Field Report on 3D Audio Capture of Solo Piano for Classical Music Productions*

EMRE EKİCİ,[1] WILL HOWIE,[2] TORU KAMEKAWA,[3]
(mrekici@outlook.com) (wghowie@gmail.com) (kamekawa@ms.geidai.ac.jp)

[1] *School of Performing Arts, University of Otago, Dunedin, New Zealand*
[2] *JSPS International Research Fellow, Tokyo University of Arts, Tokyo, Japan*
[3] *Department of Musical Creativity and the Environment, Tokyo University of Arts, Tokyo, Japan*

This field study aims to help recording engineers and producers understand the aesthetics of different techniques for 3D audio capture of solo piano in classical music productions. 3D audio gives the listener a sensation of auditory height, width, and depth compared to traditional stereo and surround sound arrangements. Since the early 2000s, there has been an increasing spread of new 3D audio formats and an increase in attention from academics, industry experts, businesses, and consumers about the different uses of 3D audio. 3D audio formats demand new content creation methods optimized to better utilize this larger sound field. Microphone placement techniques are crucial for recording engineers and producers to capture the required spatial and tonal qualities of a given auditory scene. Research on acoustic music recording for 3D audio is growing; however, classical music productions still form a small part of the literature. Since solo piano offers a common sound relevant to both chamber and orchestral music, production techniques for solo piano can be used to address this gap in 3D audio capture of classical music. In this study, a controlled recording session of solo piano for Dolby Atmos (ITU 4+7+0) is investigated. The setup is explained briefly as an overview; then, the recording methodology for microphone placements is demonstrated in detail, focusing on different spatial and timbral attributes. Furthermore, the rationale behind these microphone placements and the decision-making process during the pre-production, production, and post-production stages are discussed. Recommendations are generated for recording engineers and producers to comprehend several techniques to optimize their technical and aesthetic objectives.

Keywords: immersive audio, sound recording practices, classical music production, art of record production, recording journal

0 INTRODUCTION

3D audio has the potential to reinforce auditory experiences towards compelling experiences, provided by a wide variety of technical opportunities (Roginska & Geluso, 2017, p. xii); these technical opportunities can lead to enhanced listener envelopment (Leonard, 2017, p. 348). There have been many 3D audio formats created for commercial and broadcast purposes, many of which have been standardized by the International Telecommunication Union (ITU) (2022). Although many aspects of 3D audio production and reproduction are frequently addressed by researchers (e.g., Howie et al., 2020; Kamekawa & Marui, 2020; Lee, 2021), the increasing trend in music production for Dolby Atmos (ITU 4+7+0) format (e.g., Bresler, 2021, p. 1, Immersive Audio Album, 2023) showed a small gap regarding the reproduction of M±090 (side surround) loudspeakers at this intersection of academia and industry.

This field report describes a controlled recording session to address the overarching research question: ”How do different off-axis microphone placements on solo piano contribute to early and background spatial impressions provided by M±090 loudspeaker channels in classical music productions?” Since recording engineers mediate between artistic aspirations and the technical realization of the sound, the authors have also addressed other issues regarding the 3D audio capture of solo piano for classical music productions, which will be thoroughly described in this field report.

*3D audio recordings can be accessed at: https://doi.org/10.5281/zenodo.8023989

1 JUSTIFICATION

3D audio offers excellent potential for delivering more immersive listening experiences as compared with traditional stereo playback systems (Leonard, 2017, p. 333). The multi-dimensionality of 3D audio offers a challenge to engineers and researchers, as now there are more variables to investigate, and the relationship between these new variables in recording situations is understudied (Martin et al., 2015, p. 1). In this regard, to yield meaningful results, 3D audio experiments need to be highly specific and, ideally, should only investigate the impact of one variable at a time. This field study focuses on the context of one kind of music (classical), one sound source (piano), and one static microphone array; only one of the variables is changed (M±090 loudspeaker reproduction). Also, the performance variable is eliminated by utilizing a Yamaha Disklavier piano, which does not require a performer and can playback any MIDI file at any tempo, key, and volume. This field study focuses on investigating recording techniques for M±090 loudspeaker reproduction, the only variable in the recording setup.

