Audio Engineering Society

Convention Paper

Presented at the 141st Convention
2016 September 29–October 2 Los Angeles, USA

A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound

Will Howie, Richard King, and Denis Martin

The Graduate Program in Sound Recording, McGill University, Montreal, QC, H3A 1E3, Canada

The Centre for Interdisciplinary Research in Music Media and Technology, Montreal, QC, H3A 1E3, Canada

Correspondence should be addressed to Will Howie (william.howie@mail.mcgill.ca)

ABSTRACT

Based on results from previous research, as well as a new series of experimental recordings, a technique for three-dimensional orchestral music recording is introduced. This technique has been optimized for 22.2 Multichannel Sound, a playback format ideal for orchestral music reproduction. A novel component of the recording technique is the use of dedicated microphones for the bottom channels, which vertically extend and anchor the sonic image of the orchestra. Within the context of highly dynamic orchestral music, an ABX listening test confirmed that subjects could successfully differentiate between playback conditions with and without bottom channels.

1 Introduction

22.2 Multichannel Sound

In recent years, much work has been done to introduce and standardize various three-dimensional audio formats for cinema, home theatre, and broadcast [1]–[6]. Japan Broadcasting Corp. (NHK) has developed and introduced Super Hi-Vision, "an ultra-high definition video system with 4000 scanning lines and a viewing angle of 100°." [6] Super Hi-Vision includes a complementary, immersive audio format: 22.2 Multichannel Sound [7], now standardized by SMPTE [8] and the ITU [5]. Utilizing ten playback channels at ear level, nine above the listener (top layer), and three at floor level (bottom layer) [Fig. 1], 22.2 Multichannel Sound has been shown to significantly increase presence over a wide listening area, as compared with traditional 5.1 multichannel systems [9]. NHK has already produced a number of special programs featuring 22.2 Multichannel Sound, and plans to be broadcasting in Super Hi-Vision in time for the 2020 Tokyo Olympics [10].

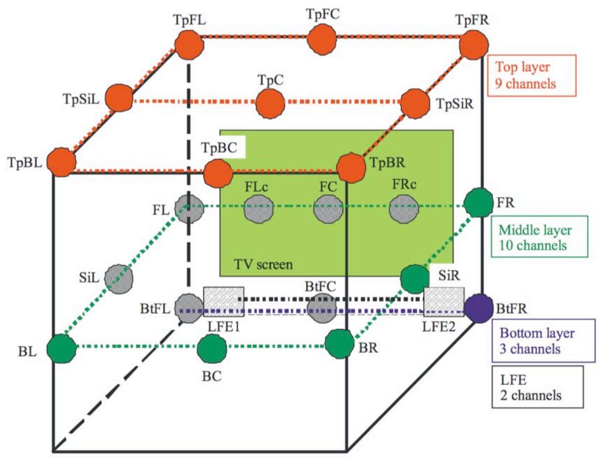

Figure 1. 22.2 Multichannel System [1]

3D Audio and Classical Music Recording

Listeners of recorded classical music have become accustomed to an idealized, realistic recreation of a live performance in an acoustic space [11] [12]. In multichannel audio, this aesthetic typically involves a “concert” perspective ie., instruments are reproduced in front of the listener, while ambience surrounds from the sides, behind and above. Hinata et al. state, “From years of experience in mixing 5.1 surround sound and 22.2 multichannel sound, it is known that close sounds heard from the sides and back create a psychological feeling of pressure, which results in a small spatial impression” [13].

A primary goal of classical music recording engineers is to capture an ideal balance of direct and diffuse sound: many microphone techniques have been developed to do just this, for both stereo and 5.1 surround sound [14] [15]. These techniques, however, do not capture the fully immersive experience of listening to a live performance in a real acoustic environment. Three-dimensional audio with height channels has been shown to improve the depth, presence, envelopment, naturalness, and intensity of music recordings [16]–[18]. Several authors have introduced three-dimensional music recording techniques and/or concepts primarily aimed at classical ensemble capture [19]–[22], but these techniques tend to be designed and optimized for smaller-scale three-dimensional audio formats. Specific to 22.2 Multichannel Sound, most publications have discussed capture methods for special events [1] [23] and live sports broadcast [13].

