Audio Engineering Society

Conference Paper 3

Presented at the International Conference on Spatial & Immersive Audio,

2023 August 23-25, Huddersfield, UK

Case Studies in Music Production for 3D Audio Reproduction with Bottom Channels

Will Howie[1], Toru Kamekawa[2], and Miki Morinaga[2]

[1]Japan Society for the Promotion of Science International Research Fellow, Tokyo University of the Arts, Tokyo, Japan
[2]Department of Musical Creativity and the Environment, Tokyo University of the Arts, Tokyo, Japan

Correspondence should be addressed to Will Howie (`wghowie@gmail.com`)

ABSTRACT

This report describes the production of several immersive musical sound scenes for a 9+10+8 (27.2) 3D audio reproduction system. 9+10+8 features an even spatial distribution of loudspeaker channels in three vertical layers: top (above the listener), main (ear-level), and bottom (below the listener). 3D sound recording and mixing methodologies are discussed with reference to four specific sound scenes: solo jazz piano, large pipe organ in a concert hall, alternative rock, and a taiko drum ensemble, as well as in more general terms. This report aims to help fill a current gap in published knowledge surrounding music production techniques for 3D audio systems featuring bottom-layer reproduction channels, such as NHK 22.2, Sony 360 Reality Audio, and various binaural audio production and rendering mediums.

1 Introduction

Numerous commercial and broadcast 3D audio formats have been introduced, many subsequently standardized by the International Telecommunications Union (ITU) [1]. 3D sound fields can be reproduced using either an array of loudspeakers, or over headphones using binaural rendering [2]. In either case, an optimal presentation of a 3D sound scene would include sound reproduction from all around the listener: not just at head-height, but also from above and below, as with real-life listening. Several currently available 3D audio formats already include bottom-layer loudspeaker-based sound reproduction, while binaural audio rendering tools theoretically allow for the panning of sound sources to any location in the horizontal or vertical plane.

Much of the previous work introducing concepts or techniques for 3D music recording has focused exclusively on how to capture audio for elevated "height channels," i.e., loudspeakers placed above the listener [2, 3]. Comparatively few authors have discussed music production techniques optimized for 3D audio reproduction systems including "bottom channels," i.e., loudspeakers placed at or near floor-level. This report describes several case studies in audio production of musical sound scenes for a 9+10+8 (27.2) 3D audio reproduction system.

2 Background

2.1 9+10+8 (27.2) audio reproduction

All recordings described herein were recorded for a custom 9+10+8 (27.2) 3D audio reproduction system. Following ITU naming conventions [1], 9+10+8 refers to a reproduction system with nine height channels, ten "main layer" channels (loudspeakers placed roughly at ear-level), and eight bottom channels. 9+10+8 is identical to 9+10+3 (also known as NHK 22.2 Multichannel Sound) [4] in terms of number and spatial positions of loudspeakers, but adds bottom channels for the Side Left and Right, and Rear Left, Centre, and Right loudspeaker positions (Fig. 1). The addition of the five bottom channels gives an even spatial distribution of loudspeakers in all three vertical layers, which allows for the potential to create highly realistic or hyper-realistic presentations of various 3D sound scenes.

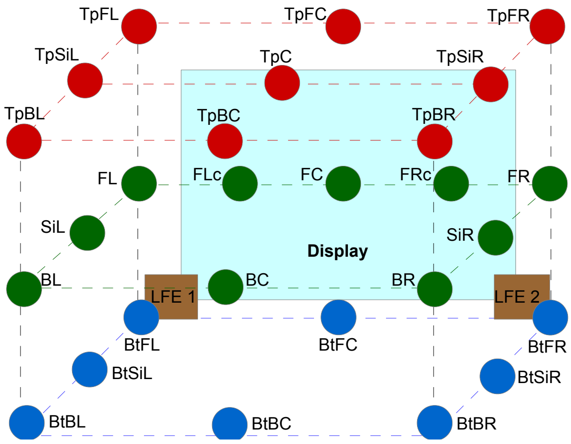

Figure 1. 9+10+8 channel/loudspeaker physical layout. Channel nomenclature as per ITU [1].

