Audio Engineering Society

Convention Paper

Presented at the 146th Convention
2019 March 20–23, Dublin, Ireland

Pop and Rock music audio production for 22.2 Multichannel Sound: A Case Study

Will Howie

CBC/Radio-Canada, Vancouver, British Columbia, Canada V6B 4A2

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

ABSTRACT

Advanced sound capture and mixing techniques, optimized for high channel-count three-dimensional audio reproduction systems, are discussed for pop/rock music production. Based on previous research and experimental recordings, newly developed complex close-microphone arrays are designed to deliver realistic sonic images of musical instruments in terms of physical size and timbre. Combined with multiple ambience microphones, these direct sound arrays can be used to create highly realistic or hyper-realistic sound scenes for 22.2 Multichannel Sound (9+10+3) reproduction, or other 3D audio formats. A specific case study highlights the aesthetic and technical considerations for production of pop/rock music for advanced audio formats such as 22.2 Multichannel Sound.

1 Introduction

22.2 Multichannel Sound

Japan Broadcasting Corp. (NHK) has developed and standardized 22.2 Multichannel Sound (22.2) as the immersive audio component for their ultra-high definition 8K resolution video broadcast format, Super Hi-Vision [1]. 24 discreet loudspeaker channels are arranged in three layers: ear level, height level, and floor level (Figure 1). Alternately known as "9+10+3", 22.2 has been further standardized by SMPTE [2] and the International Telecommunications Union (ITU) [3].

Published research on 22.2 has covered a wide range of topics, including spatial impression and listener experience [4, 5], audio production techniques [6, 7], construction of 22.2 production facilities [8], flexibility of loudspeaker azimuth and altitude [9], and transmission concerns, such as downmixing to reduced playback formats [10] and the subjective loudness of 22.2 program material [11].

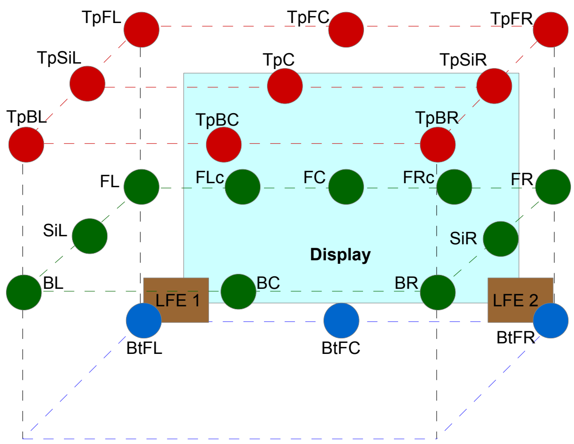

Figure 1. 22.2 Multichannel Sound speaker layout: main layer in green, height layer in red, bottom layer in blue.

Recording and mixing music for 22.2

Previous research [5, 12, 13] has shown that when musical audio content is created using recording and mixing techniques optimized for 3D audio reproduction, the resultant sound scenes are capable of delivering listeners strong impressions of presence, reality, depth, envelopment, and naturalness. Among common, currently standardized channel-based 3D audio formats, 22.2 has been shown to deliver a perceptually unique experience in terms of reproduction of acoustic music [14]. The formats' wide, 120° frontal sound reproduction angle and floor-level "bottom" channels allow for the creation of sound scenes with both realistic horizontal and vertical extent [15]. The numerous height and surrounding loudspeakers have an even spatial distribution (Figure 1), allowing for a highly immersive reproduction of a given acoustic space.

Howie et al. [15] and Hamasaki et al. [4] have discussed techniques for recording large ensemble classical music for 22.2 reproduction. Both proposed techniques that combine separate microphone arrays designed to prioritize direct sound or ambience. Such a method allows for the capture of a complete sound scene while maintaining image control at the mix stage. Beyond this, most published work addressing audio production for 22.2 has focused on non-musical sound capture [6, 7]. A notable exception is Martin and King's [16, 17] work detailing aesthetic, technical, and artificial acoustic considerations when mixing for 22.2 from traditional stereo multitrack recordings. The process of mixing 3D sound scenes from 1D material is shown to be problematic in terms of labour, DSP usage, available 3D mixing tools, and sound scene stability [16].

Optimizing close microphone techniques for 22.2

Traditional close-mic'ing techniques utilized in pop/rock recording are optimized for mono or stereo sound reproduction, typically using only one or two microphones per instrument. Stereo sound reproduction suffers from a great deal of spatial compression as compared with natural 360° listening. As such, a single microphone will often capture more than enough information to create a sonic image appropriately sized for the medium. 3D sound reproduction, by comparison, is spatially vast. As Martin and King [17] show, recorded sounds that deliver appropriately sized sonic images within stereo reproduction typically yield images that seem small and unimpactful within a dense immersive reproduction environment such as 22.2. Mixing engineers are then required to manufacture sonic images more appropriate to the reproduction environment using tools such as track duplication, 3D panning, delays, and algorithmic or convolution reverb engines. The resultant images are often unstable and lacking cohesion; a more ideal approach would be to capture sonic images appropriate to the reproduction format at the recording stage.

