays use 2 cardioids at 90° with 30cm of space between the capsules, thereby creating a near-coincident stereo field between the height layer and the main layer of my system (See Figure 9). This combination of near and non-coincident systems utilizing directional microphones provided several advantages, especially in the context of a studio recording. Overall, it minimized the size of the system as much as possible while prioritizing the ambient capture. Every microphone was oriented so that very distinct areas of the room would be captured. The front main layer only captured front signal, while rejecting rear and side reflections. The sides and surrounds only captured lateral reflection points while rejecting direct signal. The height layer was oriented to capture first reflections points from the ceiling while minimizing pickup of source material and unwanted floor reflections. The consolidation of these decorrelated ambient slices results in a holistic ambient image with minimal phasing or inter-channel crosstalk (see Figure 10).

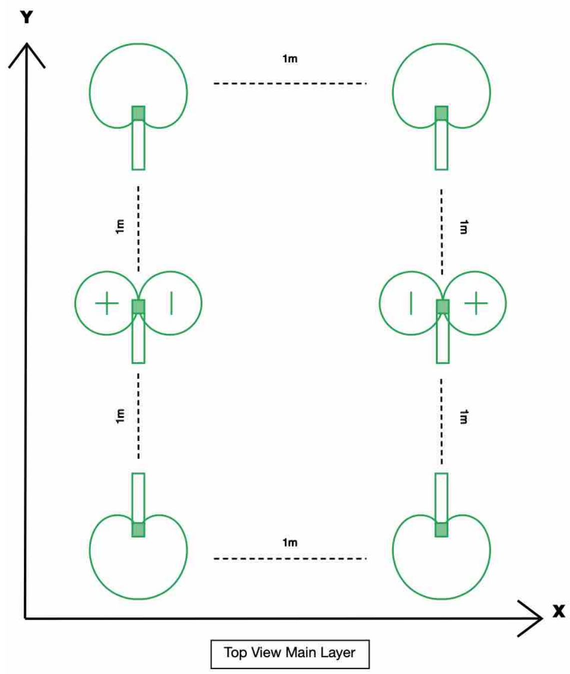

Figure 8. 2D bird's eye view of system's main layer

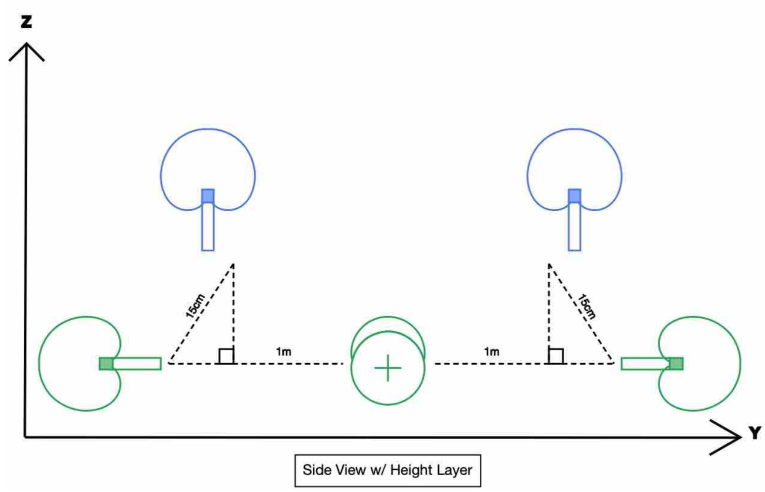

Figure 9. 2D profile view of system with main layer and height layer

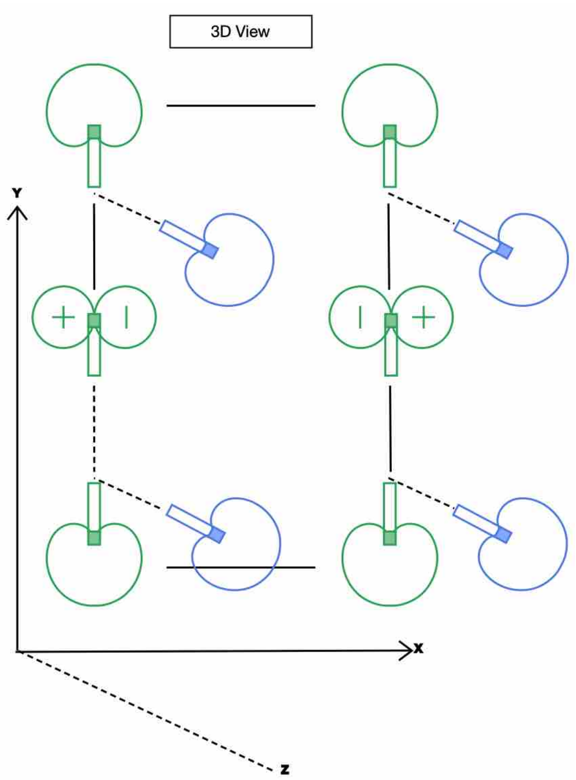

Figure 10. 3D view of system with main layer and height laye

3.2 Ensemble Case Study

In order to test the validity of the proposed microphone system, two distinct case studies were devised. The first case study sought to apply the system to a more traditional chamber/classical setting in which the room itself was to be captured along with the ensemble of instruments. The second sought to use the system in a more pop oriented setting, in which overdubs and creative post-production techniques were to be employed.

In the ensemble case study, the system was setup alongside a Neumann KU100 binaural head, and placed in front of a piano and cello duo (see figure 11). The binaural head acted as a control, allowing me to compare its' pseudo-natural encoding of spatial cues with my systems fragmented capture method. The recording apparatus was set up in the middle of the live room with the binaural head level with the main layer and in-line with the front of my system. For the L, R, C, Ls, and Rs microphones, Sennheiser MKH800's set