Audio Engineering Society

Convention Paper 9251

Presented at the 138th Convention
2015 May 7–10 Warsaw, Poland



# Immersive Content in Three Dimensional Recording Techniques for Single Instruments in Popular Music

Bryan Martin[1,2], Richard King[1,2], Brett Leonard[3], David Benson[1,2] Will Howie[1,2]

[1] The Graduate Program in Sound Recording, McGill University, Montreal, QC H3A 1E3, Canada
[2] Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), Montreal, QC, H3A 1E3, Canada
[3] The School of Music, the University of Nebraska at Omaha, Omaha, NE.

Correspondence should be addressed to Bryan Martin, (bryan.martin@mcgill.ca)

## ABSTRACT

“3D Audio” has become a popular topic in recent years. A great deal of research is underway in spatial sound reproduction through computer modeling and signal processing, while less focus is being placed on actual recording practice. This study is a preliminary test in establishing effective levels of height-channel information based on the results of a listening test. In this case, an acoustic guitar was used as the source. Eight discrete channels of height information were combined with an eight-channel surround sound mix reproduced at the listener’s ear height. Data from the resulting listening test suggests that while substantial levels of height channel information increase the effect of immersion, more subtle levels fail to provide increased immersion over the conventional multichannel mix.

## 1. INTRODUCTION

The focus of this research is to investigate recording techniques that will create three-dimensional images of individual instruments in surround playback loudspeaker arrays that include height channels. While certain recent developments in surround sound presentations have been focused on live concert recording, outdoor and sporting events, or the examination of theoretical techniques [7, 11, 12, 14]. The techniques described herein are aimed at developing tools for use in recording studio applications, specifically for single instruments in popular music.

This study is a preliminary test in establishing playback levels of height-channel information that are considered to be effective. Eight discrete channels of height information were presented in conjunction with an eight channel discrete multichannel mix of solo acoustic guitar. The latter is presented in one horizontal plane at the listener’s ear level as front L/C/R, rear L/C/R, and side channels positioned at +/-90 degrees (Fig. 1). The ring of height channels copies the number and placement of the first ring of loudspeakers, positioned 1.5 meters above. This configuration comprises the middle and top layers of the 22.2 SMPTE standard 2036-2 [8, 10] that was developed by the Japanese broadcaster NHK [4].

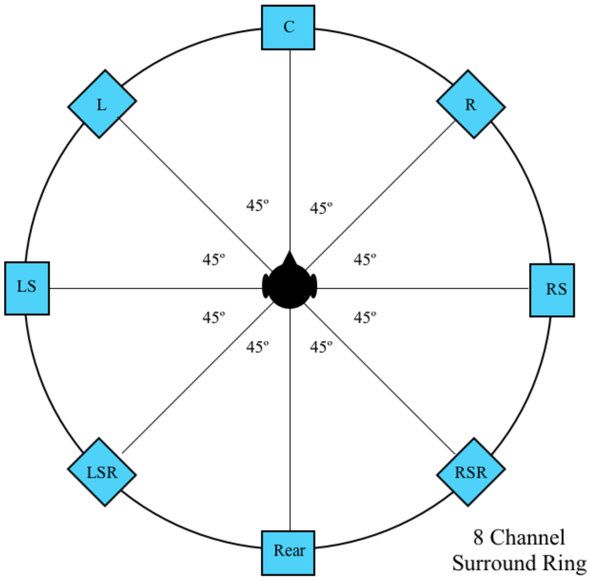


Fig. 1: Eight Channel Surround Loudspeaker Ring. (Height channels copy first ring of loudspeakers 1.5 meters above).

## 2. CONTEXT

Traditionally, the literature has abundant references to several well-adopted stereo recording techniques such as those referenced by Woram [15], but these techniques that are optimized for two-channel playback do not directly relate to the practice of 3D audio recording. Multichannel, or surround sound recording, has been commonplace since the mid to late 1990s, and although a few techniques have been described [1, 2, 6], most individual practitioners in the recording industry seem to have their own solutions for recording and mixing in this format. Moreover, the few references in the literature offer rather generic solutions to 3D audio capture, which are not optimized to specific instruments [3, 5, 9, 13, 14].

In the rather limited amount of literature that is dedicated to 3D audio presentation, there are very few references to actual 3D recording techniques. There is a general lack of research in this area, and it is the intent of this investigation to correct this void in the literature.

## 3. TEST DESIGN

### 3.1. Ambient Recording Configuration

For the ambient audio component of the study, the experiment was designed such that microphones be placed in the recording studio with positioning and spacing that mirrored the number of loudspeakers in each playback "ring" of the control room; i.e. eight microphones for each ring of eight loudspeakers; sixteen microphones in total.