Spatial Impression in Multichannel Sound

In concert hall acoustics, spatial impression is typically divided into two broad categories: Apparent Source Width (ASW), and Listener Envelopment (LEV). For multichannel music production it is envelopment (and in the case of classical music, environmental envelopment) that is the more important of the two spatial attributes. Berg and Rumsey have done much research in the area of perceived spatial quality in reproduced sound, finding that “an enveloping sound gave rise to the most positive descriptors and that the perception of different aspects of the room was most important for the feeling of presence.” – presence being defined as “The experience of being in the same acoustical environment as the sound source, e.g. to be in the same room” [24]. When examining correlation between various subjective spatial attributes, Berg and Rumsey found that “preference” was most strongly correlated with “envelopment”, while “naturalness” was most strongly correlated with “presence” [24]. In multichannel audio, creating a strong sense of envelopment is key to achieving listener enjoyment and immersion.

Hanyu and Kimura have shown that LEV “increases if there is adequate spatial balance in the direction of arriving reflections” [25]. In David Griesinger’s model of spatial impression, what he calls “background spatial impression” (BSI) is closely tied to envelopment. Griesinger asserts that in order to achieve high levels of spaciousness, large fluctuations in the interaural intensity difference (IID) and interaural time difference (ITD) at the two ears during background sound are required [26]. Griesinger suggests that maximum spaciousness will occur when the reverberant component of a recording is fully deccorlated, and recommends that component should be “reproduced by an array of decorrelated loudspeakers around the listener.” [26] Hiyama et al. showed that for loudspeakers placed at even intervals around the listener, “at least six loudspeakers are needed to reproduce the spatial impression of (a) diffuse sound field” [27].

22.2 Multichannel Sound for Orchestral Music Recording and Reproduction

Most current three-dimensional audio formats retain the traditional 60° frontal sound reproduction angle associated with stereo and 5.1 surround sound [1] [5]. 22.2 Multichannel Sound, however, has a frontal sound reproduction angle of 120°. This wider reproduction angle is ideal for reproducing the sonic image of an ensemble as large as a symphony orchestra. Hamasaki et al. have shown that the use of five frontal speakers (as opposed to three) is essential for creating the impression of presence [9].

As seen in Figure 1, the even spatial distribution of loudspeakers at both ear level and above in 22.2 Multichannel Sound is ideal for the reproduction of early and late lateral reflections, as well as a fully deccorelated reverberant field, all of which will ensure maximum listener envelopment at all frequencies, as well as contributing to the dimensional broadening of the orchestral image.

Bottom Channels in 22.2 Multichannel Sound

Within the current literature, only Hamasaki and his co-authors have specifically addressed recording orchestral music for 22.2 Multichannel Sound – however, the presented techniques do not use the bottom channels [9] [29]. These loudspeakers are typically located bellow FL, FC, and FR, at floor level. The bottom channels were originally intended to reproduce special effects germane to on-screen action. However, an ideal presentation of an orchestra would make use of these channels in order to extend the ensemble image to the floor, as it would sound from the conductor's perspective.

Through a number of experimental recordings of classical and pop/rock music at McGill University, it has been observed that the bottom channels add a great deal of "weight" to the instruments or ensembles being reproduced, providing a lower vertical extension that anchors the sonic image. This "anchoring" effect is highly useful, as correlated or semi-correlated sonic information in the height channels has the tendency to cause instrument images to move upward, which may not be desirable [30] [31]. Listeners have been shown to prefer higher levels of vertical immersive content in a three-dimensional playback environment [32]; increasing the vertical extent of the direct sound image by use of the lower channels allows the recording engineer to maximize the level of immersive ambience. Martin and King have shown the importance of vertically extending sound images using the bottom channels in re-mixing one-dimensional content for three-dimensional playback environments [33].