2.2 Music production techniques for 3D audio

Various techniques for capturing acoustic music for 3D audio reproduction have been discussed, many of which are described within Lee's [3] extensive review of multichannel 3D microphone arrays. These techniques tend to be optimized for use with smaller-scale 3D audio formats, and within Lee's review only two techniques specifically include microphones for bottom layer loudspeakers: “Hamasaki et al." [5] and “Howie et al." [6]. Both techniques are optimized for “concert" perspective sound scene reproduction, using widely spaced directional microphones to capture ambient sound, an either directional (Hamasaki) or omnidirectional and directional (Howie) microphones to capture the direct sound of the ensemble. Grew et al. [7] describe a variation on the “OCT-9” [8] 3D microphone array with three spaced directional microphones added to capture sound specifically for bottom channels. Similarly, Eaton and Lee [9] describe a variation on the “PCMA-3D" microphone array, adding three spaced directional microphones 30 cm above the floor to capture sound for the bottom layer.

A review of these sources reveals several trends for the selection and positioning of “bottom channel” microphones:

1) using microphones with directional sound pickup characteristics (e.g., cardioid or hypercardioid),
2) placing these microphones within 1 m of the floor,
3) a significant vertical spacing between the bottom layer and main layer microphones, and
4) facing these microphones towards the sound source or ensemble (except for Eaton and Lee), usually with somewhat of a downward facing angle.

Similar methods have also been observed in 3D music recording setups used by NHK and WOWOW broadcast recording engineers. The use of directional microphones for bottom channels is particularly important as it ensures a focus on capturing direct and reflected sound coming naturally from below within the sound scene: sonic information that has a logical spatial relationship with the floor-level loudspeakers that will be reproducing it.

Howie [10], and Martin and King [11] have described various concepts and considerations for recording and mixing pop and rock music for 9+10+3 reproduction. Additionally, Martin and his co-authors [12–14] have shown that complex close-microphone arrays can be used to capture and reproduce the direct sound of an instrument in a way that yields stable sonic images with perceivable horizontal and vertical extent. When these techniques are combined with ambience arrays to capture room reflections, it is possible to create pop/rock sound scenes that deliver a high degree of realism or hyper-realism, and inform the listener of the relationship between the performer and the performance space: a methodology described in detail in [10]. Most musical instruments have complex, frequency-dependent sound radiation patterns [15]. Multichannel close microphone arrays that extend both horizontally and vertically will naturally capture a more realistic impression of a given instrument’s timbre profile than what is possible when using only one or two close microphones, as is typical with stereo or 5.1 music production techniques.

3 3D Music Recording Case Studies

All recordings described herein were recorded at 96kHz/24bit resolution, by a team of recording engineers and audio researchers possessing extensive experience recording and mixing multichannel and 3D audio. Recordings were mixed for 9+10+8 reproduction in Studio B at Tokyo University of the Arts' Senju Campus: an acoustically treated studio equipped with 27 “KS Digital C5” full-range 2-way

powered studio monitors. Loudspeaker positions conform to ITU recommendations for 9+10+3 reproduction [1], with the five added bottom channels matching the horizontal angles of their corresponding main-layer loudspeakers. Additional documentation, including complete input/microphone lists, and excerpts of the referenced 3D audio recordings can be accessed at:
https://doi.org/10.5281/zenodo.7563813

3.1 Solo jazz piano

A solo jazz piano performance was recorded in Studio A at Tokyo University of the Arts' Senju Campus: floor area = 160 m^2, ceiling height = 7 m, reverb time = approx. 1.0 s at 500 Hz. A Steinway concert grand piano was placed in the north end of the room, facing to the south: the south end of the room being somewhat more acoustically resonant than the north end, with a greater density of reflections. An array of nine cardioid microphones, arranged as Left, Centre, and Right in three vertical layers, was placed near the piano to capture primarily direct sound (Figs. 2–3). Placement was designed to reproduce an image of the piano that was realistic in terms of timbre and physical extent, based on similar techniques described in [10] and [16]. The bottom channel microphones were placed specifically to capture both direct and reflected sound from underneath the piano: a perspective that has a darker, more bass-heavy tonality than what is typically captured with more traditional piano microphone placements. Microphone signals were assigned directly to their corresponding loudspeaker channels (Fig. 1), except for the “close top layer" microphones, which were panned between the top and main layers. The resultant piano image retains a natural impression in terms of size, shape, and timbre.

An array of largely spaced directional microphones was set up to capture ambient sound for the height-layer loudspeakers, and side and rear main-layer and bottom-layer loudspeakers (Figs. 2–3). Ambience microphones were generally placed at least 2 m apart to ensure a high level of decorrelation between signals [17]. This large number of ambience signals panned in three vertical layers provides the listener with a realistic and enveloping impression of the sonic characteristics of the recording venue. All microphones were routed to either ADT or RME microphone preamplifiers, then to Digital Audio Denmark (AX32) analog-to-digital converters. Microphone signals were balanced and panned to give the listener an impression of standing approximately 1.5 m in front of the piano: a close sound that retains enough depth perspective and ambience for a strong sense of space around the music. Artificial reverb (Flux IRCAM verb) was added to certain ambience channels to somewhat lengthen and smooth the recording venue's reverb tail.

