Audio Engineering Society

Convention Paper

Presented at the 142nd Convention
2017 May 20–23 Berlin, Germany

Subjective Evaluation of Orchestral Music Recording Techniques for Three-Dimensional Audio

Will Howie[1,2], Richard King[1,2], Denis Martin[1,2], and Florian Grond[2, 3]

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

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

[3] *The Input Device and Music Interaction Laboratory, Montreal, QC, H3A 1E3, Canada*

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

ABSTRACT

A double-blind study was conducted to evaluate a recently developed microphone technique for three-dimensional orchestral music capture, optimized for 22.2 Multichannel Sound. The proposed technique was evaluated against a current 22.2 production standard for three-dimensional orchestral music capture, as well as a coincident, higher order ambisonics capture system: the Eigenmike™. Analysis of the results showed no significant difference in listener evaluation between the proposed technique and the current production standard in terms of the subjective attributes "clarity", "scene depth", "naturalness", "environmental envelopment", and "quality of orchestral image".

1 Introduction

Recording Acoustic Music for 3D Playback

Howie et al. recently developed a new method for three-dimensional orchestral music recording, optimized for Japan Broadcasting Corporation (NHK)'s 22.2 Multichannel Sound (22.2) [1] [2]. The technique is designed to take advantage of several aspects of the 22.2 reproduction environment that make it uniquely suited to orchestral music reproduction. Featuring 10 playback channels at ear level, nine above the listener, and three below, 22.2 is one of the most advanced and robust of the currently standardized 3D audio formats [3]. Five frontal speakers at ear level, with a reproduction angle of 120º, and three below the listener allow for the creation of a large, stable orchestral image that gives the listener the impression of an idealized conductor's perspective. An even spatial distribution of surrounding loudspeakers allows for realistic reproduction of early and late reflections, and a reverberant field that is highly decorrelated at all frequencies. These factors are key to achieving strong levels of listener envelopment [4 – 6] as well as a dimensional broadening of the sound source image [7]. Previous research [1] concluded that while the proposed recording technique performed well in informal listening tests that took place at five different 22.2 reproduction facilities, a more formal subjective evaluation was required.

Several authors have proposed microphone techniques for three-dimensional classical music recording, optimized for 9.1 or similar formats [8 – 11]. However, few of these new techniques have been examined through formal subjective listening tests. Ryaboy investigated perceptual differences between two recording techniques: Double MS+Z (a coincident technique that captures B-format signals), and Twins Square (a mixed spaced/coincident technique) [12]. Results of a double-blind listening test were reported as showing significant differences between the two techniques regarding "localization" (horizontal and vertical) and "perceived room size".

Hamasaki et al., introduced a method for three-dimensional orchestral music recording as part of an investigation comparing 22.2 (though the three bottom channels were not used) with reduced playback conditions: 17.1 (FL and FR removed), 10.1 (mid-layer only) and 5.1 [13]. Hamasaki and Van Baelen describe an updated version of the same technique in [14]. When compared with stereo and 5.1 recordings of the same material, subjects in a listening test rated the three-dimensional recording significantly higher than stereo and 5.1 for a number of subjective attributes, including "deep", "elevation", "spaciousness", "envelopment", and "good sound". The technique described by Hamasaki and his co-authors (or variations on said technique) has been used by NHK recording engineers for numerous orchestral music recordings, and is as such, considered a current production standard optimized for 22.2. No known publication has subjectively compared multiple three-dimensional capture methods optimized for 22.2.

Recording Array Comparisons for 5.1 Surround

Within the realm of 5.1 surround sound, there is far more literature exploring subjective comparisons and evaluations of recording techniques. Kassier et al. [15], and Heitala [16] examined differences between spaced (eg. Fukada Tree) and semi-spaced (eg. OCT surround) techniques. Within the context of an informal comparison, listeners consistently preferred Fukada Tree paired with Hamasaki Square [15]. Camerer and Sold [17], Kim et al., [18], Kamekawa et al., [19], Paquier et al., [20], A. Sitek and B. Kostek [21], and Peters et. al. [22] all undertook investigations that included evaluating perceptual differences and/or preferences between spaced, semi-spaced and coincident surround recording techniques. These publications often investigated different aspects of multichannel sound, and as such, depending on the research question, certain spaced or semi-spaced recording techniques tended to perform better than others. However, a consistent trend found within these publications is that regardless of the subjective or preference attribute(s) being investigated, coincident techniques tend to be rated on the negative end of the spectrum. This was true for 1st order ambisonics techniques [17, 18, 20-22], higher order ambisonics (HOA) [20], and Double MS [19].