The complexity of tonal and directional characteristics of most musical instruments is well documented by Meyer [18]. When close-mic'ing instruments with frequency-dependent radiation patterns using one or two microphones, there is a natural trade-off between desired proximity and perspective, and completeness of the instrument's captured tonal and timbral characteristics. The spatial density and multi-layer positioning of loudspeakers within the 22.2 format would naturally allow for the design of complex microphone arrays that maintain the close perspective associated with pop/rock recordings while capturing a more complete sonic impression of a given musical instrument. Martin [19, 20] has shown that such an approach can also be used to create sonic images with strongly perceivable horizontal *and* vertical extent. Combined with large-scale ambience arrays of the type described in [4], [12] and [15], these complex close microphone arrays can be used to created highly realistic or hyper-realistic musical sound scenes optimized to exploit the unique sound reproduction capabilities of 3D audio formats such as 22.2. Such an approach, as applied to jazz and new music performance recording, is described briefly in [14].

2 General philosophy and workflow for optimized sound capture for 22.2

Direct sound capture

Following from the above considerations, as well as numerous experimental recording sessions, a general approach to close-proximity instrument capture has been developed, optimized for 22.2, but applicable to other channel-based or channel-agnostic formats. The basic workflow for designing direct sound microphones systems is as follows:

1) For a given instrument, estimate the minimum amount of captured sonic information required to yield the desired playback image.

2) Estimate how many microphone signals are necessary to represent that information and decide on a rough physical placement scheme.
3) Place microphones near a given instrument, by ear. Try to capture a complete picture of the instrument's timbral and tonal range, while maintaining complementary and non-destructive relationships between microphone signals.
4) Based on the physical placement of microphones and the desired reproduced sonic image, pan the microphone signals using a digital audio workstation or console equipped with a 3D panner.
5) Fine-tune the balance and pan of microphone signals until the sonic image is cohesive and stable.

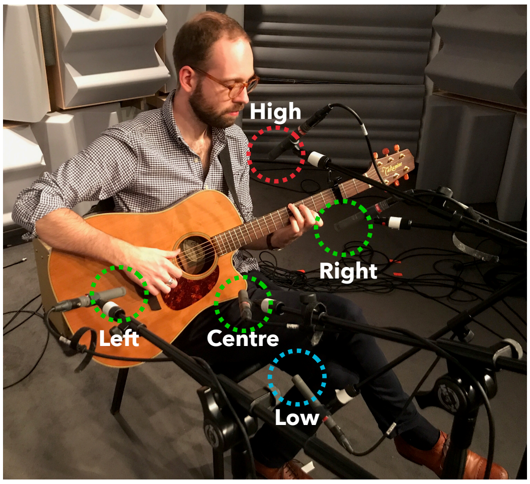

Figure 2. Acoustic guitar with close microphones.

As an example, Figure 2 shows the recording of an acoustic guitar in a hemi-anechoic environment. Five microphones have been placed near the instrument: centre, left, right, high, and low. This simple 5-point array captures enough information to reproduce a sonic image with realistic or hyper-realistic horizontal and vertical extent, and a natural tonal and timbral range. Panning and balance of signals depends largely on how large or small of an image is desired by the mix engineer. Typically, the goal is to create images whose physical stability within the sound scene is sweet-spot agnostic: this often involves "anchoring" one or more microphone signals to loudspeakers. Figures 3 and 4 show two examples of how the guitar image could range from tight and cohesive, to very large and spread apart, depending on microphone signal panning.

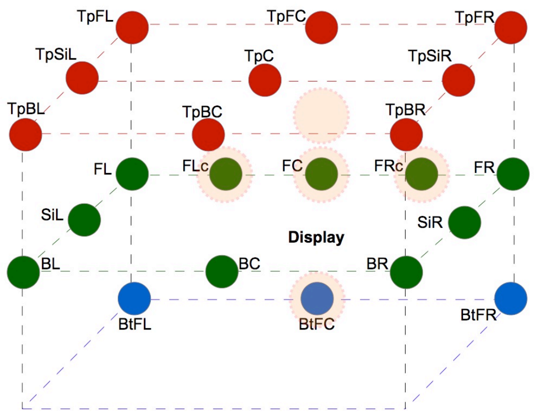

Figure 3. Example of signal panning that could yield a "realistic-sized" guitar image. Microphone signals are shown in orange.

