Audio Engineering Society

Conference Paper 27

Presented at the AES 5th International Conference on Audio for Virtual and Augmented Reality
2024 August 19–21, Redmond, WA, USA

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Case Study on High Order Ambisonics Music Production: Music and Technology within an Ambisonics Framework Using Korean Traditional Instruments

Seungryeol Paik[1], Jisu Han[2], Taehyup Kim[3], and Kyogu Lee[1,4,5]

[1]*Department of Intelligence and Information, Seoul National University*
[2]*Department of Music, Seoul National University*
[3]*Dream Scape Inc.*
[4]*Artificial Intelligence Institute, Seoul National University*
[5]*Interdisciplinary Program in Artificial Intelligence, Seoul National University*

Correspondence should be addressed to Seungryeol Paik (paik402@snu.ac.kr)

ABSTRACT

This study explores immersive music production using fifth-order Ambisonics. We addressed challenges in project planning, musical expression, and technical constraints by creating the piece "The Borderless Space". We developed efficient workflows, applied recording and mixing methods with Korean traditional instruments, and ensured binaural compatibility. The piece won the gold prize at the 2023 IEM Student 3D Audio Production Competition, and the feedback highlighted areas for improvement. Additionally, we conducted a listening test to compare different audio formats in headphone environments and identified new challenges regarding compatibility across different monitoring environments.

1 Introduction

Immersive environments like Virtual Reality (VR) and Augmented Reality (AR) have advanced significantly, enhancing the perceived sense of reality for users and creators. While these environments have primarily been associated with visual technologies, there is growing interest and research in stimuli for other senses. Immersive and spatial audio, which enhances immersion through auditory stimuli, have received substantial attention. However, immersive audio systems are not widely adopted in the music industry, where most music is still produced in mono or stereo formats. Despite advances in binaural formats like Dolby Atmos [1] using Head-Related Transfer Functions (HRTF) [2, 3, 4], barriers remain for listeners and creators, including limitations in playback devices and production resources.

We identified three main challenges in immersive audio music: (1) Project Planning: The workflow from preproduction to distribution differs significantly from traditional stereo environments, requiring new approaches

and planning for artists and engineers. (2) Musical Expression: Research is needed on methods that fully utilize immersive systems, considering performance, placement, narrative, and message delivery. (3) Technical Constraints: Creators and listeners face resource constraints, such as the need for specific monitoring equipment for immersive audio production and playback.

To tackle these challenges, we conducted a case study based on our original immersive audio music piece, created using the Ambisonics system. Ambisonics is a 3D audio production system based on the physical principles of the sound field [5, 6, 7], which covers encoding, recording, post-production, transmission, and playback. First-order Ambisonics (B-Format) is standardized in global VR solutions like Google VR and Facebook 360°.

For our research, we adopted Higher Order Ambisonics (HOA), specifically fifth-order Ambisonics, for several reasons: (1) System Perspective: Unlike channel-based approaches, HOA does not have fixed channel counts and offers geometric flexibility. Additionally, it avoids the expense and proprietary elements of object-based techniques. (2) Listening Experience: HOA provides smooth panning without the artifacts found in methods like Vector Base Amplitude Panning (VBAP) [8]. It also requires fewer speakers than Wave Field Synthesis (WFS) [9]. (3) Delivery Perspective: Ambisonics supports the transfer of mixes to various technical environments, including binaural playback. It allows music production using free software, can decode to any audio format, and enables the creation of a fully periphonic sound field [6]. (4) Spatial Realization: HOA offers superior spatial rendering, creating a substantial "Sweet Spot" in the production studio for precise sound field reproduction [10, 11]. Therefore, Ambisonics and HOA present practical applications and potential for advanced expression and listening experiences in our research.

Previous research has extensively investigated the technical and perceptual aspects of the Ambisonics system [12, 13, 14]. Additionally, the production of audio content using the system has been actively researched, providing a technical and strategic foundation [15, 16, 17, 18, 19]. Music production based on Ambisonics and HOA has also been explored. For instance, Barrett [20, 21, 22, 23] investigated music composition, performance, and production in Ambisonics, focusing on the aesthetical, technological, and perceptual aspects. Furthermore, case studies on producing ambisonic music have been conducted across various genres and with traditional instruments [24, 25, 26].