to cardioid were used. For the LH, RH, LHs and RHs microphones, Schoeps MK6's set to cardioid were used. For the Lss and Rss microphones, Coles 4038's were used due to their natural bi-directional polar pattern that would allow for the alignment of the positive lobes to face outward and the nulls of both mics to face directly towards the front and back of the room (see Figures 12 and 13).

Figure 11. Image of ensemble case study subjects (Rohan Chander and Dorothy Carlos)

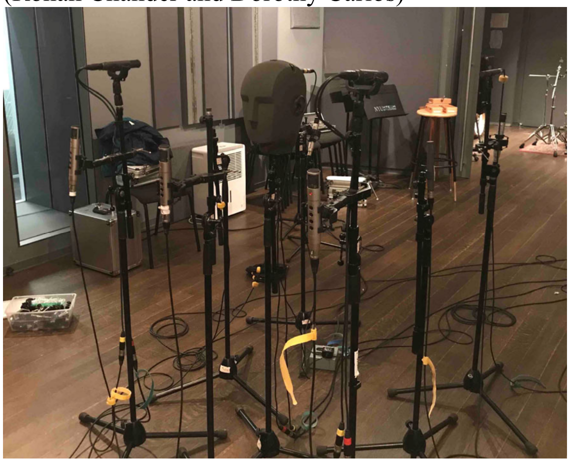

Figure 12. Image of microphone system and Neumann KU100 binaural head

Figure 13. Image exemplifying the main layer (red) to height layer (blue) relationship

Once the recording was complete and the signals were routed to the proper speaker setup, the end result was a highly satisfactory and comparable 3D image when compared to the binaural one in respects to the recordings immersiveness, depth, spaciousness, and envelopment.

3.3 Multi-track "Pop" Music Case Study

In order to apply my system to a more pop oriented recording scenario, I decided to investigate both a "native" and "non-native" spatialization process. In the context of this case study: "native" spatialization meant the instrument was being performed live in the room with the system capturing its natural placement within the 3D field, whereas "non-native" meant that the instrument was pre-recorded as a mono source and spatialized by way of re-amping. Re-amping was done by placing speakers around the immersive system in order to perform a manual pseudo-spatialization of the audio. To apply these methods, I decided to take the multitrack of a song recorded originally in stereo, and re-record elements of it using the microphone technique. The "native" recording was done by entirely re-recording the drums live using a combination of traditional spot mics and the immersive system as a room capture.

3.4 "Native" Immersive Recording

The drums were set up in the live room of the studio on the far end of the wall. I used a fairly traditional grouping of drum spot mics consisting of overheads and close mics for the snare, kick, and toms. The immersive ambient system was placed in the middle of the room, six feet from the front of the kick and was set up nearly identical to the one in the ensemble case study (see Figures 14 and 15).