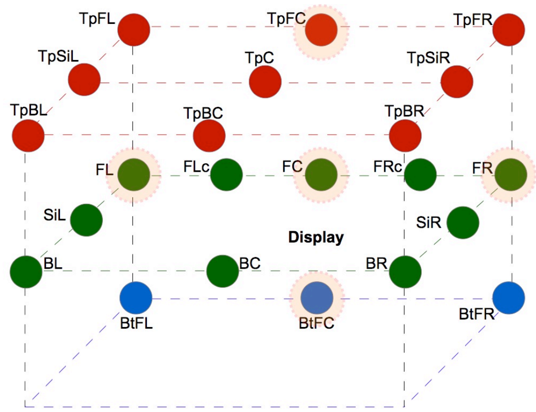

Figure 4. Example of signal panning that could yield an "unrealistically large" guitar image. Microphone signals are shown in orange.

Ambient sound capture

For each instrument or set of instruments performing in an acoustic space, ambience microphones are also used. Natural ambience captured during the tracking process conveys to the listener a sense of the relationship between performer and performance space, and gives a more complete picture of how the instrument sounded in the room. Microphone spacing that prioritizes a high degree of decorrelation between ambience signals will result in strong levels of listener envelopment [21, 22]. The number of ambience microphones necessary to give the listener a good sense of space will depend on the room and instrument.

3 Case Study: Recording

Having already explored acoustic music recording for 22.2 and other 3D audio formats [12, 14, 15, 21], a large-scale production of a pop/rock song for 22.2 was undertaken. The goal was to adapt previously developed 3D recording techniques to production aesthetics and environments more typical of pop and rock music, e.g. dryer acoustics, acoustic isolation of instruments, amplified electric instruments and synthesizers, and open-ended production workflows wherein parts are often recorded at different times in different locations. All recording and mixing was done at 96kHz/24bit resolution, using Merging Technologies' Pyramix 10 digital audio workstation.

Composition to be recorded

"Working", the composition chosen to be recorded, is a pop/rock song approximately 3 min in duration, written and arranged in the style of early 1970s R&B singer-songwriters such as Bill Withers or Bobby Womack. The instrumentation is: drums, electric bass, acoustic guitar, electric guitar (clean), Wurlitzer 206s electric piano, lead and background vocals, electric guitar solo (overdriven), and synthesizer (only for the outro section). For access to the final 24ch mix of "Working" for research or evaluation purposes, please contact the author.

Preproduction

Based on the availability and location of desired musicians, two recording venues were chosen: 1) *Humber College Recording Studio* in Toronto, Canada for drum tracking; 2) *Studio Space*, an unfinished recording studio at McGill University, Canada for recording the remaining instrumental and vocal parts. Though this decision was made for mainly practical reasons, it also reflects the reality of many contemporary pop/rock music productions, wherein the use of a professional recording studio is reserved for drum or rhythm section tracking, with overdubs being done wherever is most affordable or practical.

The most important decision to be made prior to recording was the desired spatial arrangement of instruments within the 22.2 reproduction environment (Figure 5). This information would be necessary throughout the recording in order to design and implement appropriate microphone arrays. The aesthetic concept for "Working" is primarily recreative; the layout is based on how such a group of musicians and instruments would be positioned for a live performance. This arrangement is also based on desired vertical imaging of various musical instruments: for the drums, bass, and keyboard and guitar amplifiers, the use of the bottom channels was considered to be necessary to vertically extend their respective sonic images to the floor.

Many years of experience in mixing 2D and 3D multichannel audio has led to an understanding that strongly active or obtrusive direct sound from behind the listener can contribute to feelings of pressure, disorientation, and discomfort. Similar observations are made by Hinata et al. in [6]. For "Working," the rear sound stage was primarily reserved for background vocals and ambience. It was also felt important to include at least one element that was distinctively "3D" or hyper-realistic in the mix: thus the 360° immersive synthesizer pads for the song's outro and fade-out section.

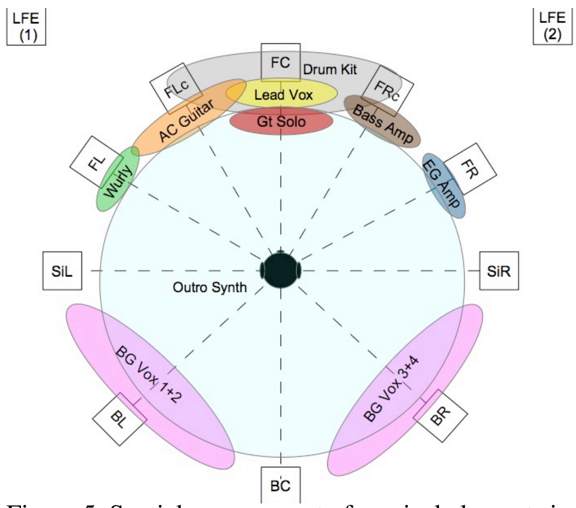

Figure 5. Spatial arrangement of musical elements in "Working," as seen from above.