Based on these researches, we produced HOA music from scratch rather than remixing already-made songs into Ambisonics. In this process, we aimed to address the challenges above of spatial audio production with the following objectives: (1) Exploring efficient planning, approaches, and workflows that assume an Ambisonics environment from pre-production, to effectively deliver our musical narratives and messages auditorily. (2) Investigating recording, placement, and performance methods to maximize the advantages of immersive audio, applying various digital and analog instruments, especially traditional Korean instruments. (3) Ensuring compatibility with headphones, considering that most listeners use headphones or earphones rather than surround speaker arrays while working with fifth-order Ambisonics.

The authors, Jisu Han and Seungryeol Paik produced a 9-minute 48-second original piece titled "The Borderless Space" using fifth-order hemispherical Ambisonics, 48kHz sampling rate, and 24bit resolution. This work was submitted to the Audio Drama, Documentary, and Soundscape categories at the 2023 IEM Student 3D Audio Production Competition (S3DAPC) and won the Gold Award. We also conducted a listening evaluation for different audio formats using headphones. Detailed explanations of the music's narrative can be found here `https://drive.google.com/file/d/17x3oydEsfHL1lm-Rjypl0BR02bcwrnoh/view?usp=drive_link`.

2 Ambisonics Music Production

In the pre-production stage, we established the spatial distinctions, narrative, and themes that form the foundation of the music and spatial expression. Based on this, we conducted the following production process: (1) Selection of Audio Materials. (2) Composition and Sound Design Planning. (3) Organization of Source Processing Methods. (4) Recording and Coloring of the Original Audio Sources. (5) Ambisonics Mixing – Spatial Arrangement. (6) Speaker and Binaural Monitoring. Composing and recording took place in the studios of the two authors. After recording, the captured sources were edited and processed into stems.

2.1 Themes and Concept

"The Borderless Space" is themed around "Space," narrating the experiences of dreams, reality, and the space in between from a first-person perspective. The goal was to fully utilize the spatial features of the Ambisonics system to successfully deliver the musical narrative, enhancing and implementing the spatial distinctions and characteristics of various audio sources. The explanations of spaces in this piece are briefly as follows. (1) Physical Space: The tangible world perceived through our senses, where sound behaves naturally, being reflected and absorbed by walls and furniture. (2) Imagenary Space: The realm of dreams and imagination, where sound often takes on exaggerated or distorted qualities. (3) Collapsing Space: The blurred boundary between physical and imaginary spaces, where sounds from the real world mix with those from our imagination, blurring the lines between reality and fantasy.

2.2 Source Materials

To create the sources, we used both digital and analog instruments, with a particular focus on traditional Korean instruments. These instruments, unlike Western ones, retain the natural flavor of raw materials, resulting in a rougher, less refined sound with unique playing methods, tuning, and tonality. For example, the friction between nails and silk strings on the Gayageum, the breathing before playing the Piri and the breathing when the sound disappears, and the gentle passing of a bow over silk threads on the Haegeum. We aimed to capture these unique sounds and ensure their harmony with the narrative and surrounding sounds, planning the source placement from an Ambisonic perspective before creation. Additionally, we used analog modular synthesizers, voices, ambient sounds, and composed or recorded additional materials.

2.3 Mixing

The overall system flow is illustrated in Figure 1. Basic sound adjustments were performed in Reaper [27], and immersive processing was carried out using SPAT Revolution (SPAT) [28]. From Reaper, audio was sent directly to SPAT without parallel or bus mixing. We conducted object-based mixing within SPAT. Instead of using an Ambisonic microphone to capture the sound, we used individual mono and stereo-recorded tracks and positioned the sources as objects in virtual space, encoding them into Ambisonics. This approach was chosen for the clarity of sources and the efficiency of immersive expression. The encoding was done in fifth-order Ambisonics using the HOA Room Module of SPAT, preferred for its detailed acoustic characteristics. The encoded sources were output for both speaker and headphone environments.

For speaker output, we used a system at Hanyang University's ICT Realistic Acoustical Laboratory (ICT Lab). The room, measuring 6.0 m x 6.0 m x 3.5 m, is equipped with a 24-channel loudspeaker (Leem C5) system. The room has an average RT60 of 0.74 seconds at 500 Hz and 1,000 Hz. The loudspeakers are installed with a diameter of 3.76 m x 3.76 m, with ceiling speakers at 2.8 m and ear-level speakers at 1.8 m. Due to the lack of loudspeakers to set up a half-sphere array, we had to arrange the monitor placement in a cube shape. Additionally, due to equipment and spatial restrictions, we paid particular attention to binaural headphone monitoring. For binaural output, Ambisonics audio tracks were sent to a Binaural Monitor track with the IEM Binaural Decoder [29]. We used multiple headphones or earphones for cross-validation, including Audio Technica M50x and R70x, Sennheiser HD 600 and HD 660S, Shure SRH840, Shure SE215, and Apple's AirPods Pro 2. Additionally, a six degrees of freedom (6DoF) setup with Supperware's Head Tracker 1 was implemented to monitor changes and distortions in the position and timbre of objects due to the listener's head movements and viewpoint changes.