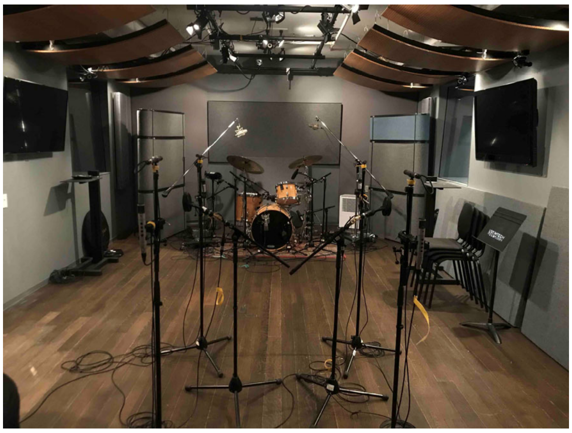

Figure 14. Image of "natively" spatialized drum recording session combining both close mics and my system to capture 3D ambience

Figure 15. Image featuring close up of drum spot mics

3.5 "Non-Native" Immersive Recording Through Manual Spatialization

Spatializing the mono signals by re-amping them into the live room fostered a large degree of control in terms of source placement. Physically moving the speakers around the microphone system effected not only the source's location around and above the 360° sound stage, but also the perceived distance of the source from the listener. In order to keep the noise floor low however, I kept the amount of spatialized layers to a minimum while maintaining one speaker per every source. Prior to recording, desired instrument placement within the sound field had to be corresponded and mapped to the speakers' physical placement in relation to the microphone system.

Through this manual panning technique, I was able to create a whole immersive sound scene using mono sources from a typical stereo production. Not only could sources such as vocals, guitars, keyboards, and bass, all be placed naturally into a 3D environment, but they could also be moved dynamically throughout the song. For example, in order to enlarge the sound stage of a certain section such as the chorus, I would simply move the speakers in the room to create the desired effect. The process was applied to instruments such as vocals and guitars where the change between verse and chorus really allowed for a creative expansion of the sonic space (see Figures 16 and 17). The ability to place instruments above the listeners' head by mounting the speakers on stands was also a highly useful creative technique.

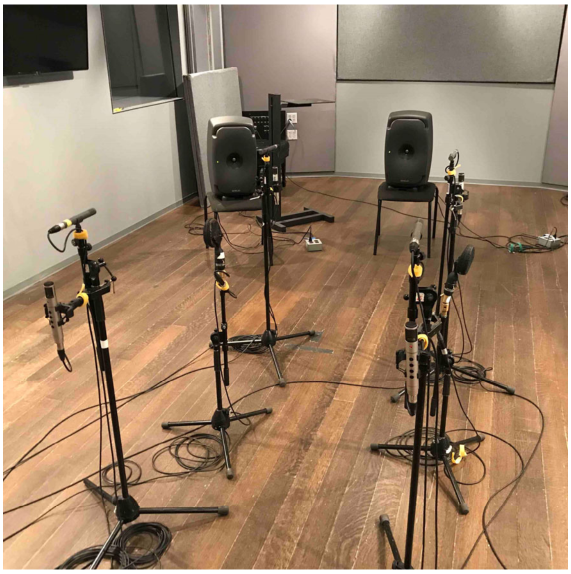

Figure 16. Image from "non-native" spatialization session featuring panning of the rhythm (L speaker) and lead guitar (R speaker)

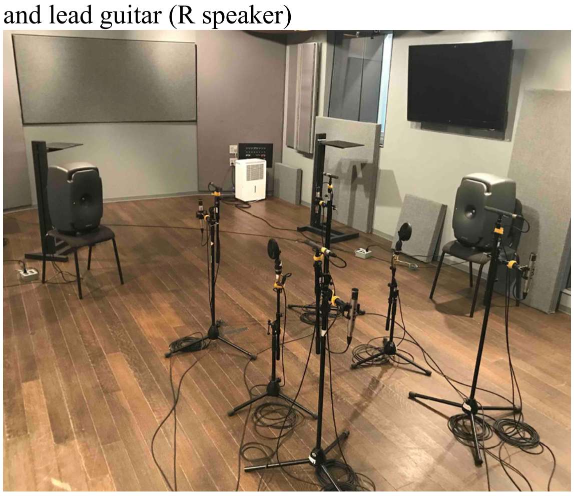

Figure 17. Image from "non-native" spatialization session featuring panning of the rhythm guitar (R speaker) and lead guitar (L speaker)

3.6 Mixing

The intention behind the mix of the pop session was to create a realistic immersive environment in which the listener felt more enveloped within the sound stage of the music. The bed of the song still followed many of the principals established in stereo recording for the past 50 years — kick, snare, and bass were kept in the middle, cymbals and toms panned stereo, etc... — but with the added depth of height and surround cues. Using the non-native immersive captures, I could imbue space and decorrelated room signal on top of the original mono recording so that some version of every instrument was in every channel.