Motivation

The primary aim of this study is to evaluate the effectiveness of the newly proposed recording technique as compared with a current production standard for 22.2 optimized orchestral music capture in terms of salient spatial sound attributes. A secondary aim is to compare the performance of these two spaced techniques with a coincident, HOA-based capture system.

2 Recording Techniques Under Investigation

A detailed explanation of **Technique 1**'s design rational can be found in [1]. Primarily direct orchestral sound is captured by a modified "Decca Tree" of five omnidirectional microphones, the middle three of which are outfitted with acoustic pressure equalizers [37]. Three directional microphones placed 1m above the stage floor provide signal for the bottom channels, vertically extending and anchoring the orchestral image. Widely spaced directional microphones capture decorrelated, spatially diffuse ambience, and are assigned to the remaining main layer and height channels. The technique is designed to retain the traditional "concert perspective" that is typical of most multichannel classical music recordings. Microphone orientation typically mirrors assigned playback channel orientation: for example, the TpFL microphone would have a horizontal orientation of around 60º, and a vertical orientation of approximately 45º.

Technique 2 was designed by Hamasaki and his co-authors, as described in [13] and [14]. The technique is a logical extension of Hamasaki's earlier publications on multichannel music recording, particularly "Reproducing Spatial Impression With Multichannel Audio", co-authored with Hiyama [23]. Direct sound from the orchestra is captured by an array of 5 supercardioid microphones, placed at equal intervals across the sound stage. In [13], ambient sound is captured with an array of laterally oriented bi-directional microphones – an extension of the well-known "Hamasaki Square" [23]. The placement and spacing of the bi-directional microphones ensures minimal capture of direct and rear wall sound, and that the ambient sound field is decorrelated across the audible frequency spectrum. Several of these ambience microphones are assigned to the front channels, to be mixed in if the recording engineer feels the orchestral sound is too “dry”.

In [14] this approach is updated, using vertically oriented supercardioid microphones as height channels. This version of the technique is representative of current 3D orchestral music recordings being made by NHK recording engineers, and thus can be considered a de-facto production standard for 22.2. Neither [13] nor [14] specify microphones for the bottom channels. For this study, three Sanken CUB-01 miniature boundary microphones have been added to the technique, each placed as far down-stage as possible (see Fig. 1). These microphones were chosen for their minimal visual impact, an important factor in broadcast sound recording, as well as to contrast with the bottom channel microphones used in Technique 1.

Technique 3
As seen in the introduction, several studies comparing multichannel recording techniques have included coincident recording systems. When considering the complexity, cost, and time associated with setting up either Techniques 1 or 2, the potential advantages to using a single-point, ambisonics-based capture system become obvious. As such, for this study, the Eigenmike (em32) was chosen as a 3rd recording technique. The em32 from Mh acoustics is a spherical microphone array where each of the 32 capsules is calibrated for magnitude and phase response. The accompanying software Eigenstudio converts the microphone signals into 3rd order ambisonics b-format. 16 channels were recorded following the ACN channel order convention with N3D normalization [31]. The em32 was placed approximately 1m above the conductor.

3 Setup and Optimization of Recording Techniques

The three techniques under investigation were installed in Pollack Hall, a medium sized concert hall with a seating capacity of 590. The hall measures 36m long by 18m wide by 12m high. Reverb times for the empty hall are listed below:

f (Hz)	63	125	250	500	1k	2k	4k
RT60	2.3s	2.0s	1.7s	1.8s	1.8s	1.7s	1.4s

Table 1: RT60 for Pollack Hall

The side and stage walls are equipped with acoustic curtains designed to decrease RT60. For this study, all acoustic curtains were “out” (removed) except for the stage curtains which were set to “¾ out” to control onstage reflections.