2.3.1 Strategy

We approached the Ambisonic mixing with two main strategies. The first strategy was to preserve the meticulously reviewed and planned sources from pre-production, which are key elements for the auditory narrative. This involved maintaining the texture of the instruments and their unique spatial sense and arrangement. The second strategy was to mix according to the classification of audio sources, dividing them into (1) sources that create a sense of space, (2) sources that express the music, and (3) sources that emphasize spatial sound effects like movement. This classification, combined with pre-production planning, formed the basis of the Ambisonics mixing stage.

2.3.2 Stems

We used a total of 93 stereo tracks, with a 48kHz sampling rate and 24-bit resolution. Mono tracks recorded

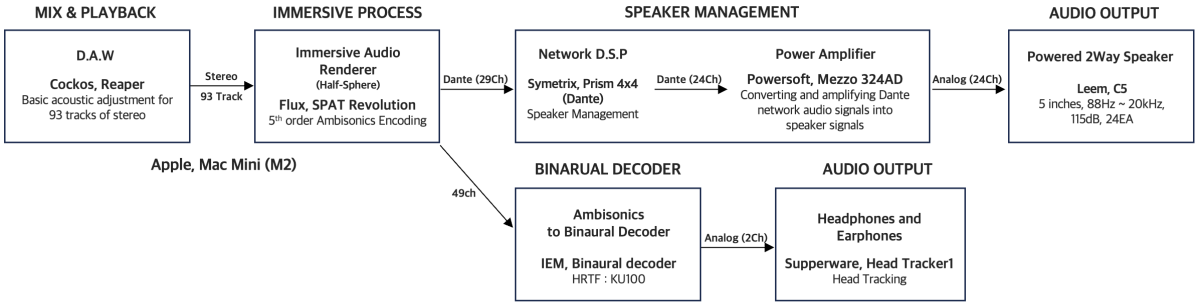


Fig. 1: Overview of our system flow

during the session were exported as dual mono. Basic tasks were performed in Reaper to process the stems. Equalizers and compressors enhanced texture and characteristics, while modulators were used minimally to maintain the unique spatial sense and context of the sources. These tasks were done conservatively to avoid compromising the original characteristics, especially the dynamics. Reverberation was not used in stem processing; instead, we used SPAT's 3D Reverb Plugin.

2.3.3 Abstract Space, Mapping and Movement

We followed two strategies and pre-established spatial sense and arrangement plans for each narrative section. The overall space was implemented using the perceptual parameters of SPAT's Room, and the spatial sense was enhanced with the global parameters of the 3D Reverb. Sources were placed according to three classifications. First, sources creating a sense of space were usually placed in the ceiling direction. We reduced their presence to allow the sources to naturally fill the space without making their sound images and directions too perceptible. Instruments like synth pads were placed on the ceiling for an enveloping feel. Second, sources expressing music focused on the harmony between instruments and were primarily distributed in the front to maintain the ensemble. Since most listening would be done facing forward, the main sources were placed in the front. We avoided placing sources directly at angles that would hit both ears directly unless they needed emphasis. Higher-pitched instruments were placed at higher elevations and lower-pitched ones at lower elevations for clearer sound imaging. We used consistent 3D Reverb characteristics for uniform spatial depth but applied different characteristics to make thematic and leading parts stand out. For strongly perceived sources, we minimized the use of reverb to express their depth closer in space. Third, sources emphasizing spatial sound effects were expressed with varying movement and placement according to their characteristics and velocity. Fast-moving sources were expressed with rapid movement, and slower sources moved more gradually. For mixed cases, the speed of the sources was adjusted according to the progression of their velocity.