Dedicated microphones for the bottom channels also have the advantage of capturing early floor reflections, as well as additional low frequency content, due to the complex radiation patterns of orchestral instruments [34]. Loudspeakers at or near floor-level are capable of more efficient low-frequency reproduction to the listener, as they do not suffer from low frequency spectral notches caused by interference between direct sound and floor reflected sound that would be found in sound reproduction from speakers at ear level and above [35].

Roffler and Butler found that tonal stimuli have intrinsic spatial characteristics – different tone bursts reproduced by a single loudspeaker will be located by listeners as being higher or lower in space, depending on their frequency [36]. Cabrera and Tilley observed that, "Having low frequencies originate from lower sources is in harmony with the pervasive pitch-height metaphor" [35]. The lower channels allow the recording/mixing engineer to concentrate low frequency power bellow and in front of the listener, which is in keeping with an apparent natural human aesthetic.

2 Design of Microphone Technique

Based on the above considerations, previous research [30] [31], established one and two-dimensional recordings techniques [11]–[14], as well as a number of experimental three-dimensional music recordings, a new technique was designed to record orchestral music, optimized for 22.2 Multichannel reproduction. The technique is designed to incorporate new considerations for creating convincing three-dimensional sound images, but retains compatibility with stereo recording techniques.

Orchestral Sound Capture

The primary component of the orchestral sound image is reproduced by the five front speakers (FL, FLc, FC, FRc, FR) and the three bottom speakers (BtFC, BtFL, BtFR). The main frontal sound capture is based on the classic "Decca Tree" model of three spaced omnidirectional microphones (assigned to FLc, FC, FRc), with two additional omnidirectional "outrigger" microphones placed at the three-quarter points of the orchestra (assigned to FL and FR). The FLc, FC and FRc microphones are fitted with acoustic pressure equalizers, diffraction attachments that increase microphone directivity at high frequencies, as well as give a natural boost in the "presence" range of the frequency spectrum (1kHz - 5kHz) [37]. The use of omnidirectional microphones is critical for capturing the complete low frequency spectrum of the orchestra, as they do not suffer from proximity effect. These "front" microphones should be placed somewhat closer to the orchestra than typical of a stereo-only recording, so as to capture less ambient sound. This is aided by the increased directivity introduced by using acoustic pressure equalizers. When also being used as the main system for a stereo mix, the "front" microphones can be combined with several ambience microphones as necessary.

Three additional microphones are placed adjacent to the FL, FC and FR microphones, ideally within a meter of the floor, angled downward and routed to the three bottom channels. These bottom microphones anchor and vertically extend the image of the orchestra, as well as capturing early floor reflections and low frequency content from instruments with frequency dependent directivity.

Ambient Sound Capture

In the proposed technique, microphones routed to the remaining fourteen playback channels are all directional (mostly cardioid), placed in such a way as to prioritize ambient sound capture, and decorrelation between channels. Cardioids are chosen for their high level of rear sound rejection, as well as being less susceptible to low frequency loss than hypercardioid or bidirectional microphones. Some amount of low frequency roll-off (due to proximity effect) is desirable, especially in the height channels, as it reinforces the above mentioned "pitch-height metaphor" [35]. In order to optimize listener envelopment, microphones that prioritize early and late reflection capture should be decorrelated at all frequencies. This can be achieved through distant spacing between microphones. Hamasaki et al. found that a distance of at least 2m between microphones was necessary to ensure decorrelation above 100Hz [11], while Griesinger has suggested that spacing microphones at a distance greater then the reverb radius of the recording venue will ensure decorrelation at low frequencies [38].

Ideally, microphone capsule direction should roughly mirror playback speaker direction – for example, microphones routed to the SiL and SiR speakers would face the side walls, primarily capturing lateral reflected sound. The exception would be the TpFL, TpFC and TpFR microphones, which may need to be oriented away from the orchestra in order to minimize direct sound capture. Microphone positions need not be dictated by layout and relative distances between reproduction loudspeakers – rather, microphone placement and optimization should be based on capturing an ideal reverberant sound field.