Drum recording: Humber College Studios

Production of "Working" began at Humber College Studios, where drums were recorded. This professional recording studio has a large tracking space, measuring 10.6 x 7.9 x 4.9 m, with an average RT60 of approximately 0.8 s. In keeping with the spirit of the song's early 1970s compositional and arrangement style, the studio's variable acoustics were adjusted to decrease reverberation as much as possible. A unique challenge of recording at Humber Studios was a lack of dedicated 3D monitoring in the control room, which is setup for 5.1 surround sound. While 22.2 monitoring was not possible, a practical solution was reached: two loudspeakers were added at floor level below the main left and right monitors. Two height loudspeakers were also added, positioned above the main left and right monitors. Combined with the existing surround loudspeakers, this ad-hoc arrangement allowed the engineer to pan microphone signals to spatial positions that approximated those that would be possible within the final mix environment. This solution proved to be adequate for placement and optimization of microphones.

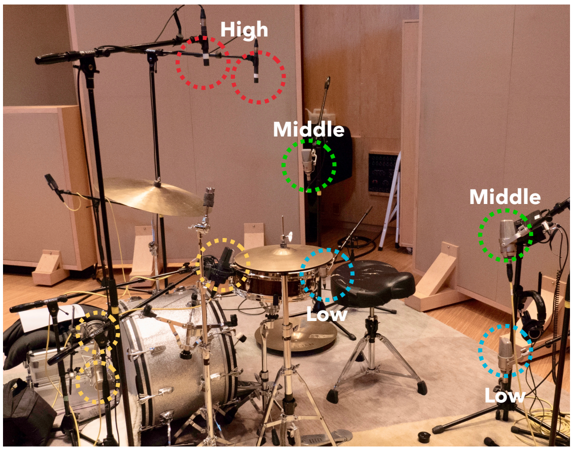

Figure 6. Close drum microphones. Red, green, and blue circles correspond to minimal hemisphere. Yellow circles demarcate kick and snare spot microphones.

Figures 6 and 7, and Table 1 show close and ambience drum microphone choice and placement. The close microphone arrangement is based on a minimal hemisphere of pairs of high, middle, and low microphones, panned accordingly. Each of these pairs is positioned equidistant to the centre "kick/snare" axis of the drum kit. High and low kick drum, and top and shell snare drum microphones aid in building a drum image wherein in the kick is positioned near the ground, the snare somewhat higher, and the cymbals somewhat higher still.

Figure 7. Complete view of tracking space, with drum ambience microphones.

Instrument	Microphone	Polar Pattern
Kick Drum Low	AT 4047	Cardioid
Kick Drum High	Neumann D-01	Cardioid
Snare Top	AKG C414 XLS	Bi-directional
Snare Shell	AKG C414 XLS	Bi-directional
Overhead Left	Schoeps MK 4	Cardioid
Overhead Right	Schoeps MK 4	Cardioid
Mid-level Left	Neumn. TML 170	Cardioid
Mid-level Right	Neumn. TML 170	Cardioid
Low-level Left	Neumn. TML 170	Cardioid
Low-level Left	Neumn. TML 170	Cardioid
Ambience SiL	Senn. MKH30	Bi-directional
Ambience SiR	Senn. MKH30	Bi-directional
Ambience BL	Neumn. KM 184	Cardioid
Ambience BR	Neumn. KM 184	Cardioid
Ambience BC	Neumn. KM 184	Cardioid
Ambience TpFL	Senn. MK40	Cardioid
Ambience TpFR	Senn. MK40	Cardioid
Ambience TpBL	Senn. MK40	Cardioid
Ambience TpBR	Senn. MK40	Cardioid

Table 1. Drum Microphones. Preamps were SSL Duality, through SSL X Logic A/D converters.

Instrument and vocal recording: Studio Space

The remaining instrumental and vocal parts were all recorded at McGill University's Studio Space. At the time of tracking, this unfinished recording studio was being used primarily for storage, and thus contained an eclectic mix acoustic surfaces. The room measures 11 x 7 x 5.7 m: RT60 averages just under 1 s. Monitoring took place in the adjacent Studio 22 (see Section 4 for details).

All the instrumental and vocal parts were performed by the composer, each recorded separately. In general, the 5-point method (Figure 2) described in

Section 2, or optimized variations of said method were used to record each instrument or loudspeaker cabinet (Figure 8, Table 2). For the lead vocal, a more compact, near-coincident array was used (Figure 9), as it was found that spacing between microphones greater than 35–40 cm contributed to a lack of image cohesion for the voice. The number and position of ambience microphones, per musical part recorded, was determined primarily by the desired level of ambience in the mix, and the ratio/amount of room acoustic that was deemed necessary for a particular instrument to "feel in a space" to the listener.

Figure 8. "5-Point" technique on guitar amplifier.