2.3.4 Experience and Aesthetic

We began the music production with the premise of a spatial audio environment, experiencing an overlap between the roles of creators and audio engineers. Many decisions made during pre-production and music planning were revisited during the Ambisonics mixing stage, sometimes requiring us to return to the planning stage for detailed adjustments. From a technical perspective, working in an Ambisonics environment required a different approach compared to stereo. Traditional constraints, such as central positioning of low frequencies or main instruments and balancing left and right channels, were more flexible in Ambisonics. During the monitoring stage, we observed significant differences in listening experiences based on the decoding method. Depending on the speaker layout system, binaural listening (with or without head tracking), or converting to another format (e.g., downmixing to less than fifth-order Ambisonics or stereo), the efficiency of mixing and the expression of music and narrative varied widely. These issues were mostly addressed based on the creators' and engineers' expertise, experience, and judgment. From an aesthetic perspective, we prioritized using spatial audio and Ambisonics as creative tools. By approaching 'Space' as the theme of the original music and crafting narratives and compositions accordingly, we were able to provide clear spatial perception, enhance sound clarity, preserve instrument timbre, and improve rhythm conveyance in musical expression. We confirmed that rhythm was

perceived more clearly when sources were dispersed in the sound field, revealing complex rhythmic textures and subtle changes. However, the impact of rhythm, familiar in a stereo environment, was psychologically reduced. We also observed harmonic clarity improved by avoiding collisions between sources occupying the same spatial position within the large acoustic space of HOA. By avoiding frequency and magnitude masking, transparency varied according to the sources' positions. Unlike traditional stereo environments, the need for equalization was reduced, allowing us to convey the natural timbre of the sources efficiently and faithfully. Especially, during the analysis of Korean traditional instruments from a spatial audio perspective, we could observe the aforementioned aspects clearly. The transparency of spatial audio was particularly effective for Korean traditional instruments, allowing us to naturally express the texture of materials like silk strings and wooden bodies.

3 Evaluation

3.1 S3DAPC

"The Borderless Space" was evaluated by five juries from the S3DAPC, with each providing assessments and feedback based on specific criteria. The collective confidence level of the reviewers was 75% expertise and 25% high confidence. Positive feedback highlighted the composition and sound design for successfully integrating layered sounds with narrative elements, affirming the cohesiveness of the Ambisonic production. Reviewers praised the clear sound of stereo mix sources and the innovative use of traditional Korean instruments. However, they noted areas for refinement: while spatial transparency, mapping, and depth were well-received, improvements in precision for sound localization were suggested to enhance the immersive quality and audience experience.

3.2 Listening Test

The purpose of the test was to compare two different binaurally rendered Ambisonic productions and stereo-downmixed results in headphone environments and to examine listener preferences. The comparison involved three formats: (1) Binaurally rendered HOA audio with a fixed listening direction, using the KEMAR HRTF in Spat Revolution [30]. KEMAR was chosen over KU100 because it includes data on the torso and shoulders, making it more effective for our content with many ceiling objects. (2) Binaurally rendered HOA audio with a changing listening direction, using the CAT3DA player, which tracks the head via a camera [31]. This player, provided by S3DAPC, applies 3 degrees of freedom (3DoF) in listening. (3) Stereo downmixed HOA audio, following the stereo panning in SPAT.

Audio Format		CAT3DA (%)	KEMAR (%)	Stereo (%)
Spaciousness	Mapping	40	60	0
	Depth & Position	40	50	10
Selectivity		40	30	30
Timbre & Dynamics	Timbre	60	40	0
	Balance	50	50	0
General Impression		60	30	10

Fig. 2: Results table for the listening test, comparing 'CAT3DA' (head-tracking binaural render), 'KEMAR' (fixed-head binaural render), and Stereo

The experiment involved 10 listeners, including 3 audio engineers, 4 music producers, and 3 instrument players. Most had experience with surround and immersive systems, with confidence levels at 60% expertise and 40% high confidence. Participants received instructions on the music narrative, evaluation criteria, interpretation methods, and response submission methods. The experiment was conducted with listeners using their headphones in their own spaces. The first and second formats were shared anonymously as files, and the CAT3DA format was shared via a web player link with a user guide. Listeners provided their confidence levels, demographic information, and details of the headphones used. They switched between audio formats without time limits and determined their preferences, providing reasons and subjective feedback for each choice.

We referenced evaluation criteria from previous research related to the subjective evaluation of spatial and binaural content [24, 32]. These criteria provided a comprehensive framework for assessing the quality and effectiveness of the immersive audio content. The evaluation items included: (1) Spaciousness: The appropriateness of the audio's spatial sense. In our test, this was divided into spatial mapping and depth and position. (2) Selectivity: The clarity and distinction of

specific sources, layers, and instruments. (3) Timbre and Dynamics: The creation and balance of dynamics and timbre. In our test, this was divided into timbre balance, and dynamic naturalness. (4) Overall Impression: A general evaluation of the audio.