3 Implementation of Design

The proposed recording technique was implemented during recording sessions for the 90-piece, National Youth Orchestra of Canada (NYOC). The recordings took place over three days in McGill University's Music Multimedia Room, a large scoring stage measuring 24.4m x 18.3m x 17m, with an RT60 of approximately 2.5s [Fig 2]. Monitoring of the recordings took place in the adjacent Studio 22, an acoustically treated listening room that fulfills ITU-R BS.1116 [39] requirements. 22 Musikelectronic Geithain GmbH *M-25* speakers are arranged for 22.2 Multichannel Sound reproduction, as per [8]. Monitoring in a 22.2 playback environment was essential in order to understand the complex sonic relationships between the different points of ambient sound reproduction. Hearing how all 22 playback channels resolved to form a single audio scene was critical to microphone placement and adjustment.

Height channel microphones were hung from various catwalks above the studio floor in such a way that positional optimization could take place during recording breaks. Microphone choice and placement were as seen Table 1 and Figures 3 and 4. The majority of the recording venue's floor space was occupied by the orchestra, with only 3.9m from the conductor to the back wall of the studio [Fig 2]. As such, certain ambience microphones were spaced extremely widely in order to grain greater distance from the "frontal" microphones, thereby ensuring an appropriate amount of depth in the audio scene. To avoid strong rear wall reflections in the BC channel, a laterally oriented bi-directional microphone was used.

Figure 2. NYOC in MMR

Channel	Microphone
FL	Schoeps MK2s
FR	Schoeps MK2s
FC	Schoeps MK2H w/APE
BL	Schoeps MK 21
BR	Schoeps MK 21
FLc	Schoeps MK2H w/APE
FRc	Schoeps MK2H w/APE
BC	Neumann KM120
SiL	DPA 4011
SiR	DPA 4011
TpFL	Schoeps MK 4
TpFR	Schoeps MK 4
TpFC	Schoeps MK 4
TpC	DPA 4011
TpBL	Schoeps MK 4
TpBR	Schoeps MK 4
TpSiL	Schoeps MK 4
TpSiR	Schoeps MK 4
TpBC	Schoeps MK 4
BtFC	DPA 4011
BtFL	DPA 4011
BtFR	DPA 4011