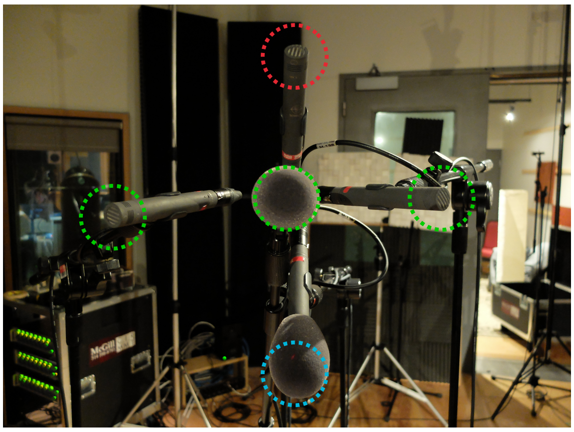

Figure 9. Near-coincident vocal microphones.

To create a hyper-realistic, enveloping synthesizer image for the composition's outro section, the instrument's output was split to 6 different loudspeakers and combo guitar amplifiers positioned at different points within the tracking space. Each loudspeaker or amplifier was facing a different direction in the room (Figure 10). A mixture of direct and ambient sound was captured using a 3-layer spaced array of 19 microphones, resulting in a highly immersive synthesizer sonic image whose tone and texture vary with listener position within the reproduction environment.

Prior to the scheduled recording of the electric guitar solo, Studio Space was closed for renovations. Therefore, the part was recorded in a home studio, with a single dynamic microphone placed near the centre of the speaker cabinet of a Fender Blues Junior guitar amplifier. Section 4 will discuss how a 3D image was constructed from this mono source.

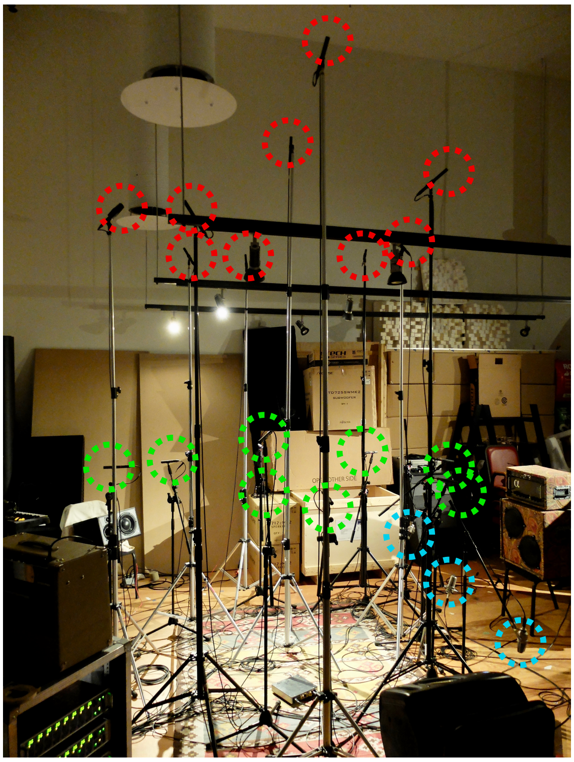

Figure 10. 19-channel, 3-layer microphone array for "immersive" synthesizer.