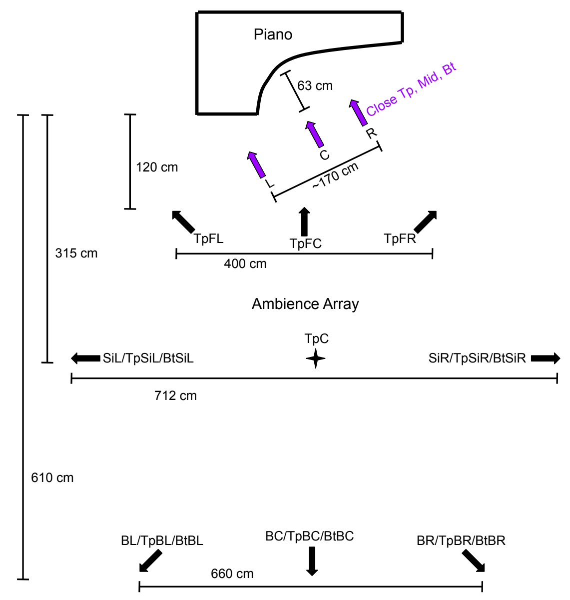

Figure 2. Piano microphone placement, overhead view.

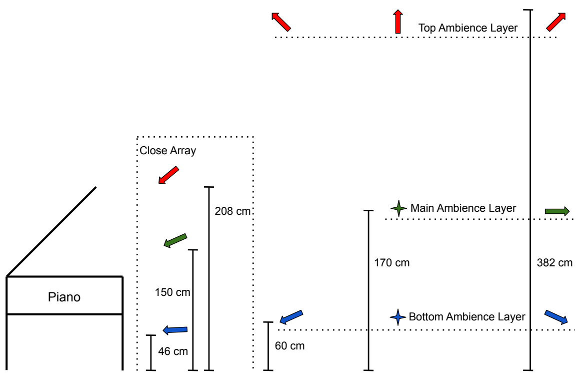

Figure 3. Piano microphone placement, side view.

Solo pipe organ in a concert hall

This example comes from a set of recording sessions for a stereo release of solo pipe organ music, recorded at Tokyo University of the Arts' Sōgakudō Concert Hall, an 1100 seat classical music venue (length = 36 m, width = 18 m, height = 15 m) with a variable reverb time that was set to approximately 2.4 s for this recording. The microphone setup used was a combination of a “Decca Tree" [18, 19] with outriggers, and a Left-Centre-Right array of “close" microphones, all optimized for stereo reproduction.

Additional microphones in both “main" and “ambient" positions were then added to create a 3D audio recording (Figs. 4–5).

The FLc, FC, and FRc omnidirectional microphones were fitted with acoustic pressure equalizers [20] for improved channel separation at higher frequencies, as in [6]. Since the pipe organ is such a physically large instrument, all of the “front" microphones captured a good deal of direct sound. When these signals are assigned to their respective loudspeakers, the resultant sonic image of the organ extends downward vertically from the top layer of loudspeakers nearly to the floor, and spans around 100 to 120º in width. Ambience microphones were generally spaced at least 2 m apart horizontally and vertically. Microphone signals were routed to a Studer Vista 5 console for preamplification and analog to digital conversion. As the organ pipes do not extend all the way to the stage, the bottom channel microphones were positioned facing upward (+45 degrees), capturing a fairly equal blend of direct and reflected sound. In this way, when combined with the main and height layer front microphones, the bottom channel signals do not “pull” or “smear” the sonic image of the organ any lower than how it would be perceived when listening in the concert hall. Due to time and equipment constraints, microphones were not placed for the following channels: TpC, TpBC, BC, BtBC.

This recording makes use of a number of omnidirectional microphones, particularly in the main layer, which were chosen primarily based on the sonic goals of the stereo recording. Directional microphones were used for all bottom channel locations, to allow for better precision in capturing specific aspects of the lower sound scene. Placement of all microphones was based largely on previous experience recording in Sōgakudō, availability of microphone hanging and patch points, and in-situ listening during rehearsal and soundcheck. Most microphone signals maintain a linear relationship between label and panning, e.g., the “FL” microphone is assigned directly to the “Front Left" loudspeaker. The exceptions are the three “close" microphones signals, which were panned 50 percent between the main and top layers. Thus, the close presence within the sound of the organ is weighted more towards the upper part of the sonic image, which corresponds to the mixing engineer's impressions from listening to the organ at various locations within the concert hall. Artificial reverb was applied to a number of ambience channels to lengthen the reverb tail (Altiverb 7: King's College), giving the mix a more cathedral-like sound that is typical for pipe organ music.