The microphones for all three techniques were installed the day before a week of orchestral rehearsals, with the goal of having all three techniques fully optimized before recording the final dress rehearsal. All microphones were routed to RME Micstacy preamps and A/D converters. Two streams of optical MADI output from the Micstacys were routed via fibre optic lines to Studio 22, a multichannel audio mixing room in an adjacent building. Studio 22 is equipped with 28 full-range, two-way loudspeakers (Musikelectronic Geithain GmbH *M-25*) and a stereo sub-woofer, arranged for reproduction of both 22.2 Multichannel Sound, and Auro 3D 9.1. The studio fulfills ITU-R BS.1116 requirements [24].

For Techniques 1 and 2, microphone choice and placement was based on [1], [13] and [14], as well as extensive experience recording orchestral music (see: Table 2 and Fig 1.) A current NHK production engineer provided valuable insight as to the optimization of Technique 2. Placement of the front five microphones for Technique 2 was based on available hanging points, and the increased “reach” of hypercardioid microphones as compared with omnidirectional microphones. Like the “Hamasaki Square”, Technique 2 included 3 frontal ambience (FrAmb) microphones to be mixed in with the direct orchestral sound as necessary. Microphones for all three techniques were either hung or placed on telescopic stands in the hall, depending on their desired height and location. Adjustments were made based on monitoring the orchestra’s rehearsals.

Channel	Technique 1	Technique 2
FL	Schoeps MK2S	Neumann KM185
FLc	Schoeps MK2H	Neumann KM185
FC	Schoeps MK2H	Senn. MKH 8050
FRc	Schoeps MK2H	Neumann KM185
FR	Schoeps MK2S	Neumann KM185
BL	Schoeps MK21	Schoeps MK8
BC	Neumann KM120	Neumann KM120
BR	Schoeps MK21	Schoeps MK8
SiL	Neumann KM184	Senn. MK30
SiR	Neumann KM184	Senn. MK30
TpFL	Schoeps MK4	Neumann KM185
TpFC	Schoeps MK4	Neumann KM185
TpFR	Schoeps MK4	Neumann KM185
TpC	Schoeps MK41	Schoeps MK41
TpBL	Schoeps MK4	Neumann KM185
TpBC	Schoeps MK4	Senn. MKH 8050
TpBR	Schoeps MK4	Neumann KM185
TpSiL	Schoeps MK4	Senn. MK50
TpSiR	Schoeps MK4	Senn. MK50
BtFL	DPA 4011	Sanken CUB-01
BtFC	DPA 4011	Sanken CUB-01
BtFR	DPA 4011	Sanken CUB-01
FrAmb L	N/A	Senn. MKH 800
FrAmb C	N/A	Senn. MKH 800
FrAmb R	N/A	Senn. MKH 800

Table 2. Microphones used per technique. For a detailed explanation of channel naming, see [5].

Technique 3 Placement and Optimization

Professional recording engineers tend to place microphones based on previous experience, known best practices, and most importantly, what they hear in the monitoring environment. Recording with the Eigenmike, as such, presents a unique set of challenges. There is little published information detailing placement and optimization strategies for music recording using spherical HOA microphones, especially where the desired sound scene utilizes the traditional "ensemble in front, ambience surrounding" perspective. Daniels discusses several experimental recordings done with spherical HOA microphones, mixed for two-dimensional playback [25]. For a large ensemble recording where the goal was to keep the ensemble imaged in front of the listener, Daniels placed a 20 capsule HOA sphere near several other (unidentified) 5.1 microphone arrays. Barrett [26] and Power [27] both used the Eigenmike for music recording as part of their respective studies, but provided no methodology for placement and/or optimization.

The 32-channel output from the Eigenmike is recorded to a computer running Eigenstudio software via firewire output from an mh acoustics EMIB termination box. There is no effective way to monitor a 22.2 rendering of these signals in real time. For this study, the beampattern of an omnidirectional microphone was sent from the Eigenstudio recording software to Studio 22 for monitoring. Though not ideal, this gave the recording team some degree of information (distance and balance of instrumental groups) for microphone placement optimization. The result was the Eigenmike being placed in the centre of Technique 1's "Decca Tree" (Fig 1).