Another important consideration is that 3D audio is still a new domain in the history of music. When the Roman architects started building large stone basilicas in the 4th century AD, the performing stage was no longer in the open air: this spatial change impacted both performer and audience expectations and aesthetics. The semantic content of the spoken words in the open air was slowly transitioning towards an experience in a reflective space (Boren, 2017, p. 43). The early days of transitioning from mono to stereo also had a similar process regarding the shift in aesthetic understandings. A new medium required novel aesthetics. The traditional ways of producing stereo recordings are still valid to some extent in 3D audio; however, these ways are not supposed to be limitations for 3D audio (Howie, 2019, p. 9). While exploring M±090 loudspeaker reproduction of various microphone placements, this field study is also helping to build new understandings of aesthetics and/or validate some previous practical experience in sound recording.

Recent developments in the accessibility of 3D audio production are fundamental for content creators composing new music specifically for new audio medium formats. New works will be designed for the latest audio playback formats: current recording engineers need to adapt to these specifications from content creators. This field study reflects the perspective of recording engineers in such circumstances where they are supposed to adapt to a specific repertoire, instrumentation, and room.

The authors value documenting this transition period from stereo to 3D audio to future-proof their work, hoping to give an idea about how the recordings were assembled for future researchers, taking into consideration that there is relatively little analysis of the processes and systems by which production decisions are implemented by recording engineers in the 3D audio production of classical music. Since we can only reverse engineer earlier recordings done in the past and the transition from mono to stereo from the perspective of recording engineers was skipped in the accessible related literature, this field report contributes to the historical aspect of the art of record production.

2 PRE-PRODUCTION

The pre-production stage was mainly about deciding upon theoretical microphone placements of possible reproductions for M±090 loudspeakers and the repertoire (test stimuli) that will be recorded.

2.1 Suggested Microphone Placements for M±090

As ITU (2022) suggests, M±090 loudspeakers are pointed almost directly (+85○ to +110○ and -85○ to -110○) at the listeners' ears and should not compromise the front image quality (Howie, 2018, p. 59). In contrast to commercial pop music, the performance of classical music creates a stable visual cue, and the sound localization combined with this image should not be disregarded by sound reproduction (King, 2016, p. 51). While preserving the main localization, the information sent to the M±090 loudspeakers should help to spread the frontal sound image, ensuring the prevalence of spatial information and reflections while maintaining the stability of the frontal sound. Also, not creating a phantom image on M±090 loudspeakers is recommended to maintain frontal image quality. Another issue is that, similar to vertical interchannel crosstalk (Wallis & Lee, 2020), unwanted localization blur/shift and image spread in the horizontal axis may occur when significant delayed direct information of the sound source is sent to M±090 loudspeakers. Overall, to have increased envelopment, the information sent to the M±090 loudspeakers need to meet different parameters.

Only two recommendations were made about M±090 placements in earlier research. Hamasaki Hiyama (2003) argue that since ambiance arrays are placed only to capture diffuse sounds, they are recommended to be placed beyond the critical distance of the recording venue. Also, Kamekawa Marui (2020) state that the null of the microphone pointing to the sound source can suppress the on-axis information from the stage while focusing on early reflections and reverberation from side walls.

In summary, to record solo piano in 3D audio, microphone placements for M±090 loudspeakers need to satisfy the recommendations (ideally all): 1) contain diffused and decorrelated signal that is mostly lateral reflected energy, 2) maintain frontal image quality and stability, 3) spread out the total environmental image by enhancing the environment width, 4) do not create a conflicting phantom image with the frontal sound image.

2.2 Repertoire / Test Stimuli

Three MIDI pieces from contrasting style periods were suggested for Yamaha Disklavier's playback to understand possible aesthetic implications better: 1) Johann Sebastian Bach, Prelude and Fugue: No. 6 in D Minor, BWV 851, 2) Wolfgang Amadeus Mozart, Piano Sonata in F major, K.

332, 2nd movement, 3) Franz Liszt, Hungarian Rhapsody No. 10 in E major, S. 244/10.