Table 1. Microphones used for test recording

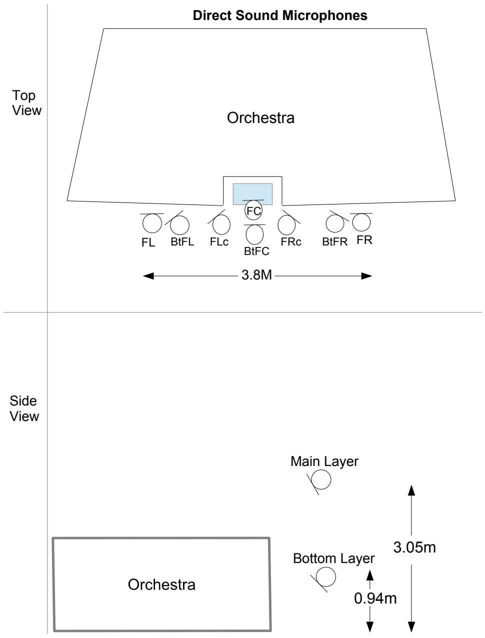

Figure 3. Orchestral Sound Capture

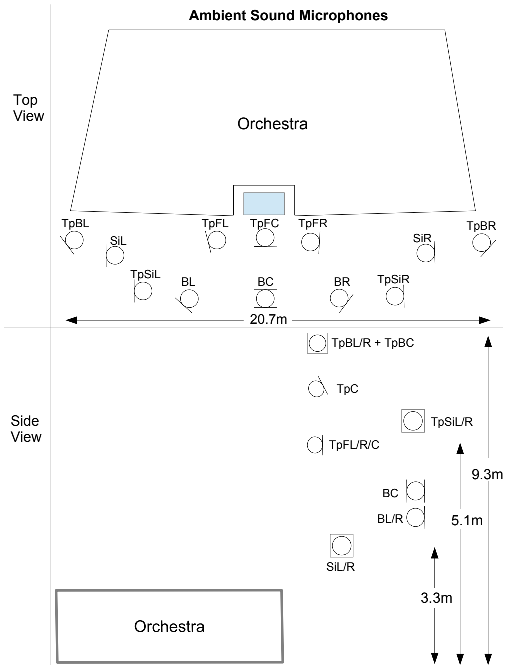

Figure 4. Ambient Sound Capture

4 Evaluation of Recording

A balanced mix was created of "Mars, The Bringer of War" from Gustav Holst's *The Planets*, using only the above described 22-microphone "main system". A number of informal listening sessions have taken place to form a preliminary evaluation of the recording technique. Because this is the first 22.2 orchestral recording made at McGill University, there are no "reference recordings" with which to compare it in a formal listening test. An excerpt of this mix was recently used as part of research performed by the BBC, investigating various current immersive audio standards [40].

The NYOC recording has been heard by a number of fulltime and visiting faculty of the Graduate Program in Sound Recording at McGill University, as well numerous visiting researchers and recording engineers. In general, comments have been positive. Many have noted the large, coherent, and realistic orchestral image, natural depth of field, excellent instrument and sectional image clarity, and enveloping reverberation. The overall impression seems to be a naturalistic listening experience.

The same recording has been heard as a part of informal listening sessions in four studios in Japan designed and/or equipped for 22.2 Multichannel Sound reproduction: NHK STRL, NHK's Shibuya production center, Tokyo University of the Arts, and Yamaha Corp. The recording was well received, with similar observations as noted above. It was observed that the sound of the mix was generally consistent across all playback venues, which themselves varied extensively in terms of size, acoustic treatment, and speaker radius.

5 Evaluation of Bottom Channels

Listening Test

A double blind ABX listening test was designed to determine whether or not, within the context of a highly dynamic orchestral music recording, subjects could successfully differentiate between playback conditions with and without the three bottom channels. Although a seemingly simple task, this was felt to be a good first question to answer, before moving on to a more complex perceptual evaluation of the contribution of the bottom channels to orchestral music recording. All testing took place in McGill University' Studio 22.

Test stimuli consisted of three 25s excerpts from "Mars, The Bringer of War": 1) a relatively soft passage, 2) a passage ranging from mezzo-forte to forte, and 3) the very loud ending of the piece. Mixes of each stimulus were prepared with and without bottom channel content. Bottom and non-bottom channel mixes were then level matched within 0.2LUFS using a Neumann KU 100 dummy head at the listening position, whose output was measured using a HOFA LUFS meter.

Test subjects were seated at the listening position in Studio 22. Prior to performing the listening test, each subject took part in a brief training session, in which they were given time to familiarize themselves with the three musical excerpts, the Pro Tools session being used as the testing interface, and the "with bottom channels" and "without bottom channels" stimuli mixes.

For each test trial, subjects were presented with one of the three musical excerpts, on a repeating loop, and asked to compare mixes labelled "X" "A" and "B" by selecting between three VCA groups in Pro Tools. Subjects were instructed to identify the mix that was "different from X". (During preliminary tests, subjects found this to be a more logical task then identifying the mix that was "the same as X".) Subjects recorded their answers on an online form that also included a short demographic survey and comments section to be completed after the listening test. Subjects were also asked to rate the difficulty of the task. Each participant saw each excerpt three times, for a total of nine trials. Mix stimuli assignment to VCA groups (X, A and B) was randomized, as was the order of excerpt presentation.

Results

14 subjects performed the listening test – all had at least two years of experience and/or training as recording engineers, and reported having normal hearing. Each subject completed nine trials for a total of 126 trials. The participant's response was

marked "correct" if they successfully discriminated between the A and B stimuli (with X as reference) and "not correct" if they did not. Throughout the entire analysis the overall success rate was examined as well as the success rates for each musical excerpt (soft, medium, and loud). The percentage of correct responses is shown in Figure 5. An overall success rate of 69% was achieved and a binomial test (Table 2) shows that this result is highly significant. When looking at the three musical excerpts individually, significant discrimination rates were achieved for both the medium (81%) and soft (67%) excerpts. However, the 60% discrimination rate for the loud excerpt was not significantly above chance. Finally, participants rated the task difficult overall, giving it a mean rating of 3.9 (S.E. 3.8-4.1) on a scale from 1 (Easy) to 5 (Hard).