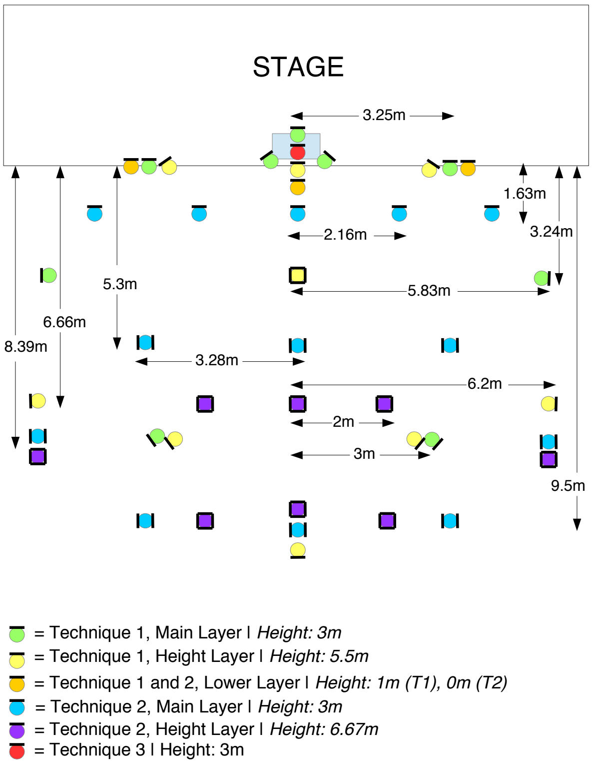

Figure 1. Microphone placement, overhead view. Height is referenced to stage floor.

4 Experimental Design

Creation of Stimuli

All three techniques simultaneously captured the final orchestral dress rehearsal: Techniques 1 and 2 were recorded to Pro Tools 10 at 96kHz/24bit resolution. Spot microphones for the woodwinds, bass and tympani were also recorded. Technique 3 was recorded to a separate laptop computer, whose audio interface was locked to the RME Micstacys' master clock. A single, 30-second musical excerpt was chosen as stimuli – the passage contains dense orchestration representative of the piece it was derived from (Tchaikovsky's 5th Symphony), and has a fairly even dynamic envelope.

The techniques under investigation were balanced by a team of three recording engineers with extensive professional experience recording and mixing orchestral music. It was observed that Technique 2 did not contain enough low frequency content for a satisfying mix, largely due to the low frequency roll-off typical of highly directional microphones. In accordance with [23], the FL and FR omnidirectional channels from Technique 1 were added to Technique 2's mix, low-passed at 200Hz. Once ideal balances were achieved, 24-channel mixes of the musical excerpt were made for each technique.

To create an optimal 22.2 mix of the Eigenmike recording, a custom-made decoder for the speaker positions in Studio 22 was built. By using the Ambisonic Decoder Toolbox by Heller [33] the decoder matrix for a dual band All-Round decoder [32] was calculated, which allowed for adjustment of balance between high and low frequencies with phase matched filters per [34]. The crossover frequency (400Hz) and the gain for the balance (+1dB HF) were chosen to perceptually match the mixes from Techniques 1 and 2.

The three resultant stimuli were level matched by ear. These results were then confirmed by objective means. A Neumann KU-100 Dummy Head microphone was placed in the listening position at ear level, and used to record the playback of each stimulus. Integrated loudness measures (LUFS 9) were then performed for each recording. All stimuli were found to be within 0.5dB of each other.

Design and Implementation of Listening Test

A double-blind listening test was designed to identify possible salient perceptual differences between the three techniques. The test was implemented using Cycling 74's Max/MSP software. Twenty-three subjects performed the test: all were either current students or faculty within the Graduate Program in Sound Recording at McGill University – all reported having normal hearing.

Subjects were seated in Studio 22's listening position, were explained the testing conditions, and given time to familiarize themselves with the testing interface and stimuli (Fig 2). Definitions of the perceptual attributes being investigated were provided both verbally and in written form (see: Appendix A). Based on previous research into spatial audio evaluation [28, 29], Clarity, Scene Depth, Naturalness, Environmental Envelopment and Sound Source Envelopment were chosen. "Quality of Orchestral Image", a new term, was included based on the sonic imaging goals of Technique 1.