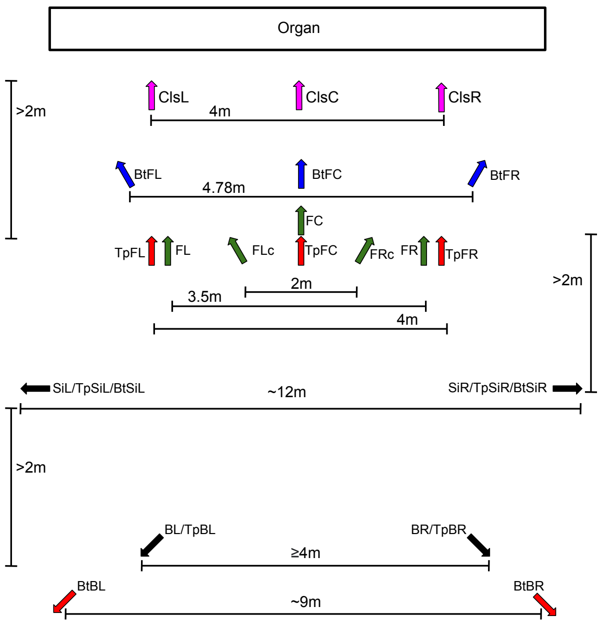

Figure 4. Organ microphone placement, overhead view.

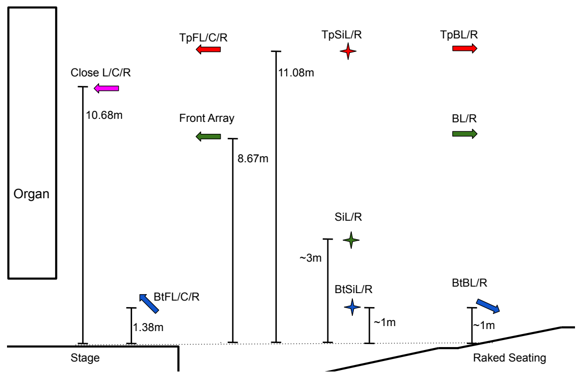

Figure 5. Organ microphone placement, side view.

3.3 Alternative rock

An alternative rock song was recorded and mixed in the above-described Studios A and B. The song features a dense musical arrangement with instruments spread around the listener in horizontal and vertical space (Figs 6–7): drums, electric bass, two electric guitars, double-tracked acoustic guitar, thirteen mono and stereo synthesizer parts (keyboard and Akai EWI 4000s), piano, tambourine, and vocal. The recording methodologies for these various instruments were developed from previous experience recording pop/rock music for 4+5+0 and

9+10+3 reproduction, as well as extensive in-situ experimentation.

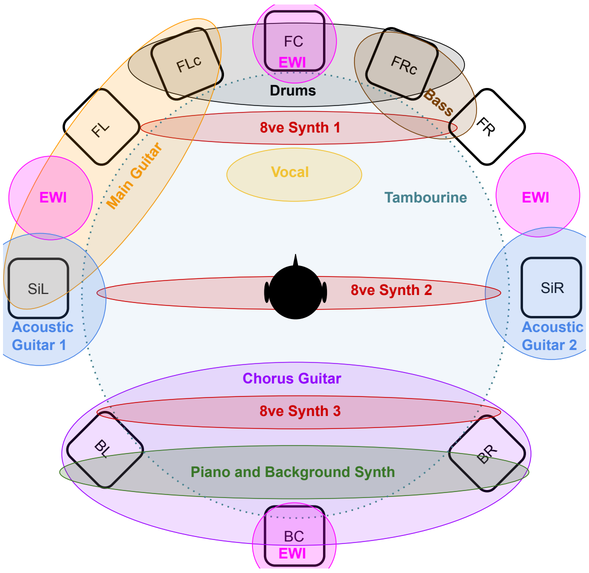

Figure 6. Rock sound scene layout, top view.