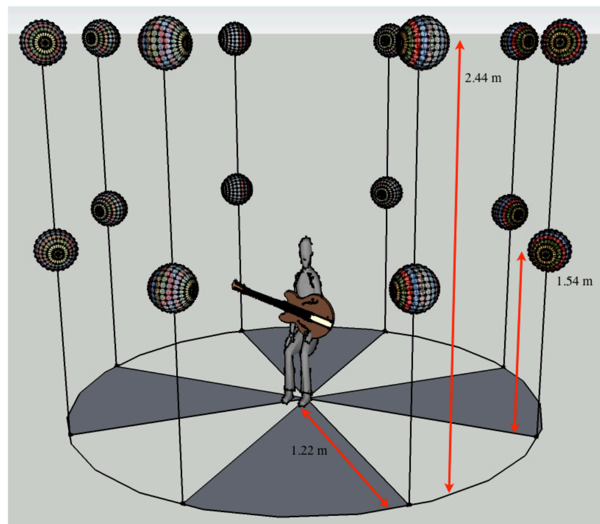


Fig.2: Recording studio microphone positions corresponding to control room loudspeaker position.

In this test, an acoustic guitar was recorded in the center of the studio. In early test recordings, several distances to the source and microphone heights were compared. The authors decided on a medium distance of ambient capture for this pilot study. The radius decided upon was 1.22 m. The mid-microphone ring was placed 1.54 m from the floor (corresponding to the control room mid-ring loudspeaker height), and the high ring was positioned at 2.44 m (Fig. 2). All microphones were pointed at the guitar. An additional close microphone was used to capture the direct sound of the instrument, as would be common in popular music production. This microphone was placed for optimum sound quality as determined by the recording engineer, and carefully balanced into the center channel of the multichannel mix presented at ear level. The microphones used for this recording are listed in Table 1. All were cardioid or sub-cardioid types.

All microphones were recorded through a Sony SIU 100 interface which provided microphone pre-amplification

(DMBK-S101 cards) and A/D conversion. Careful attention was paid to match the input gains of the sixteen microphones in the mid and high rings. The session was recorded at a 96kHz sample rate.

| | Mid Ring | High Ring |
|---|---|---|
| L | Schoeps CMC62U / MK 4 | Schoeps CMC62U / MK 21 |
| C | Schoeps CMC62U / MK 4 | DPA 4011-TL |
| R | Schoeps CMC62U / MK 4 | Schoeps CMC62U / MK 21 |
| LS | Schoeps CMC62U / MK 4 | ADK HA-TL-II Cardioid |
| RS | Schoeps CMC62U / MK 4 | ADK HA-TL-II Cardioid |
| LSR | Schoeps CMC62U / MK 4 | Schoeps CMC62U / MK 21 |
| REAR | Schoeps CMC62U / MK 4 | DPA 4011-TL |
| RSR | Schoeps CMC62U / MK 4 | Schoeps CMC62U / MK 21 |
| Close | ADK C-LOL-67 capsule, ADK HA-TL-II body | |

Table 1: Microphones used in recording.

### 3.1.1. Recording Studio

The recording studio was rectangular in shape (11 m x 7 m) with an RT60 of just over 1 second (Fig. 3, 4). The ceiling height was 5.7m. The wall treatment was a combination of absorption and diffusion with the upper walls and ceiling being more reflective.

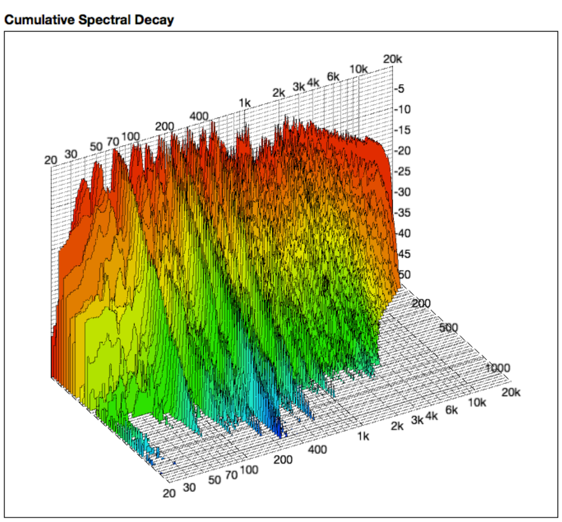


Fig. 3: Spectral plot of recording studio.

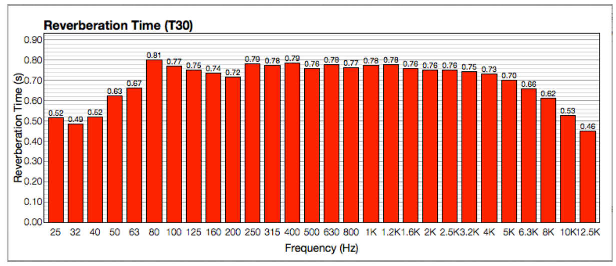


Fig. 4: Reverberation time (RT30) of recording studio.