For each trial, subjects were asked to evaluate mixes labelled "A", "B", and "C" for a given attribute, using a set of continuous sliders (0-100). Anchor words were provided at the extremes of each slider. Since absolute anchors were not given at intervals along the scales, these measurements are relative and not absolute. To reduce scaling bias, subjects were instructed to always rate the mix they felt was the "most" or "best" of a given attribute as 100%, then using that as a reference, rate the other two accordingly. More than one mix could be rated 100%. After completing ratings for a given attribute, the subject was asked to choose the mix they preferred, regardless of the perceptual attribute being investigated. Subjects could switch between playback of A, B, and C or stop the audio at any point – playback was continuously looped. The test was administered in blocks of three trials per attribute, for a total of 18 trials. This was done to allow subjects to focus on one perceptual attribute at a time. For each trial, stimulus assignments to A, B, and C were randomized. The order of trial blocks was also randomized. Subjects were instructed to set a comfortable listening level before completing the first trial, and then leave the level unchanged for the remainder of the test. At the test's midway point was an enforced rest period of 1 minute. Subjects took an average of 25 minutes to complete the test. Upon completion, subjects were instructed to fill out a short demographic survey.

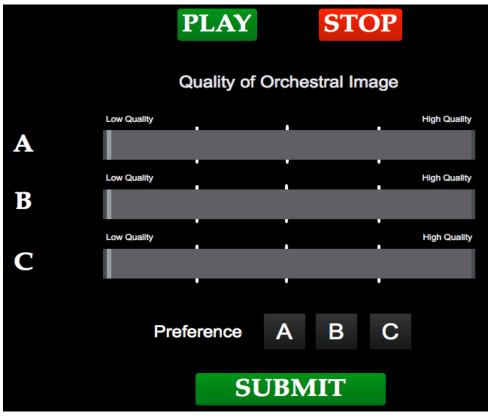

Figure 2: Testing GUI

5 Results

Attributes

The mean ratings, for each attribute, for each technique are shown in Figure 3. For all attributes, Techniques 1 and 2 were rated quite high and similar, whereas Technique 3 was rated quite low. The results of a one-way repeated measures ANOVA on each attribute can be seen in Table 3, and show that the differences seen in the ratings for each attribute is significant. Post-hoc Bonferroni corrected pair-wise t-tests show that significant differences exist between Technique 3 and both other techniques for all attributes. Post-hoc tests also show that a significant difference between Techniques 1 and 2 exists only for the attribute "sound source envelopment". It should also be noted that significance at the 95% level was almost achieved for the "naturalness" attribute as well.

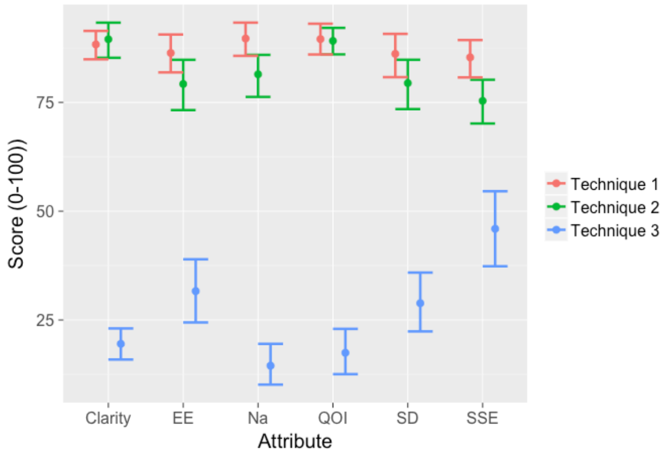

Figure 3: Average rating for each attribute. Colour represents the three different recording techniques. EE = environmental envelopment, Na = naturalness, QOI = quality of orchestral image, SD = scene depth, SSE = sound source envelopment.