The results are presented in the table in Figure 2. Overall, binaurally rendered Ambisonics were preferred over stereo, especially the fixed-head binaural format in spatial mapping and depth. For selectivity, the head-tracking format was most preferred, though differences among the formats were minimal. In terms of timbre and dynamics, binaural formats were preferred over stereo, with a slight preference for head-tracking binaural over fixed-head binaural. The overall impression was highest for the head-tracking format, followed by the fixed-head format, with some preference for the stereo. Feedback indicated that most participants found that binaural formats implemented physical and realistic spatial sense well, with head-tracking providing particularly positive experiences. One participant found it difficult to empathize with the sound aesthetically, making preference marking challenging. Some found it hard to differentiate between the two Ambisonic binaural renders, while others noted the successful implementation of ceiling sources. Additionally, some found that the stereo downmix was perceived as most prominent at the same volume.

3.3 Discussion

Positive feedback and suggestions from the S3DAPC evaluation will guide us in improving tonal balance and dynamic range for better quality and proposing more refined auditory stimuli. We also realized the need to simplify the narrative to avoid abstract elements that could hinder the listening experience. For the additional listening test, we had specific expectations for the comparative evaluation of binaural and stereo environments using headphones. The results for spatial sense met our expectations, with binaural renderings showing a clearer spatial sense than stereo. Stereo received high evaluations for selectivity, likely because the spatial sense provided by binaural in a headphone environment made the sources less distinct. Stereo was not preferred for timbre and dynamics, due to the significant degradation of position, space, movement, and volume differences included in the Ambisonics mix during the downmixing. A key aspect we tested was the difference between head-tracked and fixed-head binaural environments. Although there were no significant differences in spatial sense, selectivity, timbre, and dynamics between the two formats, fixed-head binaural was slightly ahead in spatial sense, providing clear spatial perception and accurate reverberation simulation. The head-tracking format was slightly preferred for selectivity, as it allowed listeners to discern directionality and subtle sound attributes by changing their viewpoint, though overall clarity was inferior to stereo. In timbre and dynamics, head-tracking was slightly preferred due to its realistic feel. Overall, head-tracking received higher ratings due to the experience of sound changes according to viewpoint, and the narrative was more dramatically conveyed.

4 Conclusion

In this study, we produced immersive music using fifth-order Ambisonics and conducted case studies and evaluations. We investigated all stages from pre-production to post-production to maximize the advantages of immersive audio. Additionally, we analyzed traditional Korean instruments from a spatial perspective, which had not been done in previous studies. We aimed to intertwine the musical and narrative space with the technical space. We observed the potential for spatialization tools to be widely applied in both music and commercial media, especially at the intersection of composition and technology. This approach allowed us to reflect on spaces and objects from both technical and musical perspectives. However, despite the increasing use of Ambisonics in VR/AR audio, compatibility with various monitoring environments, including binaural, still requires sophisticated technical solutions. This presents a new challenge of developing technical approaches to monitor more consistently and reduce variance in the listening environment. Our future work will focus on creating new Ambisonics-based content with refined spatialization and internal object spatial implementation. We will also explore compatibility in binaural environments, which are more accessible to the general public. Furthermore, we plan to conduct more thorough listening evaluations under different conditions.

References

[1] Sergi, G., "Knocking at the door of cinematic artifice: Dolby Atmos, challenges and opportunities," *The New Soundtrack*, 3(2), pp. 107–121, 2013.

[2] Ellis-Geiger, R. J., "Music production for Dolby atmos and auro 3D," in *Audio Engineering Society Convention 141*, Audio Engineering Society, 2016.

[3] Bresler, Z., "Immersed in Pop: 3D Music, Subject Positioning, and Compositional Design in The Weeknd's "Blinding Lights" in Dolby Atmos," *Journal of Popular Music Studies*, 33(3), pp. 84–103, 2021.

[4] Scott-Deuchar, S., "Dolby Atmos Music and the Production of Risk," in *Creative Tools and the Softwarization of Cultural Production*, pp. 243–262, Springer, 2024.

[5] Zotter, F. and Frank, M., *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*, Springer Nature, 2019.

[6] Arteaga, D., "Introduction to ambisonics," *Escola Superior Politècnica Universitat Pompeu Fabra, Barcelona, Spain*, pp. 17–18, 2015.