Instrument	Microphone	Polar Pattern
Electric Bass Amp		
High	AT 4047	Cardioid
Middle	Neumn. TML 170	Cardioid
Low	Neumn. TML 170	Cardioid
Ambience SiL	AKG C414 XLS	Bi-directional
Ambience SiR	AKG C414 XLS	Bi-directional
Ambience BC	Schoeps MK 4	Cardioid
Acoustic Guitar		
Centre	Schoeps MK 4	Cardioid
High	Schoeps MK 4	Cardioid
Low	Schoeps MK 4	Cardioid
Left	Schoeps MK 4	Cardioid
Right	Schoeps MK 4	Cardioid
Ambience SiL	AKG C414 XLS	Bi-directional
Ambience SiR	AKG C414 XLS	Bi-directional
Ambience BL	Schoeps MK 4	Cardioid
Ambience BC	Schoeps MK 4	Cardioid
Ambience BR	Schoeps MK 4	Cardioid
Ambience TpFL	DPA 4011	Cardioid
Ambience TpFR	DPA 4011	Cardioid
Ambience TpBL	DPA 4011	Cardioid
Ambience TpBR	DPA 4011	Cardioid
E. Guitar Amp		
Centre	Schoeps MK 4	Cardioid
High	Schoeps MK 4	Cardioid
Low	Schoeps MK 4	Cardioid
Left	Schoeps MK 4	Cardioid
Right	Schoeps MK 4	Cardioid
Ambience BL	Schoeps MK 4	Cardioid
Ambience BC	Schoeps MK 4	Cardioid
Ambience BR	Schoeps MK 4	Cardioid
Ambience TpFL	DPA 4011	Cardioid
Ambience TpFR	DPA 4011	Cardioid
Ambience TpBL	DPA 4011	Cardioid
Ambience TpBR	DPA 4011	Cardioid
Wurlitzer Amp		
Centre	Schoeps MK 4	Cardioid
Low	Schoeps MK 4	Cardioid
Left	Schoeps MK 4	Cardioid
Right	Schoeps MK 4	Cardioid
Ambience BL	Schoeps MK 4	Cardioid
Ambience BC	Schoeps MK 4	Cardioid
Ambience BR	Schoeps MK 4	Cardioid
Ambience TpFL	DPA 4011	Cardioid
Ambience TpSiR	DPA 4011	Cardioid
Ambience TpBC	DPA 4011	Cardioid
Vocal Ld. + Bkg.		
Centre	Schoeps MK 4	Cardioid
High	Schoeps MK 4	Cardioid
Low	Schoeps MK 4	Cardioid
Left	Schoeps MK 4	Cardioid
Right	Schoeps MK 4	Cardioid
Ambience SiL	AKG C414 XLS	Bi-directional
Ambience SiR	AKG C414 XLS	Bi-directional
Ambience BL	Schoeps MK 4	Cardioid
Ambience BC	Schoeps MK 4	Cardioid
Ambience BR	Schoeps MK 4	Cardioid
Ambience TpFL	DPA 4011	Cardioid
Ambience TpFR	DPA 4011	Cardioid
Ambience TpFC	AKG C414 XLS	Cardioid
Ambience TpSiL	DPA 4011	Cardioid
Ambience TpSiR	DPA 4011	Cardioid
Ambience TpBL	DPA 4011	Cardioid
Ambience TpBR	DPA 4011	Cardioid
Ambience TpBC	AKG C414 XLS	Cardioid
E. Guitar Solo		
Centre	Shure SM57	Cardioid
Outro Synth		
FL	Schoeps MK 4	Cardioid
FR	Schoeps MK 4	Cardioid
FC	Schoeps MK 4	Cardioid
SiL	Schoeps MK 4	Cardioid
SiR	Schoeps MK 4	Cardioid
BL	Schoeps MK 4	Cardioid
BR	Schoeps MK 4	Cardioid
BC	Schoeps MK 4	Cardioid
TpFL	DPA 4011	Cardioid
TpFR	DPA 4011	Cardioid
TpFC	AKG C414 XLS	Cardioid
TpC	Neumn. KM140	Cardioid
TpSiL	DPA 4011	Cardioid
TpSiR	DPA 4011	Cardioid
TpBL	DPA 4011	Cardioid
TpBR	DPA 4011	Cardioid
TpBC	AKG C414 XLS	Cardioid
BtFC	Neumn. TML 170	Cardioid
BtFL	Neumn. TML 170	Cardioid
BtFR	Neumn. TML 170	Cardioid

Table 2. Rhythm Section and Vocal Microphones. Channel abbreviations are as per Figure 1 and [3]. Preamps and A/D converters were RME Micstacy.

4 Case Study: Mixing

The final mix was constructed from 185 audio tracks spread across three separate Pyramix sessions, owing to track-count limitations at 96kHz. In each session, a different section of the song was mixed, with all three sections being edited together to create the final 24ch master.

Monitoring environment

All mixing took place in McGill University's Studio 22. The acoustically treated studio is equipped with 28 full-range, two-way loudspeakers (ME Geithain *M-25*) powered by Flying Mole class D amplifiers, and an Eclipse TD725SWMK2 stereo sub-woofer. The loudspeakers are arranged for reproduction of both 9+10+3 and 4+5+0, as per [3].

Direct sound images

Through the production of this case study, as well as several other 3D music recordings, it has been observed that when working with the large sound stage afforded by 22.2, very little "mixing" is required, in the traditional sense. The focus with this and other 3D music recordings has been on fine-tuning the balance and pan of the various microphone signals that make up each direct sound image. It is important, at this stage, for the mixer to leave the sweet spot as often as possible: direct sound images should ideally remain stable in terms of spatial location and physical extent, regardless of the listener's location.

When each instrument is given its own physical space to exist (Figure 5), tools such as compression and EQ appear to be less relevant. For the mix of "Working," no compression was used. Compressing microphone signals within a multi-microphone direct sound image can be problematic, as the image may distort as the compressors engage. Because a great deal of time and care was given to microphone selection and placement per instrument, EQ was also largely unnecessary. The main exception was the vocal image: the closer spacing of the vocal microphones (as compared with the other instruments recorded) caused some audible mid-frequency resonances.

In general, what has been found through this and other recordings made using similar microphone techniques, is that if the direct sound images are well captured, well balanced relative to each other, and afforded adequate physical space within the reproduction environment, the natural dynamics within the performances need not be adjusted. Again, the primary exception within "Working" was the lead vocal, whose microphone signals required fairly extensive level automation.