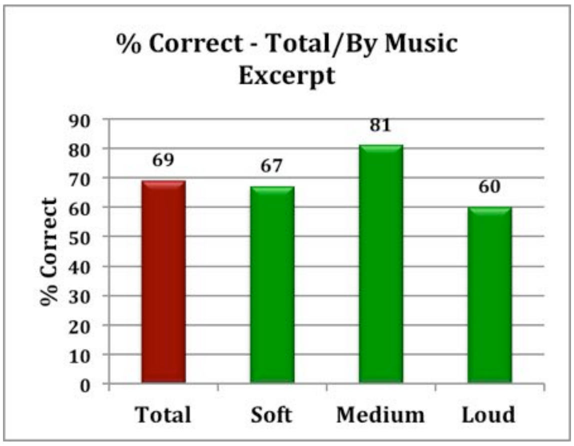

Figure 5. Percentage of correct responses.

Data Group	Prob-ability	95% Conf. Interval	*p*-value
Total	.6905	0.60-0.77	**<0.001**
Soft	.6667	0.50-0.80	**0.044**
Medium	.8095	0.66-0.91	**<0.001**
Loud	.5952	0.43-0.74	0.28

Table 2. Binomial Test

6 Discussion and Future Work

Informal Evaluations

Based on preliminary observations, it can be said that the proposed recording technique captures a broad, vertically anchored orchestral image with a natural depth of field and clear image localization, as well as highly enveloping ambience. More formal and rigorous subjective evaluations are now required (see Future Work).

Bottom Channel Evaluation

In comments left in the post-test survey, as well as those made to the examiners afterwards, a majority of subjects commented on the subtlety and difficulty in detecting when the bottom channels were in use. This is also reflected in the analysis of the perceived difficulty rating. For orchestral music, this is not surprising, especially considering how the bottom channels were mixed in comparison with the main front channels (typically 7LUFS lower in output). As such, a 69% probability of success ($p<0.001$) is considered a valid result in demonstrating the ability of listeners to discriminate between playback conditions with and without bottom channels.

Many subjects also commented on what they were listening for when attempting to discriminate between playback conditions. A number of subjects felt they could discern more low frequency information when the lower channels were active, particularly in the mezzo forte/forte excerpt. Not surprisingly, it was often observed that the image of the orchestra extended further towards the floor when the bottom speakers were active. Several subjects commented that the bottom channels contributed to a broadening of the orchestral image, similar to what is often described in concert hall acoustics as Apparent Source Width. This is quite interesting, as ASW is typically associated with lateral reflections.

Source Material for Spatial Audio Evaluation

Analysis showed that the "loud ending" musical excerpt had the lowest percentage of correct differentiation of playback conditions, and that the discrimination rate was not above chance. It is likely that for this loud passage, the difficultly experienced by the listeners was due to the dynamic envelope of the music. This passage was made up of brief fortissimo tutti orchestra chords, separated by moments of silence. The large variance in the overall dynamic envelop makes it very difficult for the participants to find an appropriate place to “switch” between A, B and X. Rumsey has discussed the importance of choosing appropriate source material as stimulus within the context of listening tests designed to evaluate sound quality, noting that the choice of source material “can easily dictate the results of an experiment, and should be chosen to reveal or highlight the attributes in question.” [41] Similarly, ITU-R BS.1116-1 states that “Only critical material is to be used in order to reveal differences among systems under test. Critical material is that which stresses the systems under test” [39]. For these types of listening tests, which seek to reveal subtle differences in sound quality, it is advisable to use source material that is relatively static in both dynamic envelope as well as spectrum for the entire length of the excerpt, thus making any differences between stimuli more apparent to the listener.

Future Work

Evaluation of the proposed orchestral recording technique is still very much in its early stages. A major hindrance to any serious subjective evaluations of the technique (and subsequent recording) is a lack of “reference” material with which to compare the NYOC recording. As such, a large scale research recording session is planned for Fall 2016. The proposed technique will be setup and optimized to record several days of orchestral rehearsals. As a comparison, several other three-dimensional orchestral music recording techniques will be setup for simultaneous recording. This should yield a number of different recordings that can be used for extensive subjective comparison between large-scale three-dimensional recording techniques, as well as further validation of the technique proposed herein.