3.6.1 Matrixing

The first step of the mixing process involved the matrixing of each channel within the Dolby Atmos mixing environment. The Dolby Atmos Render was used in conjunction with ProTools Ultimate in send/return mode. In this mode, I utilized a 7.1.2 bed in conjunction with a supplemental stereo bed for the surround height channels as well as objects. The 7.1.2 bed, as well as the stereo surround height bed, were routed directly to their respective speaker channels.

The "natively" recorded drums were mixed so that the spot mics were kept within their normal routing in a stereo environment. The stereo overheads were panned hard L and R. The kick, snare, and supplemental spot mics were all panned to the center channel. The only mics that were panned slightly wider were the toms, which were placed somewhere between the front and side-surround channels for a nice image pop. The immersive room system was then panned so that each microphone was matrixed to its respective channel.

The "non-natively" recorded instruments such as the bass were oriented so that the original direct signal was panned up the middle, while the ambient system (similar to the matrixing of the drums) was matrixed so that each microphone had a respective channel. This matrix layout was rinsed and repeated for all individual instruments that were also spatialized "non-natively." Pan automations of specific sources were made to mirror the manual movement of the speakers in the "non-native" spatialization session, keeping the widening effect during certain sections consistent across the 3D image (see figures 18 and 19). Instruments that were spatialized above the listener were panned respectively and combined with the ambient system (see figure 20), and sound sources that required automation were assigned as objects in the Atmos renderer (see figure 21).

Not every channel of the room system was used on every instrument. In cases where a mic and a direct signal were in identical placements within the 3D matrix, the redundant ambient channels would only produce phasing problems when paired with the original signal in mixing. It was in these scenarios that I "replaced" certain microphones entirely with their original signal during the mixing stage. For example, the vocals needed a prominent and steady frontal image that came from the original signals, thereby making the immersive systems L and R mics (which sat directly in front of the speakers) obsolete. By replacing the L and R channels of the immersive system with the original signals, but incorporating the rest of the ambient microphone channels, the system still functioned as intended in terms of its physical spatialization of the pre-recorded signal in the room, but without any phasing problems.

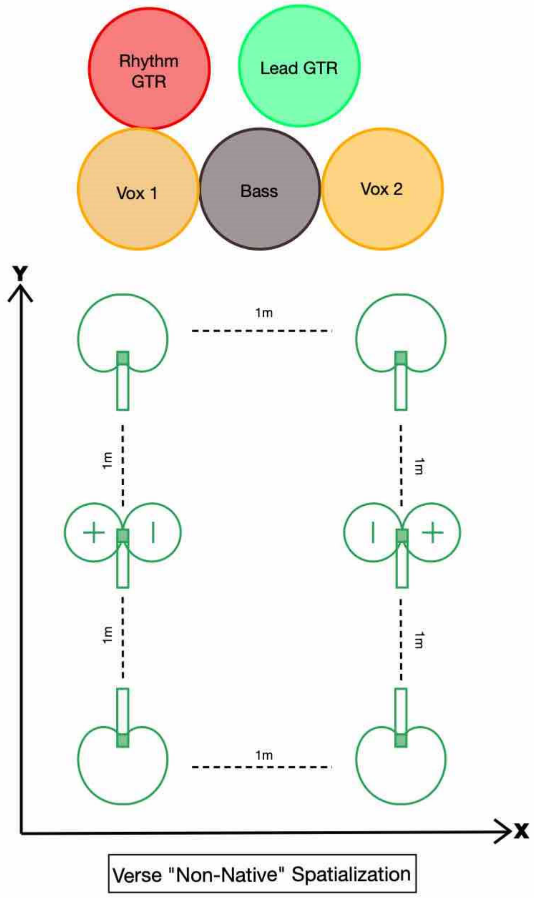

Figure 18. 2D birds-eye-view rendering of main layer immersive panning during the verses