Bach's BWV 851 is selected to investigate transients, auditory streams, and sounds with shorter decay times. Mozart's K. 232, 2nd movement is preferred for a blend of sustained chords and faster melodies. Liszt's S. 244/10 is chosen due to having a wider range.

3 PRODUCTION / RECORDING TEST STIMULI / METHODOLOGY

All channels described in this field report were recorded at 96kHz/24bit resolution in Studio A at Tokyo University of the Arts' Senju Campus: floor area = 175.5 m², ceiling height = 4.8 m on the north end, 7 m on the south end, reverb time = approx. 1.0 s at 500 Hz. All microphones were routed to ADT microphone preamplifiers, then to Digital Audio Denmark (AX32) analog-to-digital converters. The microphone array used in this field study is based on one-to-one routings between microphones and loudspeakers. Microphone choice and placement are described in Table 1, and Figures 1 and 2, where each microphone is colored based on resistor color codes (UL/CSA) assigned in the order in Table 1. Omnidirectional microphones are marked with a circle, and bi-directional microphones are marked with a double arrow. The arrows point to the direction of the front of the microphone. Placements with only an arrow show that these microphones have a cardioid polar pattern. Photos of the sessions with the same color coding can be seen in Figure 3 and Figure 4.

The first approach was to decide upon the piano's placement. Yamaha Disklavier was placed on the north end of the room, facing the south. The expansion of the ceiling from the north end to the south end created more balanced reflections than placing the piano on the south end while facing the north end. The acoustics in the studio are relatively dry, given the size of the room, and this helped the recording team better analyze and understand the piano's sound.

Once the recording team agreed upon the piano placement and repertoire, the main array placements were decided. Before implementing M±090 variations, the main array placement was improved iteratively during in-situ evaluations. For instance, the initial placement for the Main L-C-R group created a bass buildup on the center channel. This buildup was resolved by placing the Main R microphone farther, around 40 cm. Another issue encountered was the slight incompatibility between the sonic characters of different microphones. Multi-pattern microphones (e.g., AKG C414) were initially preferred since they offer more flexibility. However, these microphones were changed to Schoeps (single diaphragm and single polar pattern) to have a more integrated character between the remaining microphones, and the recording team committed to current microphone choices, placements, and polar patterns. During another round of in-situ evaluations, the recording team realized that the piano would benefit from a closer representation and decided to add a pair of spot microphones.

Table 1. Microphones for solo piano

Placement	Microphone	Polar Pattern	Color
Spot L	Sony C100	Cardioid	Black
Spot R	Sony C100	Cardioid	Black
Main L	Schoeps MK2	Omni	Brown
Main C	Schoeps MK2	Omni	Brown
Main R	Schoeps MK2	Omni	Brown
B 1 L	Schoeps MK4	Cardioid	Red
B 1 R	Schoeps MK4	Cardioid	Red
B 2 L	Schoeps MK6	Bi-directional	Orange
B 2 R	Schoeps MK6	Bi-directional	Orange
Top F L	DPA 4011	Cardioid	Yellow
Top F R	DPA 4011	Cardioid	Yellow
Top B L	DPA 4011	Cardioid	Green
Top B R	DPA 4011	Cardioid	Green
Si 1 B	Schoeps MK2	Omni	Blue
Si 1 F	Schoeps MK2	Omni	Blue
Si 3 L	MKH8040	Cardioid	Purple
Si 3 R	MKH8040	Cardioid	Purple
Si 4 L	Schoeps MK2	Omni	Grey
Si 4 R	Schoeps MK2	Omni	Grey
Si 5 L	OC818	Custom	White
Si 5 R	OC818	Custom	White

Once the main array sound was satisfactory, the recording team started to implement M±090 variations over the main array. These variations also required a similar workflow mentioned above related to integrating various microphone models. During the final in-situ evaluation, one of the M±090 variations was changed to an ambiance variation (renamed as B 2 pair) since it did not offer anything new for M±090 loudspeaker reproduction, but once placed farther than the B 1 pair, this pair gave another aesthetic option for surround, which will be described in the post-production section.