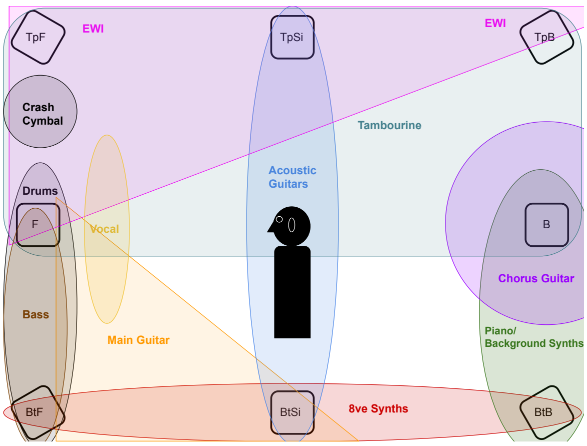

Figure 7. Rock sound scene layout, side view.

For the core parts of the musical arrangement (drums, bass, "main guitar" and vocal), the direct sound of each instrument was recorded with a set of directional microphones positioned to capture the physical extent of the instrument or amplifier's sound image, and overall timbre profile in a realistic or hyper-realistic manner:

Pairs of "high", "mid", and "low" perspective close microphones captured the basic image of the drum kit, which were then combined with other close microphones to complete the sound image. Pairs of vertically spaced close microphones were used for both the kick and snare drums, which aids in achieving more accurate vertical localization and spread for those instruments within the mix [10].

The bass amp close sound was captured with 4 directional microphones (centre, bottom, left, right), spaced ~25–35 cm apart.

The "main guitar" close image is a composite sound from two different combo-amplifiers: a Peavy Bandit placed on top of a Roland Jazz Chorus (Fig. 8). Four microphones, arranged as Top, Centre, Left and Right, were placed near the Bandit, while two more microphones were placed near the stereo loudspeakers of the Jazz Chorus. While the Bandit close microphones were panned to retain a realistic image size and focus for the guitar amp sound, the Jazz Chorus signals were panned widely to the BtSiL and BtFC loudspeakers, resulting in a guitar image that expands horizontally and vertically in a triangular shape (Figs. 6–7).

Figure 8. "Main guitar" amp and microphone setup.

The vocal was captured with a five-microphone spaced array: left, centre, right, top, bottom, with signal panning largely following those conventions. Approximately 70 cm further back from this set of microphones was a "Blumlein" pair of ribbon microphones, which were delayed and assigned to the BL and BR channels to create an immersive echo effect.

A set of widely spaced directional microphones was set up to capture ambient room sound throughout the recording sessions (Fig. 9). This microphone array remained largely unchanged regardless of which instrumental part was being recorded, though the number and physical positions of microphones varied depending on the desired level and quality of ambience for a given instrument, as well as considerations to time delays between microphone signals.

Figure 9. Rock drums, with ambience microphones.

Once the primary parts of the musical arrangement had been recorded, it was felt that there was enough sonic information within the 9+10+8 environment to give the listener a strong sense of presence. Thus, the remaining instrumental parts could often be recorded with simpler microphone setups, while still achieving desired physical, timbral, and textural characteristics. The “chorus guitar" part was recorded through the Roland Jazz Chorus, with three close microphones (BL, BC, BR) and two distant microphones. The distant microphone signals were panned to the TpFL and TpFR loudspeakers, and delayed by 240 ms, creating a vertical echo effect.

For each of the monophonic synthesizer parts, the direct signal was routed to a combo guitar amplifier, with one ribbon microphone placed near the amp, and another several meters away. When these pairs of signals are assigned to loudspeakers that face each other on a diagonal (e.g., TpFR and BtBL), a sense of depth and spaciousness within the sound scene is achieved, especially when multiple monophonic parts panned to different spatial locations are being performed at the same time.

The piano was recorded with three pairs of microphones set up successively further and higher away. Three microphones were placed close to the acoustic guitars at “bright", “neutral", and “dark" sounding positions, which were then assigned to the TpSiL/R, SiL/R, and BtSiL/R right channels respectively, creating tall, column-like guitar images. The tambourine was recorded with ambience microphones only, to create a more diffuse sound image.

Within the primary instruments of the musical arrangement, panning of microphone signals to the bottom layer was largely based on the desire for those instruments to localize in a way that reflects real-life listening. In contrast, other musical parts, such as the bass “8ve synthesizer” in the song’s chorus, were panned to the bottom layer for purely aesthetic reasons. In general, the mix engineer tried to construct a sound scene wherein the more diffuse and “airier" sounds localized above the listener, mid-range instruments localized nearer to ear level, and low frequency elements localized more towards the bottom layer (Fig. 7). This approach was inspired by personal aesthetics, but also a desire to respect the well documented “pitch-height” metaphor within human hearing [21–24].