A more comprehensive evaluation of the effectiveness of bottom channels in music reproduction is also required. It would be valuable to extend this investigation beyond orchestral music, and include recordings of other genres of music, such as chamber music, jazz, and pop/rock.

7 Conclusions

A three-dimensional orchestral music recording technique, optimized for 22.2 Multichannel Sound reproduction has been developed. The technique prioritizes the capture of a natural orchestral sound image, with realistic horizontal and vertical extent, as well as a highly diffuse, enveloping reverberant sound field. Using the proposed technique, a recording was made of the National Youth Orchestra of Canada. Preliminary evaluations of the recording have been positive. A subsequent listening test showed that subjects can successfully differentiate between playback conditions with and without the use of the bottom channels in an orchestral music mix. More test recordings using the proposed technique are required, as well as further, more formal subjective evaluation of its effectiveness.

8 Acknowledgements

This work was supported by The Social Sciences and Humanities Research Council (SSHRC) and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT). Special thanks to the 2015 National Youth Orchestra of Canada.

References

[1] "Multichannel sound technology in home and broadcast applications", *Report ITU-R BS.2159-4*, ITU, Geneva, 2012, pp. 1-54.

[2] J. Riedmiller and N. Tsingos, "Recent Advancements in Audio - How a Paradigm Shift in Audio Spatial Representation and Delivery Will Change the Future of Consumer Audio Experiences," in *2015 Spring Technical Forum Proceedings*.

[3] "Productions in Auro-3D," *White Paper*, Auro Technologies, Feb 18 2012, pp. 1-16.

[4] "Dolby Atmos for Home Theatre", White Paper, April 2015, Dolby Laboratories. pp. 1-18.

[5] "Advanced sound system for programme production," ITU-R BS.2051-0, International Telecom Union: Geneva, Switzerland, 2014.

[6] K. Hamasaki, "The 22.2 Multichannel Sounds And Its Reproduction At Home And Personal Environment," in *AES 43rd International Conference*, Pohang, Korea, 2011, pp. 1-4.

[7] K. Hamasaki and K. Hiyama, "Development of a 22.2 Multichannel Sound System, *Broadcast Technolgy*, no. 25, Winter 2006, pp. 9-13.

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

[9] K. Hamasaki et al., "Advanced multichannel audio systems with Superior Impressions of Presence and Reality," in *AES 116th Convention*, Berlin, Germany, May 2004, pp. 1-12.

[10] I. Sawaya et al., "Dubbing Studio for 22.2 Multichannel Sound System in NKH Broadcast Centre," in *AES 138th Convention*, Warsaw, Poland, 2015, pp. 1-9.

[11] K. Hamasaki et al, "Approach and Mixing Technique for Natural Sound Recording of Multichannel Audio," in *AES 19th International Conference*, Schloss Elmau, Germany, 2001, pp. 1-6.

[12] A. Fukada, "A Challenge in multichannel music recording," in *AES AES 19th International Conference*, Schloss Elmau, Germany, 2001, pp. 1-8.

[13] T. Hinata et al, "Live Production of 22.2 Multichannel Sound For Sports Programs," in *AES 40th International Conference*, Tokyo, Japan, 2010, pp. 1-10.

[14] M. Dickreiter, *Tonmeister Technologies*. New York, NY: Temner Enterprises Inc., 1989.

[15] F. Rumsey, *Spatial Audio*. Burlington: Focal Press. 2013.

[16] K. Hamasaki et al, "Effectiveness for height information for reproducing presence and reality in multichannel audio system," in *AES 120th Convention*, Paris, France, 2006, pp. 1-15.

[17] T. Kamekawa et al, "Evaluation of spatial impression comparing 2ch stereo, 5ch surround, and 7ch surround with height for 3D imagery," in *AES 130th Convention*, London, UK, 2011, pp. 1-7.