4 POST-PRODUCTION / COMPARING TEST STIMULI

A good deal of balancing was done on the production day, and the recording team already had an idea of the overall levels. Before the A/B comparison of test stimuli, some manipulations of the channels were required. Initially, HPF was applied to the Top Front channels at 230 Hz with a 12 dB per octave slope and to the Top Back channels at 100 Hz with a 12 dB per octave slope. These EQ moves helped to remove overall bass buildup and gave a better impression of space and localization. Also, unwanted resonance was removed from Top channels at 900 Hz with a Q of 1.85 for about -6 dB. The next step was to apply LPF to Main L-C-R at 300 Hz with a slope of 12 dB per octave to utilize the intimacy of the Spot pair. This move helped the recording team capture the piano's clarity without disrupting localization. Lastly, the B 1 pair had an EQ notch at 490 Hz with a Q of 1.8 for about -3 dB to minimize boxiness. These manipulations were the only EQ interventions.

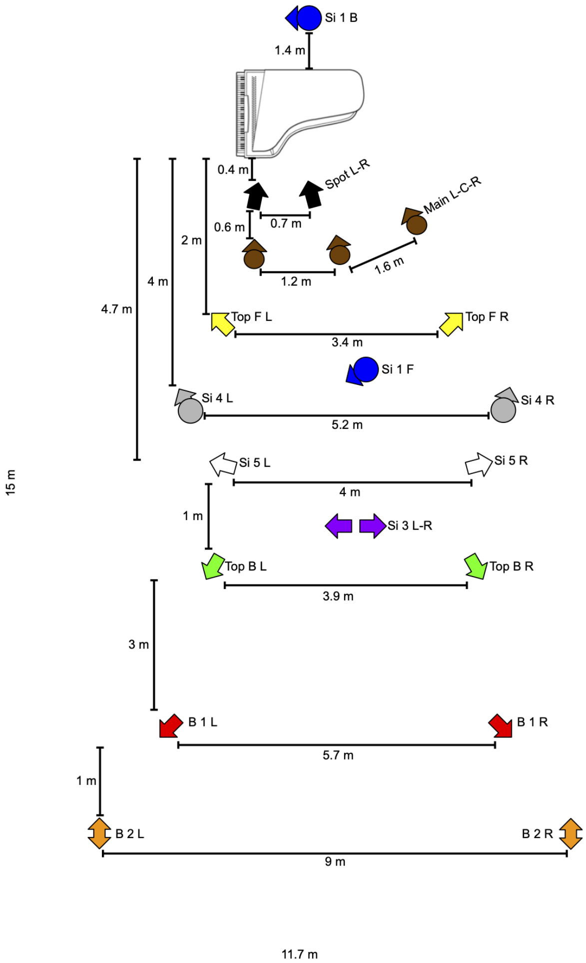

Fig. 1. Piano microphone placement, overhead view

Once the rough mixing of the main array was completed, M±090 variations were level matched for A/B comparison. The recording team first noticed that Si 3 pair was narrowing the envelopment. Si 1 and Si 4 pairs were found to be decorrelated compared to the main microphone array, to the extent that they did not contribute significantly to the overall sound scene. Si 5 option was recorded with a multi-output multi-pattern microphone (Austrian Audio OC818), which allowed for comparison of cardioid, omni, and bi-directional polar patterns from a single microphone position. In this regard, Si 5 bi-directional variation consisted of significant early reflections and dragged the central image towards the sides. This option seemed to create a slight localization blur between the frontal image and M±090 loudspeakers. Si 5 cardioid option was considered to improve listener envelopment, but it captured more direct signal than anticipated, which was dragging the center image towards the sides. Overall, the consensus was to blend Si 5 bi-directional and Si 5 cardioid signals for the ideal balance of on-axis/off-axis signals on M±090 loudspeaker reproduction.

As mentioned in the production section of this report, B 2 pair was initially considered for M±090 variations; nonetheless, during in-situ evaluations, this pair served better as ambiance microphones. Comparing B 1 and B 2 pairs was also beneficial since they offered different characteristics due to their placement and polar pattern. After balancing the levels of B 1 and B 2 pairs, the B 1 pair gave an impression of an equal circle, whereas the B 2 pair created an extended impression towards the back. The difference was minor, but they served different aesthetic purposes. Overall, B 1 was a better fit for the selected repertoire.