Compared with the complexity of capturing these sound images, the mixing process was simpler. As has been described anecdotally by many other practitioners recording or mixing for large-scale 3D audio environments, dynamic range compression was found to be largely unnecessary, and was only used sparingly to even out the dynamics within the lead vocal performance. EQ, when needed, was largely subtractive, removing unwanted or masking resonances. Rather than using existing multichannel or “3D” algorithmic reverb tools, mono instances of reverb plugins would be inserted directly on ambience channels, adjusting the wet/dry mix, balance of early and late reflections, timbral properties, and reverb time of each channel until a desirable multichannel ambience was achieved. This method gave the mixing engineer greater control over the spatial aspects of ambience within the mix than what they had found to be possible with available 3D reverb plugins.

3.4 Taiko drum ensemble

A recording was made of a taiko drum ensemble (one ōdaiko and two shime-daiko) in the above-described Studio A. To ensure focused direct sound images of the drums, the performers were positioned in the least acoustically resonant area of the room, with large sound absorbing baffles placed approximately 2 m behind each of the instruments (Fig. 10). For this recording, a simple one microphone per loudspeaker setup was used. The direct sound of the three instruments was captured by a set of eight microphones, corresponding to the FL, FLc, FC, FRc, FR, BtFL, BtFC, and BtFR channels. For the FLc, FC, and FRc channels, omnidirectional microphones were used to ensure a linear capture of low frequencies from the ōdaiko. Placement of the bottom channel microphones focused on capturing the sound coming off the bottom of the three drums, as well as early reflected sound from the floor. Careful balancing of the main and bottom layer close microphones results in a sonic image of the drums that accurately reflects both the horizontal and vertical locations of the instruments in the room. Two omnidirectional microphones were placed on the floor in areas that were deemed to contain rich low frequency content: their signals were low-passed at 80Hz, then used as Left and Right LFE channels.

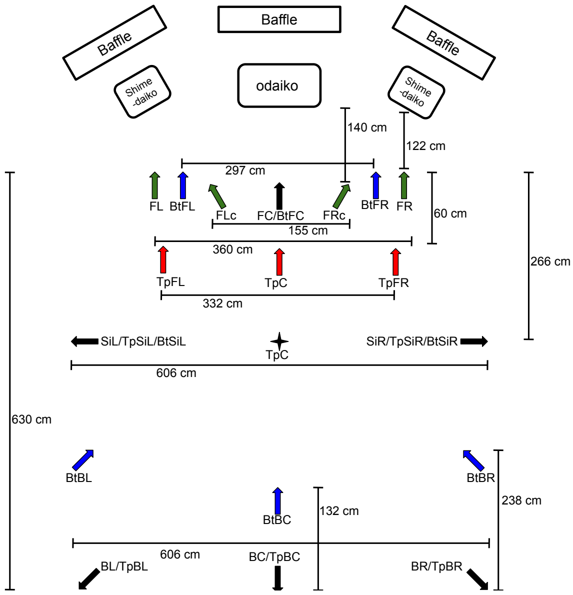

Figure 10. Taiko microphones, overhead view. Height of layers: 71–87 cm (Bottom), 157–170 cm (Main), 367 cm.

Most of the ambience microphone horizontal positions remained relatively unchanged between vertical layers (Fig. 10). The primary exception is the three rear bottom microphones: the recording team found that when placed at the same distance as the main and top-layer microphones, the rear bottom microphones contributed an unpleasant ambience to the recording that blurred the impact of the large ōdaiko drum, particularly in the low frequencies. By moving those microphones closer to the instruments and facing them towards the ōdaiko, a better balance between direct and diffuse sound was achieved in the lower layer. Microphone input routing followed the same signal path as for "Solo jazz piano". Once a satisfactory balance had been achieved between all microphone signals, low or low-mid frequency filtering was applied to certain ambience channels to increase focus and definition within the transients of the drums. Artificial reverb (LiquidSonics' "Seventh Heaven") was applied sparingly to a number of ambience channels to lengthen and smooth the reverb tail in a way that better matched with the musical intentions of the performers.