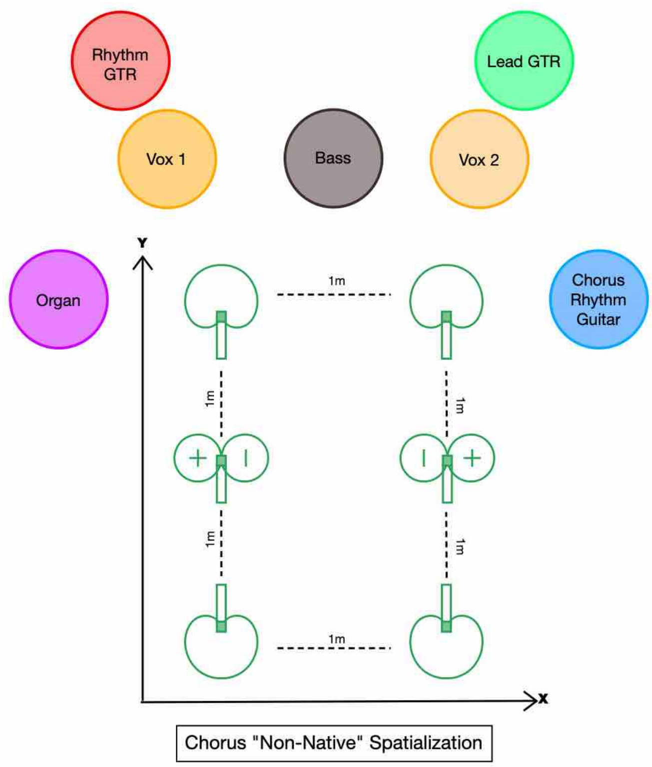

Figure 19. 2D birds-eye-view rendering of main layer immersive panning during the choruses

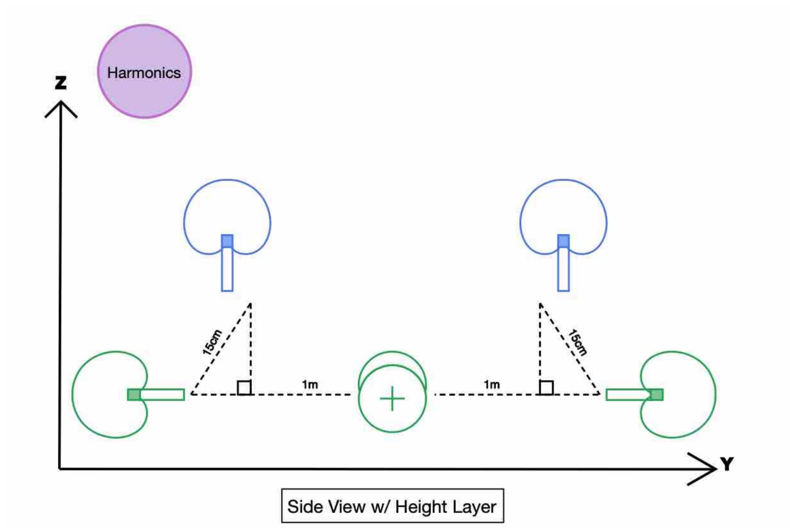

Figure 20. 2D side-view rendering of main and height layer immersive panning during the choruses

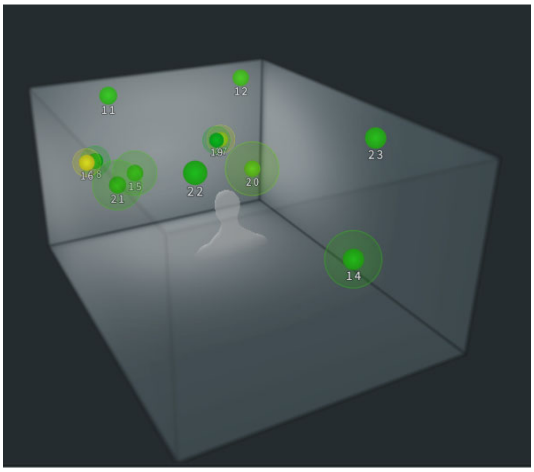

Figure 21. Image of objects in the Atmos Renderer

3.6.2 Processing

While Dolby Atmos is an extremely flexible system when it comes to production and signal processing, like any immersive environment it is still a certain degree more complicated to apply traditional mixing techniques. Tools such as compression and delay-based effects become much more cumbersome when one has 12 channels to contend with as opposed to two — especially when there are very few plugins that work beyond two-channel stereo without being run as native multi-mono. That being said, over the course of the mixing process, I developed several techniques for processing tracks so that I could come within a certain degree of how I would want to process a stereo mix.