At this point, the recording team was satisfied with the decisions about which tracks to use and started to consider implementing time-based effects. First, convolution reverb (Audio Ease Altiverb 7) on each channel (except M±090 variations and Spot pair) were added as inserts. The same hall preset with alternative options were used on each instance of the plugin and adjusted by ear accordingly to the related microphone's position. The dry/wet ratio was kept low in each instance, and both damping and time parameters were adjusted. EQ on the plugin was also necessary to remove bass buildup on several occasions. Time alignment can be done at the start of the mixing by delaying signals. However, in this case, the recording team was already satisfied with the results without time alignment, which also could be done later to support the current sound.

5 DISCUSSION

This field study started as an experiment to try different M±090 variations for 3D audio capture of solo piano. During in-situ evaluations in the production stage, it was realized that initial plans deviated partially. Therefore, it is vital to keep the planning as flexible and adaptive as possible. This mindset is crucial for recording engineers: improvisation is also a part of the craft since recording engineers have to work for the room during their decision-making process and let go of their ego/intuition if the desired sound is not captured as planned. It is not possible to think through some variables in recording situations. As 3D audio capture requires more setup time compared to stereo recordings, flexible and adaptive planning become even more important.

For M±090 variations, although the recording team agreed that the Si 5 cardioid option provided the ideal listener envelopment, most other options were also deemed acceptable for different aesthetics or repertoire. Nonetheless, the recording team also agreed that Si 3 pair (upwards-facing ORTF, near coincidental) was unusable because it decreased the listener envelopment by narrowing down the whole image. The recording team observed that this microphone pair disrupted the whole experience in this case. Also, excessive direct information from multiple directions in the horizontal plane in one-to-one routings creates competing/conflicting early reflection points in the sound scene, which might not be desired results for most aesthetics. The final decision should be based on what serves the integration to the sound field best.

During the production stage, the recording team agreed that critical listening for 3D audio is slightly more demanding compared to stereo / traditional formats in terms of cognitive effort. In order to address this challenge, the record-