Attribute	Tech1 Mean (SD)	Tech2 Mean (SD)	Tech3 Mean (SD)	F (df)	*p*	*p Tech1 vs Tech 2*
Clarity	88.3 (13.4)	89.5 (16.4)	19.5 (15.4)	184.3 (2, 44)	<.001	
EE	86.3 (19.0)	79.2 (23.4)	31.6 (30.1)	32.8 (2, 44)	<.001	
QOI	89.5 (15.5)	89.5 (15.5)	17.5 (22.9)	165.1 (2, 44)	<.001	
Na	89.7 (15.8)	81.5 (21.5)	14.5 (19.9)	190.7 (2, 44)	<.001	0.066
SD	86.1 (22.0)	29.4 (24.6)	28.9 (28.3)	31.1 (2, 44)	<.001	
SSE	85.3 (18.9)	75.3 (21.5)	46.0 (35.5)	14.9 (2, 44)	<.001	<.01

Table 3 ANOVA and Post-Hoc on Attribute Ratings

Preference

Preference for each technique was measured by counting the number of times a given technique was chosen as the most preferred (Table 4)

	Technique 1	Technique 2	Technique 3	Total
Count	243	167	3	413
% of total	59%	40%	0.7%	

Table 4 Contingency Table for Preference

A Chi-Square test shows that difference in preference is significant, $x^2(2) = 218.6$, $p < .001$.

Technique 3's large deviation from the expected random frequency (~33.3%) is likely the cause of this significant difference. Given that Technique 3 received so few preference counts, it can be dropped from the analysis, and a binomial test on the counts for Techniques 1 and 2 can be performed. In this case, Technique 1 was significantly preferred over Technique 2, $p < .001$ with a confidence interval of 0.54-0.64.

With this test design, attributes and preference were rated by the participants simultaneously. It is therefore important to know if the attribute being rated for a given trial influenced preference. In this case, the attribute being rated did not have a significant effect on the preference ratings $x^2(2) = 13.18$, $p = 0.21$.

Correlation of Attributes

Having the same three stimuli (techniques) rated along several different attributes allows for investigating rating correlation between said attributes. The results (Table 5) show that there is a high positive correlation between all pairs of attributes. The correlation coefficients are significant to at least the $p = 0.05$ level.

	Clarity	EE	QOI	Na	SD	SSE
Clarity	1.00	-	-	-	-	-
EE	0.75	1.00	-	-	-	-
QOI	0.92	0.72	1.00	-	-	-
Na	0.84	0.65	0.81	1.00	-	-
SD	0.67	0.53	0.72	0.58	1.00	-
SSE	0.49	0.41	0.50	0.37	0.54	1.00

Table 5 Pearson correlation matrix between attributes

The relationship between attribute ratings for each stimuli and preference ratings is visualized in Figure 4. It shows that when a given technique is preferred, it also receives much higher ratings along all attributes. The magnitude of this difference appears to be similar between all attributes.

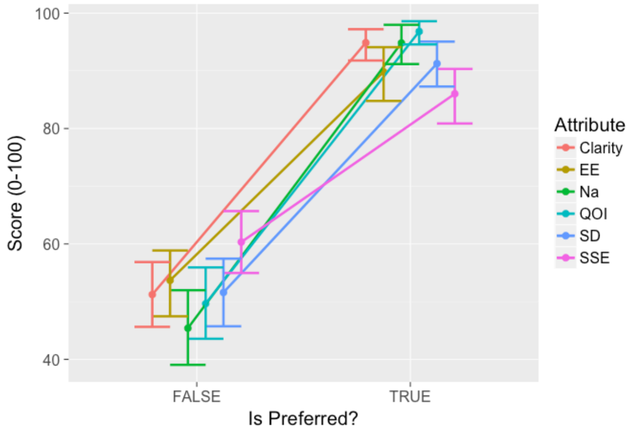

Figure 4: Average rating for each attribute according to preference. Colour represents the attribute.

6 Discussion

Overall Performance of Recording Techniques

Figure 3 shows a clear similarity of ratings between Techniques 1 and 2 for all the subjective attributes under investigation: clarity, scene depth, naturalness, environmental envelopment, sound source envelopment, and quality of orchestral image. Given this, and its consistently high mean scores across all attributes, the three-dimensional recording technique proposed in [1] should be considered a well-performing, valid production technique for three-dimensional orchestral music recording. Concepts from both Technique 1 and 2 could also easily be combined to form any number of hybrid techniques. For example, broadcast recordings involving picture would benefit from the bottom channel microphone design from Technique 2, which is more visually transparent.