The biggest goal in mixing a project like this is the externalization of the immersive sound stage. While my microphone system did a favorable job creating a sense of space natively within the recording, that sense could always be enhanced and highlighted for great effect in post. An effective asset for ambience shaping in mixing is compression, in which the initial transient can be manipulated and the decay and room ambience can be heightened. In order to compress the immersive ambient mics, I decided to work in pairs by creating stereo busses for each LR pair, i.e. an Ls/Rs buss or an LHs/RHs buss. These busses allowed me to treat each microphone pair as a stereo unit before I sent it from ProTools to the Atmos rendering system. I used several plugins in order to expand the space of the live room. On drums, I found that aggressive compressor plugins were useful for crushing a signal by using a fast attack (around 15ms) and a fast release (around 15ms), and then blending the wet and dry signals using a mix knob. This allowed the room to pump and the natural decay time to grow while not losing control of the original space. I used this technique on the sides and surrounds to lengthen the room and provide more space behind the listener to contrast the front. This approach was furthered through the use of reverb. I found that by using artificial reverberators in parallel with the original signal, I could gain more depth through the addition of decorrelation, while also widening my surround and height channels. Keeping the reverb at about 40% wet with a relatively quick decay (around 0.5s), and some pre-delay (around 18ms), allowed the perception of the room within the sound stage to be elongated and rectangular in shape rather than square. I used similar compression techniques for other instruments such as the bass and guitars, in which the rooms were compressed using a less severe compressor plugin (high ratio with slow attack and fast release) in order to enhance the ambience and increase the rooms audible decay time.

Other highly effective assets for externalizing and enhancing the ambience of the recorded tracks were stereo width and image manipulation. Using a combination of micro pitch shift plugins and imagers, I was able to enhance the width and decorrelation factor of the paired signals. For example, on the bass room heights I wanted the signal to feel more enveloping and wide without being muddied by the elongated decay time of a reverb, so I added a micro pitch shift effect to enhance the decorrelation factor. These processes were crucial for elements like the bass where I wanted the low end to feel more present and in front of the mix, but the mids and highs to feel diffuse and engrossing. The imager was useful for surround microphones, such as those on the guitars, as it widened the sound stage and created a deeper and more realistic image.

Another type of tool that was crucial in creating the immersive mix was EQ and spectral manipulation. Proper EQ of different microphones within an immersive system is crucial when there are so many signals to contend with. Cutting competing frequency ranges, especially between the spot mics and the ambient system, allowed the low end to breathe and avoid phase cancellations. Overall, I found that hi-pass filters were on almost every channel of the immersive system. For example, cutting frequencies below about 120Hz on the main layer drum mics and below 150-200Hz on the height layer, allowed the low end of the spot mics to remain uncluttered by phasing problems caused by the time delay of the room system. The height system was also more easily detached from the main-layer with the subtraction of low end, thereby mitigating any additional inter-channel crosstalk. I applied this same technique to the bass and guitars as well. Utilizing spectral shapers like distortion was also a crucial step in fitting together so many signals at once. By adding harmonic complexity to individual signals, I was able to shift the spectral weight of tracks into different frequency ranges. For example, on the bass tracks, I found that the combination of compression, EQ, image manipulation, and distortion gave the bass' room sound definition and stability, allowing the low end to emanate from the spot mics more effectively, and for the overall image to feel more enveloping as a whole.

4 Analysis and Conclusions

In the proposition of a new Dolby Atmos-specific immersive recording technique, it was imperative that I weigh its capabilities against other systems that are already in use. All microphone systems have their innate utility within certain contextual boundaries, and are designed for use within specific recording scenarios. The design of this system was centered around the idea that Native-D format recording systems with maximum decorrelation and minimal inter-channel crosstalk posed the most ideal circumstances for a quality immersive recording, in which ambient sound fields were the objective. This project was centered around capturing the most optimum recorded signal, even at the sacrifice of portability and simplicity. And, while systems like this have been used in the past for various 3D playbacks, no microphone system has been developed to work in concert with the Dolby Atmos framework.