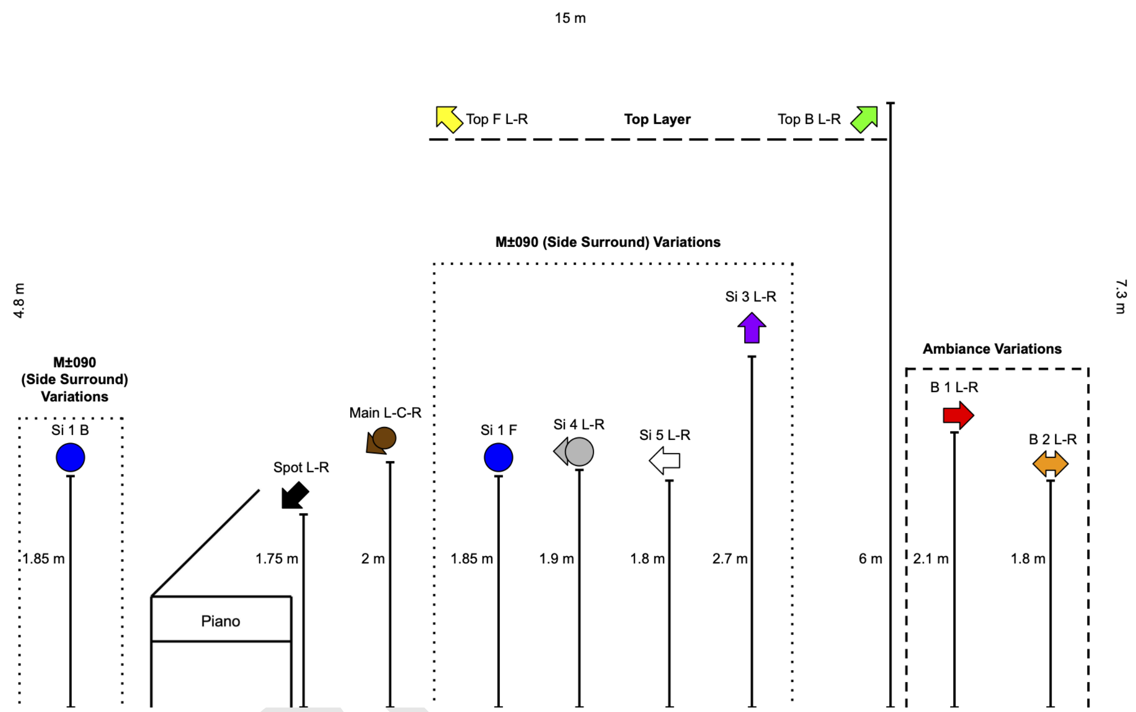

Fig. 2. Piano microphone placement, side view

Fig. 3. Piano microphone placement, view from the north end (upstage right)

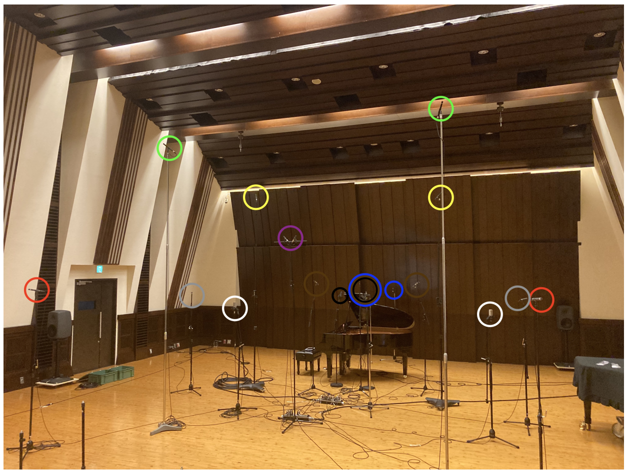

Fig. 4. Piano microphone placement, view from the south end (house right)

ing team gave a cognitive break from 4+7+0 format the following morning. This mental break was allocated to listening to 3D recordings made for NHK 22.2 (ITU 9+10+3) and NHK 27.2 (ITU 9+10+8) formats. Since the afternoon was allocated to the mixing (around 24 hours break from the production), these evaluations of other mediums during the morning provided a change of focus, and returning to listening 4+7+0 was now a new experience.

Another realization of this field study was the advantage of using Yamaha Disklavier for controlled recording sessions. Although the performance of this instrument does not offer any style or character, it does perform any material precisely, which helps recording engineers to focus on the recorded sound and possible improvements. For instance, this experiment would be difficult to conduct if there was a need to book a pianist for an entire day and ensure that the pianist was ready to perform any piece continuously. However, when a musician hears themselves right after the recording, if the relationship between the performer and place is well captured, the player gets inspired and immediately reflects on this on the next take. This lack of human touch is a disadvantage of using Yamaha Disklavier since the medium could not interfere with the player's performance.

The authors were interested in sharing this field report not only as an experiment but also as an example log of the craft for future proofing their work. The authors consider this work an extension of the primary idea behind recording sound. "How we assembled the music in the past?" is a question that historical musicologists are interested in since it is not possible to find any formal or written sources in most of the released / commercial music. This field report exemplifies increasing the value of metadata in record production.

This field report should not be perceived as a microphone comparison test since authors think such comparisons have little significance in academic and practical work. These comparisons give an idea about the microphones, but usually, they are not applicable in real-life situations. They are considered more of an exercise that puts the recording engineer in a situation to listen critically and judge what serves well to the situation. Then, making decisions from these comparisons – deciding which microphone, placement, height, polar pattern, and angle-azimuth – gets easier. Such practice is helpful in the long run as the practical knowledge gained from these virtual scenarios is usually needed for recording engineers working in unfamiliar situations. In such comparisons, learning what does not work is also important since, in this process, we also learn to validate some findings/experiences in sound recording, e.g., the microphone capsule direction should roughly mirror the playback speaker direction (Howie et al., 2016, p. 4) or wider horizontal microphone spacing would lead to a stronger spatial impression in reproduction due to a greater magnitude of interchannel decorrelation (Hamasaki & Hiyama, 2003; Rumsey & Lewis, 2002).