4 Discussion

4.1 Considerations for music production with bottom channels

Stereo audio reproduction is a format that, when compared with real-life listening, suffers from a great deal of spatial, tonal, and timbral compression of sounds and sound scenes. Equalization, dynamic range reduction, and level automation are tools commonly used by engineers to help "fit" the various sounds of a group of instruments into the limited spatial environment of the stereo field. 3D audio formats, particularly large-scale formats such as 9+10+3 and 9+10+8, offer vastly more spatial resolution than stereo or 5.1 "surround sound," allowing engineers more options and choices in terms of building individual sound images, ambient sound fields, and complete musical sound scenes. When the sonic images of various musical instruments can be constructed (or re-constructed) and positioned in physical space in a way that gives an impression approximating real-life (or beyond) listening, the use of EQ and dynamic range compression becomes far less important than in stereo production, allowing the engineer to use those tools for more artistic or aesthetic reasons. This was found to be particularly true for the Alternative Rock case study, where the spatial positioning of "background" or "secondary" instrumental parts helped inform their role within the arrangement and mix without having to reduce their level, tonal presence, or position within the mix's depth perspective. Also, the final mix required relatively little level-automation of the various instrumental parts, which is unusual for a rock recording with a dense arrangement.

While examining the perceptual influence of floor level-loudspeakers, Grewe et al. [7] found "that the addition of the lower layer loudspeakers is particularly recognizable and beneficial for content containing floor-related signals, such as footsteps." This is interesting to consider in tandem with Eaton and Lee's [9] experiment where subjects were asked to balance three vertical playback layers of four different recordings made for 9+10+3 reproduction.

For the "orchestra" example, subjects generally balanced the bottom layer signals higher than for other recordings under test. The microphone signals assigned to the orchestra recording's bottom channels contained a mixture of direct sound from the strings and early floor reflections, which Eaton and Lee hypothesize "might have been found [by subjects] to be more useful to create a more realistic sound field in the 9+10+3 reproduction" [9]. These conclusions are in line with the authors' experience creating the recordings described herein. The presence of direct and early reflected sound in the bottom layer seems to contribute to a greater sense of presence and realism within the sound scene, whereas overly diffuse or reverberant bottom channel information, even in the rear channels, can at times contribute to a general lack of clarity.

Many 3D audio content creators, the authors included, have observed that when working with height channels, pushing the level of those signals too strongly in the mix can result in an upward "smearing" of the sound stage. Mixing with bottom channels allows for the anchoring of sonic images at the floor-level, which can aid in creating a more "vertically natural" sound scene. Also, sound signals reproduced from floor-level loudspeakers, especially those with strong low-frequency content, avoid the problem of phase interference that can occur between sounds being reproduced from head-height loudspeakers and their respective floor reflections.

4.2 Evaluation of recordings

Several informal evaluations of these recordings have taken place in 3D audio production studios at Tokyo University of the Arts, Kyushu University, and McGill University. Listeners have included professional recording engineers, researchers in the field of 3D audio, and graduate students in audio production. Generally, the recordings have been well received, particularly the Alternative Rock and Taiko recordings, both of which seem to impart a very strong sense of "being there" to the listeners. One professional listener said this of the Alternative Rock recording: "The moment the first drum break hits, you know exactly where you are, exactly what the room is like." It may be that effective use of bottom channels within 3D audio reproduction can contribute to a greater ease of what Baldwin calls "auditory space perception" – "The ability to develop a spatial representation of the world (or the area one is experiencing) based on auditory information." [25]. Most listeners also commented on the clear extension of sound images to the floor, particularly for the Alternative Rock recording, and that this seemed to give a greater sense of spatial and tonal realism to the sound scenes.

Two of the more experienced professional listeners felt that the use of many omni-directional microphones for ambience capture within the Organ recording resulted in a somewhat confusing, unfocused sound scene. They speculated that even though these microphones were spaced quite far apart from each other, they still contained too much shared information in terms of reflected energy, and that directional microphones likely would have been a better choice.

Excerpts of the recordings described herein are being used as a part of a new study examining the perceptual effects of bottom channels within 3D audio reproduction across various musical and non-musical sound scenes, including examining reproduction conditions where the bottom-layer signals are folded into the main-layer loudspeakers. The results of this study will be described in a forthcoming paper.

5 Summary

This report describes the production of four musical sound scenes for a 9+10+8 advanced 3D audio reproduction system: solo jazz piano, pipe organ in a concert hall, alternative rock, and taiko drums. The immersive audio production techniques described herein are based on extensive previous experience recording for multichannel and immersive audio systems, and previous research in 3D audio content creation. It is hoped that this report can help serve as a guide to other content creators producing work for immersive audio systems including bottom-layer sound reproduction. There recordings will be made available for use by other researchers upon request. Additional documentation and audio examples can be accessed at: https://doi.org/10.5281/zenodo.7563813

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science and Tokyo University of the Arts. Thank you to assistant engineers Megu Tokunaga, Tsubasa Ikeda, and Moe Mukaitani, and to performers Shimon Ohno, Shunsuke Omori, Rie Hiroe, Sam Davidson, Masaharu Nagano, Hitomi Nagano, Yui Komura.