[7] Furness, R. K., "Ambisonics-an overview," in *Audio Engineering Society Conference: 8th International Conference: The Sound of Audio*, Audio Engineering Society, 1990.

[8] Zotter, F. and Frank, M., "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, 60(10), pp. 807–820, 2012.

[9] Spors, S. and Ahrens, J., "A comparison of wave field synthesis and higher-order ambisonics with respect to physical properties and spatial sampling," in *Audio Engineering Society Convention 125*, Audio Engineering Society, 2008.

[10] Kessling, P. and Görne, T., "Studio for immersive media research and production: Immersive Audio Lab at HAW Hamburg," in *Audio Engineering Society Convention 145*, Audio Engineering Society, 2018.

[11] Frank, M. and Zotter, F., "Exploring the perceptual sweet area in ambisonics," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[12] Sato, K. and Haneda, Y., "Filter design of a circular loudspeaker array considering the three-dimensional directivity patterns reproduced by circular harmonic modes," in *Audio Engineering Society Convention 142*, Audio Engineering Society, 2017.

[13] Wittek, H. and Theile, G., "Development and application of a stereophonic multichannel recording technique for 3D Audio and VR," in *Audio Engineering Society Convention 143*, Audio Engineering Society, 2017.

[14] Riitano, L. and Medina Victoria, J. E., "Comparison between different microphone arrays for 3d-audio," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.

[15] Neukom, M., "Ambisonic panning," in *Audio Engineering Society Convention 123*, Audio Engineering Society, 2007.

[16] Frank, M., Zotter, F., and Sontacchi, A., "Producing 3D audio in ambisonics," in *Audio Engineering Society Conference: 57th International Conference: The Future of Audio Entertainment Technology–Cinema, Television and the Internet*, Audio Engineering Society, 2015.

[17] Lennox, P. P., Myatt, T., and Vaughan, J. M., "From surround to true 3-D," in *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*, Audio Engineering Society, 1999.

[18] Benjamin, E., Heller, A., and Lee, R., "Why ambisonics does work," in *Audio Engineering Society Convention 129*, Audio Engineering Society, 2010.

[19] Chapman, M., "New dimensions for Ambisonics," in *Audio Engineering Society Convention 124*, Audio Engineering Society, 2008.

[20] Barrett, N., "Spatio-musical composition strategies," *Organised sound*, 7(3), pp. 313–323, 2002.

[21] Barrett, N., "Ambisonics spatialisation and spatial ontology in an acousmatic context," in *Proc. from the Electroacoustic Music Studios conference*, 2010.

[22] Barrett, N., "The perception, evaluation and creative application of high order ambisonics in contemporary music practice," *Ircam Musical Research Residency report*, 2012.

[23] Barrett, N., "Composing Images in Space: Schaeffer's Allure Projected in Higher-Order Ambisonics," in *Proceedings of the International Computer Music Conference*, pp. 47–52, New York, 2019.

[24] Malecki, P., Piotrowska, M., Sochaczewska, K., and Piotrowski, S., "Electronic music production in ambisonics-case study," *Journal of the Audio Engineering Society*, 68(1/2), pp. 87–94, 2020.

[25] Karadogan, C. and Görne, T., "Auditory Scenography in Music Production: Case Study Mixing Classical Turkish Music in Higher Order Ambisonics," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.

[26] Barboza, J. R., Magalhães, E., and Bernardes, G., "Towards Best Practices in Spatial Audio Post Production: A Case Study of Brazilian Popular Music," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–7, IEEE, 2021.

[27] Cockos, "REAPER," 2005, `https://www.reaper.fm/`.

[28] FLUX, "SPAT Revolution," 2017, `https://www.flux.audio/project/spat-revolution/`.

[29] IEM, "IEM Plug-in Suite," 2018, `https://plugins.iem.at/` [Accessed: (2023-05-05)].

[30] Braren, H. and Fels, J., "A High-Resolution Head-Related Transfer Function Data Set and 3D-Scan of KEMAR," *Institute for Hearing Technology and Acoustics, RWTH Aachen University, Technical report*, 2020.

[31] Gölles, L. and Frank, M., "Cat3DA-Camera-Tracked 3D Audio Player," in *DAGA*, 2023.

[32] Kleczkowski, P., Krol, A., and Malecki, P., "Multichannel sound reproduction quality improves with angular separation of direct and reflected sounds," *Journal of the Audio Engineering Society*, 63(6), pp. 427–442, 2015.