Multi-output multi-pattern microphones would be ideal when the microphone placement was committed but not the polar pattern. In this case of microphone placement for M±090 loudspeaker reproduction, a blend of cardioid and bi-directional signals might be necessary to balance the amount of early reflections and decorrelated signals.

6 CONCLUSION

This field report narrates the 3D audio capture of solo piano for classical music productions in studio settings. The techniques and decision-making processes described in this field report are based on technical and practical knowledge supported by in-situ evaluations. The purpose of the report is to provide guidance for recording engineers and producers who are creating content for immersive audio systems; since many distinct recording scenarios exist, a single technique is generally insufficient to satisfy the technical and aesthetic needs. Optimizing technical and aesthetic objectives during recording sessions is essential for superior craftsmanship in emerging immersive audio systems. The recorded material will be made available for other researchers upon request.

7 ACKNOWLEDGMENTS

The authors would like to thank the support of Japan Society for the Promotion of Science and Tokyo University of the Arts. Thank you to Julius Johansson and Dr. Atsushi Marui for their contributions and evaluations during the field study.

8 REFERENCES

Boren, B. (2017). History of 3D Sound. In A. Roginska P. Geluso (Eds.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (pp. 40–62). Routledge. https://doi.org/10.4324/9781315707525

Bresler, Z. (2021). 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), 84–103. https://doi.org/10.1525/jpms.2021.33.3.84

Hamasaki, K., & Hiyama, K. (2003). Reproducing Spatial Impression With Multichannel Audio. *24th Audio Engineering Society International Conference: Multichannel Audio, The New Reality*. Paper 19.

Howie, W. (2018). *Capturing orchestral music for three-dimensional audio playback* (Doctoral dissertation, McGill University, Schulich School of Music, Department of Music Research).

Howie, W. (2019). Pop and Rock music audio production for 22.2 Multichannel Sound: A Case Study. *146th Audio Engineering Society Convention, AES 2019*, 1–11. Paper 10150.

Howie, W., King, R., & Martin, D. (2016). A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound. *141st Audio Engineering Society Convention, AES 2016*, 1–11. Paper 9612.

Howie, W., Martin, D., Kim, S., Kamekawa, T., & King, R. (2020). Effect of Skill Level on Listener Performance in 3D Audio Evaluation. *Journal of the Audio Engineering Society*, 68(9), 628–637. https://doi.org/10.17743/jaes.2020.0050

Immersive Audio Album. (2023). *Coming soon in 2023: Immersive Music Preview*.

International Telecommunication Union. (2022). Multichannel Sound Technology in Home and Broadcasting Applications BS Series. ITU-R BS.2159-9.

Kamekawa, T., & Marui, A. (2020). Evaluation of Recording Techniques for Three-Dimensional Audio Recordings: Comparison of Listening Impressions Based on Difference Between Listening Positions and Three Recording Techniques, *Acoustical Science and Technology*, 41(1), 260–268. https://doi.org/10.1250/ast.41.260

King, R. (2016). *Recording Orchestra and Other Classical Music Ensembles*. Routledge.

Lee, H. (2021). Multichannel 3D Microphone Arrays: A Review. *Journal of the Audio Engineering Society*, 69(1/2), 5–26. https://doi.org/10.17743/jaes.2020.0069

Leonard, B. (2017). Applications of Extended Multichannel Techniques. In A. Roginska P. Geluso (Eds.), *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* (pp. 333–355). Routledge. https://doi.org/10.4324/9781315707525

Martin, B., King, R., Leonard B., Benson, D., & Howie, W. (2015). Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music. *138th Audio Engineering Society Convention, AES 2015*. Paper 9251.

Roginska, A., & Geluso, P. (Eds.). (2017). *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Routledge. https://doi.org/10.4324/9781315707525

Rumsey, F., & Lewis, W. (2002). Effect of Rear Microphone Spacing on Spatial Impression for Omnidirectional Surround Sound Microphone Arrays. *112th Convention of the Audio Engineering Society, AES 2002*. Paper 5563.

Wallis, R., & Lee, H. (2020). Localisation of Vertical Auditory Phantom Image with Band-limited Reductions of Vertical Interchannel Crosstalk. *Applied Sciences*, 10(4), 1490. https://doi.org/10.3390/app10041490