[18] S. Kim et al, "Subjective evaluation of multichannel Sound with Surround-Height channels," in *AES 135th Convention*, New York, USA, 2013, pp. 1-7.

[19] P. Geluso, "Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays," in *AES 132nd Convention,* Budapest, Hungary, 2012, pp. 1-5.

[20] G. Theile and H. Wittek, "Principals in Surround Recording with Height," in *AES 130th Convention*, London, UK, 2011, pp. 1-12.

[21] M. William, "The Psychoacoustic Testing of the 3D Multiformat Microphone Array Design, and the Basic Isosceles Triangle Structure of the Array and the Loudspeaker Reproduction Configuration," in *AES 134th Convention*, Rome, Italy, 2014, pp. 1-14

[22] D. Bowles, "A Microphone Array for Recording in Surround-Sound with Height Channels," in *AES 139th Convention*, New York, USA, 2015, pp. 1-7.

[23] T. Miura et al, "The Production of 'The Last Launch of Space Shuttle' by Super Hi-Vision," *NHK*, Tokyo, Japan, pp. 1-8.

[24] J. Berg and F. Rumsey, "Systematic Evaluation of Perceived Spatial Quality," in *AES 24th*

International Conference on Multichannel Audio, Banff, Canada, 2003, pp. 1-15.

[25] T. Hanyu and S. Kimura, “A new objective measure for evaluation of listener envelopment focusing on the spatial balance of reflections,” *Applied Acoustics*, Vol. 62, 2001, pp. 155-184.

[26] D. Griesinger, "Spatial Impression and Envelopment in Small Rooms," in *AES 103rd Convention*, New York, USA, 1997, pp. 1-12.

[27] K. Hiyama et al, "The minimum number of loudspeakers and its arrangement for reproducing spatial impression of diffuse sound field," in *AES 113th Convention*, Los Angeles, USA, 2002, pp. 1- 12.

[28] T.D. Rossing, "Acoustics in Halls for Speech and Music," *Springer Handbook of Acoustics*. Springer: New York, 2007. Chapter 9, pp. 302-315.

[29] K. Hamasaki and Wilfried Van Baelen, "Natural Sound Recording of an Orchestra with Three-dimensional Sound," in *AES 138th Convention*, Warsaw, Poland, 2015, pp. 1-8.

[30] W. Howie and R. King, “Exploratory microphone techniques for three-dimensional classical music recording,” in *AES 138th Convention*, Warsaw, Poland, 2015, pp. 1-4

[31] W. Howie et al, "Listener preference for height channel microphone polar patterns in three-dimensional recording," in *AES 139th Convention*, New York, USA, 2015, pp.1-10.

[32] B. Martin et al, "Immersive content in three-dimensional recording techniques for single instruments in popular music," in *AES 138th Convention*, Warsaw, Poland, 2015, pp. 1-7.

[33] 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, UK, 2015, pp. 1-10.

[34] C. Hugonnet and P. Walder, *Stereophonic Sound Recording: Theory and Practice*. New York: Wiley, 1998.

[35] D. Cabrera and S. Tilley, "Vertical localization and image size effects in loudspeaker reproduction," in *AES 24th International Conference on Multichannel Audio*, Banff, Canada, 2003, pp.1-8.

[36] S. Roffler and R. Butler, “Localization of Tonal Stimuli in the Vertical Plane,” *Journal of the Acoustical Society of Amercica*, vol. 43, 1968, pp. 1260-1266.

[37] W. Woszczyk, "Acoustic Pressure Equalizers," *Pro Audio Forum*, pp. 1-24.

[38] D. Griesinger, “The Science of Surround,” New York, USA, 1995.

[39] “Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems”, ITU-R Recommendation BS.1116-1, International Telecom Union: Geneva, Switzerland, 1997, pp. 1-26.

[40] N. Zacharov et al., “Next Generation Audio System Assessment using the Multiple Stimulus Ideal Profile Method,” in *8th International Conference on Quality of Multimedia Experience*, Lisbon, Portugal, 2016, pp. 1-6.

[41] F. Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *JAES*, Vol 50, No 9, 2002, pp. 651-666.