Overall, the mix incorporating this system produced a sound stage that was immersive and external, with a high degree of flexibility when it came to the processing and manipulation of tracks with the most minimal degradation of sound. While the high channel count made the system complicated in its setup and calibration, the lack of any up-mixing or format conversion made it streamlined in its mixing. The layout of the system also allowed it to excel in the context of studio live rooms, a difficult feat when compared to the large ambiences of concert halls. The system's use of directional capsules in specific orientations that minimized problems like phase and inter-channel overlap, allowed for success in a smaller room while also logically being adaptable to bigger environments. Furthermore, the usage of this system under both "native" and "non-native" immersive recording situations, exemplified the systems adaptability and utility in a multitude of circumstances. The ability to naturally spatialize a mono signal into an immersive environment by capturing natively decorrelated room ambience can be a highly useful technique in any 3D recording scenario. The combination of traditional recording techniques and the immersive system in this case study, displayed how a system like this could be used to capture instruments for a pop music scenario.

5 Future Work

Now that the system has been designed and tested using a case study, more objective testing can be performed in order to further contextualize my system within the working field. Applying my technique to other environments, sources, and situations could provide further metrics on its abilities and areas of improvement. One test that I would find highly useful would be a side-by- side comparison study between this system and "Native-A" or "B" format coincident systems in both a highly reverberant concert hall and minimally reverberant studio environment. This would provide data on the system's utility and capabilities against systems with entirely different inherent qualities.

References

[1] Alexander, Robert Charles. *The Inventor of Stereo: the Life and Works of Alan Dower Blumlein* (Massachusetts: Focal Press, 1999).

[2] Dickretter, Michael. *Tonmeister Technology: Recording Environments, Sound Sources, and Microphone Techniques* (Temmer Enterprises, 1989)

[3] Hamasaki, Kimio and Hiyama, Koichiro. "Reproducing Spatial Impressions With Multichannel Audio", paper presented at the *AES 24th International Conference on Multichannel Audio*, Banff, June 26-28, 2003. http://www.aes.org/e-lib/browse.cfm?elib=13880

[4] Lee, Hyunkook and Gribben, Christopher. "Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array", *Journal of the Audio Engineering Society* Volume 62*,* Issue No. 12 (2014), pp. 870-884. https://doi.org/10.17743/jaes.2014.0045

[5] Lee, Hyunkook and Gribben, Christopher. "A Comparison Between Horizontal and Vertical Interchannel Decorrelation", *MDPI Applied Sciences,* (2017). https://doi.org/10.3390/app7111202

[6] Nishiyama, Tomoyuki and Nagata, Takanobu and Ogata, Shinichirou and Hasue, Tetsuya and Kashiwagi, Mina. "Recording and Mixing Techniques for Ambisonic Sound Production", paper presented at the *AES International Conference on Spatial Reproduction - Aesthetics and Science*, Tokyo, Japan, August 6–9, 2018. http://www.aes.org/e-lib/browse.cfm?elib=19644

[7] Neukom, Martin. "Ambisonic Panning", paper presented at the *Audio Engineering Society 123*, New York, Oct 5-8, 2007. http://www.aes.org/e-lib/browse.cfm?elib=14354

[8] Zotter, Dr. Franz. *Spherical harmonics used for third-order HOA.* Image. https:// commons.wikimedia.org/wiki/File:Spherical_Harmonics_deg3.png

[9] *Mapping of B-format channels.* Image. https://www.pathpartnertech.com/introduction- ambisonics-360-degree-audio/

[10] Geluso, Paul and Lee, Kathleen "Ying-Ying". "The 3DCC Microphone Technique: A Native B-format Approach to Recording Musical Performance", paper presented at the *Audio Engineering Society 147*, New York, Oct 16-19, 2019. http://www.aes.org/e-lib/ browse.cfm?elib=20668

[11] Ryaboy, Alex. "Exploring 3D: A subjective evaluation of surround microphone arrays catered for Auro-3D reproduction system", paper presented at the *Audio Engineering Society 139*, New York, Oct 29-Nov 1, 2015. http://www.aes.org/e-lib/browse.cfm?elib=17987

[12] Dolby. *Home Theater Speaker Guide 7.1.4*. Dec 2018. Image. Dolby Atmos® Home Theater Installation Guidelines

[13] Munro, Andrew. "The Design of ATMOS Post Production Theatres", paper presented at the *Audio Engineering Society 137*, Los Angeles, Oct 9-12, 2014. http://www.aes.org/e-lib/ browse.cfm?elib=17414

[14] Rumsey, Francis. "Immersive audio, objects, and coding", *Journal of the Audio Engineering Society* Volume 63*,* Issue No. 5 (2015): pp. 394-398. http://www.aes.org/e-lib/ browse.cfm?elib=17642