Also very clear are the consistently low scores across all perceptual attributes for Technique 3. This matches well with the trend observed in previous research comparing two-dimensional recording techniques (see: introduction). For example, [17] observed a lack of depth and adequate spatial impression for the Soundfield MKV, a 1st order ambisonics recording system. These observations are echoed in the current study, with the Eigenmike performing poorly for Scene Depth, Environmental Envelopment and Sound Source

Envelopment. Spherical HOA microphones, although a convenient alternative to large spaced microphone arrays, may not yet be suited to professional 3D music recording, especially given the monitoring difficulties discussed in section 3.

Naturalness and Sound Source Envelopment

"Naturalness" appears frequently as a subjective attribute in multichannel audio evaluation, and has been shown to correlate strongly with the impression of "presence" [28] and overall preference of sound quality [36]. Frequently observed by both subjects and researchers were unpleasant and unnatural "out of phase" sonic artefacts present in Technique 3. This may explain why amongst all attributes, Technique 3's mean rating was lowest for "naturalness". In this study, a lack of perceived "naturalness" may also be an issue of perspective bias. Techniques 1 and 2 deliver a "cinematic" perspective for reproduced orchestral music, with the orchestra appearing entirely in front of the listener – a perspective most listeners have grown accustomed to. Technique 3, however, presents a much wider orchestral image, with direct sound sources covering almost 180° of frontal sound, likely due to the spherical nature of the Eigenmike. It is possible that the more "wrap-around" direct sound perspective delivered by Technique 3 is also perceived as being "unnatural".

Rumsey has written, "Envelopment, on the other hand, must be subdivided into environmental envelopment and source-related envelopment, the former being similar to LEV in concert halls and the latter to envelopment by one or more dry or direct foreground sound sources." [30] It was assumed that Technique 3's wider orchestral image would be rated highly for Sound Source Envelopment. However, although that attribute represented Technique 3's highest rated mean, it still scored well below Techniques 1 and 2. Clearly, listeners did not find a wider "wrap around" orchestral image to be more enveloping. In this study, the listeners' impression of Sound Source Envelopment may be closer to Griesinger's concept of Continual Spatial Impression [5] – a fusion of continuous direct sound and reflected energy that results in a sense of envelopment connected to the sound source. Technique 1 seems to best represent this type of spatial impression.

Cultural Bias in Preference

Technique 1 was created by a graduate of McGill University's Graduate Program in Sound Recording. That this technique was significantly preferred by current students and faculty within that same program could point to a strong bias within the results. Arrangements have been made to perform the same listening test at Tokyo University of the Arts to investigate possible cultural trends towards 3D microphone technique preference.

7 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 Julien Boissinot for his technical support, Kensuke Irie (NHK), and Maestro Alexis Hauser and the McGill Symphony Orchestra.

References

[1] W. Howie et al., "A Three-Dimensional Orchestral Music Recording Technique, Optimized for 22.2 Multichannel Sound," in *AES 141st Convention*, Los Angeles, USA, 2016, pp. 1-10.

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

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

[4] 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. 155-185

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

[6] 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.

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

[8] 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.

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

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

[11] 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.

[12] A. Ryaboy, "Exploring 3D: A subjective evaluation of surround microphone arrays catered for Auro-3D reproduction," in *AES 139th Convention*, New York, USA, 2015, pp. 1-10.

[13] 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.

[14] 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.

[15] R. Kassier et al., "An Informal Comparison Between Surround-Sound Microphone Techniques," in *AES 118th Convention*, Barcelona, Spain, 2015, pp. 1-17.

[16] M. Hietala, "Perceived differences in recordings produced with four surround microphone techniques," M.S. Thesis, Music, Mind & Technology, Univ. of Jyväskylä, Finland, April 2007, pp. 1-45.

[17] F. Camerer and C. Sodl. *Classical Music in Radio and TV - a Multichannel Challenge*. http://www.hauptmikrofon.de/stereo-3d/orf-surround-techniques, March 30 2015. The IRT/ORF Surround Listening Test.