References

[1] International Telecommunications Union, "Advanced sound system for programme production," (2022).

[2] A. Raginska and P. Geluso, *Immersive Sound: The Art and Science of Binaural and Multichannel Audio*, Routledge (2018).

[3] H. Lee, "Multichannel 3D Microphone Arrays: A Review," *J. Audio Eng. Soc.*, vol. 69, no. 1/2, pp. 5–26 (2021).

[4] K. Hamasaki and K. Hiyama, "Development of a 22.2 Multichannel Sound System," *Broadcast Technology*, vol. 25, no. Winter, pp. 9–13 (2006).

[5] K. Hamasaki, T. Nishiguchi, K. Hiyama, and K. Ono, "Advanced Multichannel Audio Systems With Superior Impression of Presence and Reality," presented at the *116th Convention of the Audio Engineering Society* (2004).

[6] W. Howie, R. King, and D. Martin, "A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound," presented at the *141st Convention of the Audio Engineering Society* (2016)

[7] Y. Grewe, A. Walther, and J. Klapp, "Evaluation on the Perceptual Influence of Floor Level Loudspeakers for Immersive Audio Reproduction," presented at the *147th Convention of the Audio Engineering Society* (2019).

[8] G. Theile and H. Wittek, "Principals in Surround Recording with Height (v2.01)," presented at the *130th Convention of the Audio Engineering Society* (2011).

[9] C. Eaton and H. Lee, "Subjective evaluations of three-dimensional, surround and stereo loudspeaker reproductions using classical music recordings," *Acoust. Sci. Tech.*, vol. 42, no. 2, pp. 149–161 (2022).

[10] W. Howie, "Pop and Rock music audio production for 22.2 Multichannel Sound: A Case Study," presented at the *146th Convention of the Audio Engineering Society* (2019).

[11] B. Martin and R. King, "Three-Dimensional Spatial Techniques in 22.2 Multi-channel Surround Sound for Popular Music Mixing," presented at the *139th Convention of the Audio Engineering Society* (2015).

[12] B. Martin, R. King, and W. Woszczyk, "Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three- Dimensional Capture of Acoustic Instruments, Part I," presented at the *141st Convention of the Audio Engineering Society* (2016)

[13] B. Martin, J. Kelly, and B. Leonard, "Subjective Assessment of the Versatility of Three-Dimensional Near-Field Microphone Arrays for Vertical and Three-Dimensional Imaging," presented at the *147th Convention of the Audio Engineering Society* (2019).

[14] B. Martin, D. Martin, R. King, and W. Woszczyk, "Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments, Part II," presented at the *147th Convention of the Audio Engineering Society* (2019).

[15] J. Meyer, *Acoustics and the Performance of Music* Springer (2009).

[16] W. Howie, D. Martin, and R. King, "Listener Discrimination between Common Channel-Based 3D Audio Reproduction Formats," *J. Audio Eng. Soc.*, vol. 65, no. 10, pp. 796–805 (2017).

[17] K. Hamasaki and K. Hiyama, "Reproducing Spatial Impression With Multichannel Audio," in *AES 24th International Conference: Multichannel Audio, The New Reality* (2003).

[18] M. Gray, "The Decca Sound: Secrets Of The Engineers," Polymath Perspective, 13 June 2012. [Online]. Available: http://www.polymathperspective.com/?p=2484. [Accessed 4 January 2023].

[19] F. Rumsey, *Spatial Audio*, Focal Press (2013)

[20] W. Woszczyk, "Acoustic Pressure Equalizers," *Pro Audio Forum*, pp. 1-24 (1990).

[21] C. Pratt, "The Spatial Character of High and Low Tones," *Journal of Experimental Psychology*, vol. 13, no. 3, pp. 278–285 (1930).

[22] S. Roffler and R. Butler, "Localization of tonal stimuli in the vertical plane," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1260–1266 (1968).

[23] S. Roffler and R. Butler, "Factors That Influence the Localization of Sound in the Vertical Plane," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1255– 1259 (1968).

[24] J. Blauert, *Spatial Hearing: The Psychoacoustics of Human Sound* Localization, MIT Press (1997).

[25] C. L. Baldwin, *Auditory Cognition and Human Performance* (CRC Press, New York, USA, 2012).