Ambience

For most instruments, captured ambience signals were left "as is", and simply panned and balanced. Staying within the relatively "dry" aesthetic associated with the composition's production influences, ambience signals were not overly prominent in the mix. For the drums and electric guitar, some ambience channels were modified using the convolution reverb plugin "Altiverb". Two of the drum ambience channels (TpBL and TpBR) were convolved through an EMT plate impulse response (IR). With the wet/dry mix set fairly low, this gave a slightly extended, "shimmering" effect to the drum room. For the electric guitar, several of the side and rear ambience channels were convolved through a large studio chamber IR to extend the room's reverb tail and give it a "richer" sound. Occasionally, other instrumental or vocal ambience channels were delayed slightly, typically under 10 ms, which helped make ambience more pronounced.

Building the guitar solo image

To build an appropriately sized direct sound image from the mono guitar microphone signal, the track was duplicated four times. These five tracks were then panned and EQ'd in such a way as to attempt to simulate what would have been captured from the guitar amplifier from five different, spaced microphones (Figure 11). Introducing a slight delay between the "Left" and "Right" signals gave the guitar image a horizontal broadening. The original mono signal was then duplicated an additional nine times: Altiverb was applied to each of these tracks. A large room was selected that contained a number of different IRs from different microphone positions. IRs applied to the guitar signals could then be based on corresponding physical panning of signals. For example, the IRs captured furthest away from the sound source were applied to the TpBL and TpBR ambience channels.

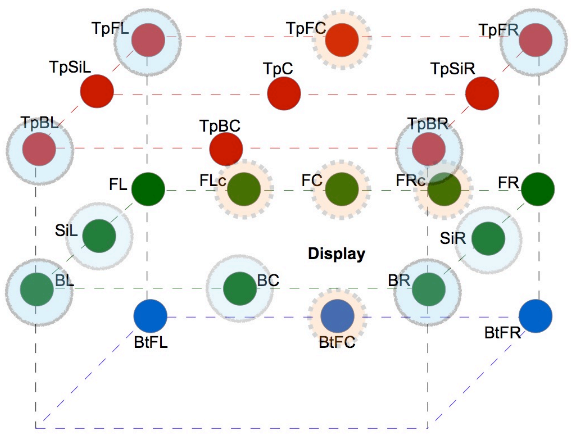

Figure 11. Guitar solo direct (orange) and ambient (light blue) signals.

5 Informal Evaluations

Several informal evaluations of the final mix of "Working" were held at four different 22.2 reproduction facilities: Studio 22 at McGill University, Studio B at Tokyo University of the Arts, and Studios CD607 and AP-E at NHK's Shibuya Production Centre. In general, the mix was well received, with impressions largely consistent between venues. The spatial stability, physical shape, and timbre of instrument and vocal sonic images were found to be consistent in most listening positions, save for within a very close proximity of an individual loudspeaker. These images were also noted as having realistic horizontal and vertical extent. It was often observed by listeners that when moving within the reproduction environment, it was their perceived perspective within the sound scene that changed, as opposed to the sound scene itself. The relatively dry acoustic and subtle levels of ambience were felt to be aesthetically appropriate to the composition.

6 Discussion

General aesthetic and technical considerations

The experience of recording and mixing classical, jazz, electroacoustic, contemporary, and pop/rock music for 22.2 and other 3D audio formats has provided valuable insight regarding general aesthetic and technical considerations for immersive music production. Some key points follow:

1) When instruments are well captured as stable, coherent sound images, very little EQ, compression, or level automation is necessary. This is especially true for formats such as 22.2, whose greater number of loudspeakers allows for clearer physical separation of sound images.
2) Building a 3D sound scene is a question of density of information: how many microphone signals are required to construct realistic or hyper-realistic images and ambiences.
3) Ideally, as the listener moves away from the sweet spot, the environment should not change drastically: what should change is the perspective of the listener.
4) Not all instruments need to be captured within the same physical space. However, all captured spaces should coalesce to form a unified sound scene. This may require certain ambience fields, or parts thereof, to be transformed using tools such as delays, EQ, or artificial acoustics.
5) Recording and mixing tools and techniques optimized for stereo can be irrelevant or counterproductive in immersive audio.
6) When using sound capture methods similar to those described in the current case study, high track counts may be inevitable. Recording multiple instruments in the same space will help alleviate this problem, as they can all share a single ambience array.
7) A 3D panner for every channel is essential to fine tune microphone signal placement and optimize sonic images.

Adaptation of techniques

The recording techniques described in this case study are applicable to many other situations. Specific to 22.2, variations on this approach have already been used to record solo instruments, small jazz ensembles, and an electroacoustic piece for immersive chamber orchestra. The general approach to sound capture and mixing described in Section 2 has worked well across these various scenarios. Also, it has been observed that musical content originally recorded for 22.2 using these methods translates well when remixed for other, reduced channel-count formats [14]. Recently, several jazz recordings have been made for 4+5+0 reproduction using simplified versions of some of the capture systems described in this case study. The process of designing these complex close-microphone systems has also provided some insight as to how to fabricate 3D images from 1D material, as shown with the guitar solo discussed in Section 4.