[18] S. Kim et al., "An Examination of the Influence of Musical Selection on Listener Preferences for Multichannel Microphone Technique," in *AES 28th International Conference*, Piteå, Sweden, 2006, pp. 1-14.

[19] T. Kamekawa et al., "Correspondence Relationship between Physical Factors and Psychological Impressions of Microphone Arrays for Orchestra Recording," in *AES 123rd Convention*, New York, USA, 2007, pp. 1-13.

[20] M. Paquier et al., "Subjective assessment of microphone arrays for spatial audio recording." Forum Acusticum 2011, Jun 2011, Aalborg, Denmark. pp.2737-2742, 2011. <hal-00606210>

[21] A. Sitek and B. Kostek., "Study of Preference for Surround Microphone Techniques Used in the Recording of Choir and Instrumental Ensemble," *Archives of Acoustics*, Vol 36, No. 2, 2011, pp. 365–378

[22] N. Peters et al., "Recording Techniques and their Effect on Sound Quality at Off-Center Listening Positions in 5.1 Surround Environments," *Canadian Acoustics*, Vol. 41, No. 3, 2013, pp. 37-49.

[23] K. Hamasaki and K. Hiyama, "Reproducing Spatial Impression With Multichannel Music," in *AES 24th International Conference on Multichannel Audio*, Banff, Alberta, 2003, pp. 1-6.

[24] “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.

[25] J. Daniel, “Evolving Views on HOA: From Technological to Pragmatic Concerns,” in *Ambisonics Symposium 2009*, Graz, Austria, June, 2009, pp. 1-18.

[26] N. Barrett, “The Perception, Evaluation and Creative Application of Higher Order Ambisonics in Contemporary Music Practice,” *Ircam Composer in Research Report 2012*, pp. 1-6

[27] P. J. Power, “Future Spatial Audio: Subjective Evaluation of 3D Surround Systems,” PhD Dissertation, College of Science and Technology, University of Salford, Salford, UK [1].

[28] J. Berg and F. Rumsey, “Verification and correlation of attributes used for describing the spatial quality of reproduced sound,” in *AES 19th International Conference*, Schloss Elmau, Germany, 2001, pp. 1-19.

[29] 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.

[30] F. Rumsey, “Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm,” *J. Audio Eng. Soc.*, Vol. 50, No. 9, 2002, pp. 651-666.

[31] M. Chapman et al., “A standard for interchange of Ambisonic signal sets”, in *Ambisonics Symposium 2000*, Graz, Austria, 2009.

[32] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," J. Audio Eng Soc, vol. 60, no. 10, pp. 807–820, Nov. 2012.

[33] A. Heller and E. M. Benjamin, "The Ambisonic Decoder Toolbox: Extensions for Partial-Coverage Loudspeaker Arrays," in *Linux Audio Conference 2014*, Karlsruhe, 2014.

[34] A. Heller, R. Lee, and E. M. Benjamin, "Is My Decoder Ambisonic?," in *AES 125th Convention*, San Francisco, USA, 2008. pp. 1–21.

[35] S. Choisel and F. Wickelmaier, “Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference”, *JASA*, Vol 121, No. 1, 2007 pp. 388-400.

[36] R. Mason and F. Rumsey, “An assessment of the spatial performance of virtual home theatre algorithms by subjective and objective methods,” in *AES 108th Convention*, Paris, France, 2000, pp. 1-39.

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

Appendix A: Attribute Definitions

Sound Source Envelopment: “The sense of being enveloped by a group of sound sources.” [30]

Environmental Envelopment: “The sense of being enveloped by reverberant or environmental sound.” [30]

Clarity: “The clearer the sound, the more details you can perceive in it.” [35]

Naturalness: “A sound is natural if it gives you a realistic impression, as opposed to sounding artificial.” [35]

Quality of Orchestral Image: A “high quality” orchestral image is defined as being a cohesive, anchored sound image, with well-defined horizontal and vertical extent.”

Scene Depth: “The overall impression of the depth of the sound image. Takes into consideration both overall depth of scene, and the relative depth of the individual sound sources.” [29]