The approach of using complex multi-microphone arrays per instrument would be equally valid for object-based environments. A set of microphone signals that make up a single sound image could be thought of as a multichannel-object. These objects could then theoretically be panned to any location within the 3D environment. It would be valuable to investigate how such an approach might impact the perceived stability and cohesion of a given sonic image.

Using large-scale spaced ambience arrays for recording each instrument or group of instruments requires a significant number of additional microphones and setup time, as compared with stereo production. As seen in Tables 1 and 2, the ambience channels tend to contribute disproportionally to track count, which may be problematic depending on the content creators' workstation. It would be worth experimenting with combining the direct sound microphone arrays developed for this and similar recording projects with large-scale 3D convolution reverb systems: a similar approach is described by Kim et al. in [23].

Acknowledgements

This work was supported by the Social Sciences and Humanities Research Council (SSHRC), The Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), and Humber College. Special thanks to Andrew Mullin, Mike Ardagh, and Jack Kelly.

References

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

[2] "Ultra High Definition Television - Audio Characteristics and Audio Channel Mapping for Program Production," *SMPTE*, ST 2036-2-2008 (2008).

[3] "Advanced sound system for programme production," ITU-R BS.2051-2, Geneva (2018).

[4] K. Hamasaki et al., "Advanced multichannel audio systems with Superior Impressions of Presence and Reality," in *AES Convention 116*, Berlin (2004).

[5] K. Hamasaki et al., "Effectiveness for height information for reproducing presence and reality in multichannel audio system," in *AES Convention 120*, Paris (2006).

[6] T. Hinata et al., "Live Production of 22.2 Multichannel Sound for Sports Programs," in *AES 40th International Conference*, Tokyo (2010).

[7] K. Irie and T. Miura, "The production of 'The Last Launch of the Space Shuttle' by Super Hi-Vision TV," in *Broadcast Engineering Conference, NAB show 2012*, Las Vegas (2012).

[8] I. Sawaya et al., "Dubbing Studio for 22.2 Multichannel Sound System in NHK Broadcasting Center," in *AES Convention 138*, Warsaw (2015).

[9] I. Sawaya et al., "Discrimination of Changing Loudspeaker Positions of 22.2 Multichannel Sound System Based on Spatial Impressions," in *AES Convention 134*, Rome (2013).

[10] T. Sugimoto, "Downmixing Method for 22.2 Multichannel Sound Signal in 8K Super Hi-Vision Broadcasting," *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 590-599 (2015).

[11] T. Komori et al., "Subjective loudness of 22.2 multichannel programs," in *AES Convention 138*, Warsaw (2015).

[12] W. Howie et al., "Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio," in *AES Convention 142*, Berlin (2017).

[13] S. Kim et al., "Subjective Evaluation of Multichannel Sound with Surround-Height Channels," in *AES Convention 135*, New York (2013).

[14] W. Howie et al., "Listener Discrimination Between Common Speaker-Based 3D Audio Reproduction Formats," *J. Audio Eng. Soc.*, vol. 65, no.10, pp. 796-805 (2017).

[15] W. Howie et al., "A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound," in *AES Convention 141*, Los Angeles (2016).

[16] B. Martin and R. King, "Three Dimensional Spatial Techniques in 22.2 Multi-channel Surround Sound for Popular Music Mixing," in *AES Convention 139*, New York (2015).

[17] B. Martin and R. King, "Mixing Popular Music in Three Dimensions: Expansion of the Kick Drum Source Image," in *Innovation In Music (InMusic'15)*, Cambridge, 2015.

[18] J. Meyer, Acoustics and the Performance of Music, Braunschweig: Springer (2009).

[19] B. Martin et al., "Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments," in *AES Conference on Sound Field Control*, Guilford (2016).

[20] B. Martin et al., "Subjective Graphical Representation of Microphone Arrays for Vertical Imaging and Three-Dimensional Capture of Acoustic Instruments, Part I," in *AES Convention 141*, Los Angeles (2016).

[21] W. Howie et al., "Subjective and objective evaluation of 9ch three-dimensional acoustic music recording techniques," in *AES International Conference on Spatial Reproduction – Aesthetics and Science*, Tokyo (2018).

[22] D. Griesinger, "Spatial Impression and Envelopment in Small Rooms," in *AES Convention 103*, New York (1997).

[23] S. Kim et al, "A Cross-Cultural Comparison of Salient Perceptual Characteristics of Height-Channels for a Virtual Auditory Environment," *Virtual Reality*, vol. 19, no. 3, pp. 149-160 (2015).