### 3.1.2. Testing Environment

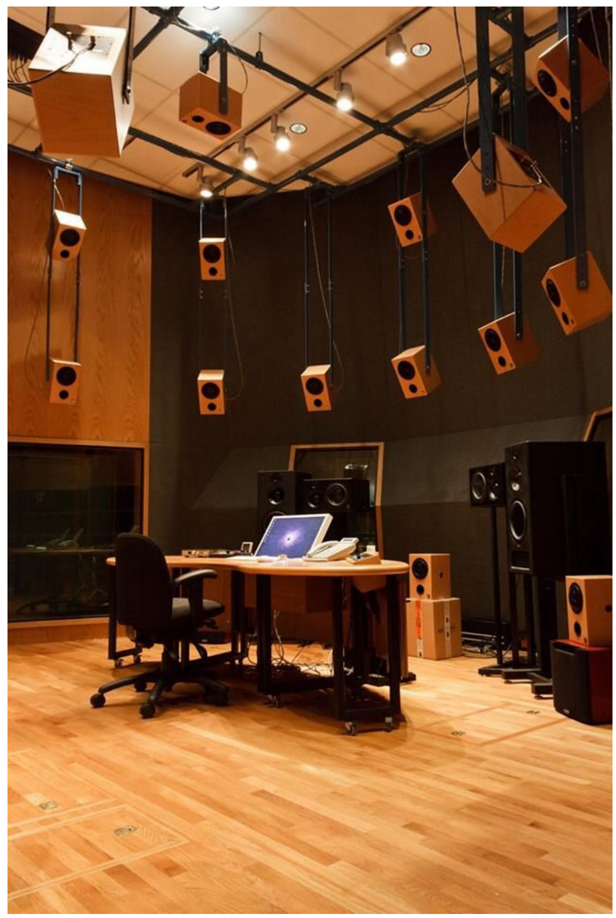

Fig. 6: Schulich School of Music's Studio 22.

This research was conducted at McGill University Schulich School of Music's Studio 22 (Fig. 5). This studio is optimized for multichannel recording and playback, with up to 30 discrete channels and loudspeakers available in the control room.

Studio 22 is a music mixing control room with an RT60 of 200 milliseconds ±50 ms at all frequency bands (Fig. 6). This is coupled with a full-range playback system comprised of two-way loudspeakers. This room/reproduction system displays a flat response, with a level deviation of ±3dB between 20 Hz and 18 kHz.

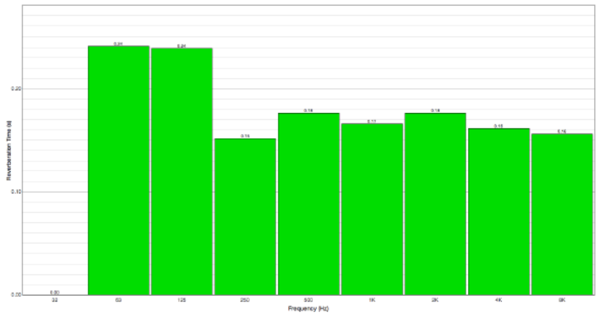

Fig. 5: Studio 22 reverberation time (T30). The lowest octave band is excluded due to noise introduced by the measurement computer.

### 3.1.3. Musical Material

The music used for testing was played on a 1959 Harmony Monterey arch-top acoustic guitar. A cyclical chord progression in E major was played using primarily open voicing.

## 3.2. Testing Software & Methodology

Testing was achieved using a software patch developed in Max MSP. The testing software managed audio playback, data collection, and treatment order shuffling.

The treatments to be evaluated for immersion consisted of eight discrete channels of height information presented at five different volume levels, in conjunction with an eight channel discrete multichannel mix of solo guitar. The five height channel levels in dB were 0, -6, -16, -22, and -144 (no signal). These levels were determined by the authors and a select group of expert listeners to be fairly equal steps between "full immersion" and "very subtle" immersion. All five upper ring levels were presented randomly in combination with the main ring, and without repetition in each subsequent trial. Listener-ranked preference and treatment presentation order was captured in the resulting data.

The software patch presented users with a graphical user interface (GUI) allowing for basic control of audio playback, as well as a rating "slider" in order to rate each treatment's "immersiveness", as experienced by the listener. The GUI's sliders were completely without scale or numeric indicators and were labeled only as "less immersive" and "more immersive" from left to right (Fig. 7). Each slider allowed for an immersive rating of 0-100 and the default starting position for each trial was 50. Additionally, users were able to vary any slider in a given trial, regardless of which of the five levels was selected for playback, allowing for flexibility in adjustment during the test.

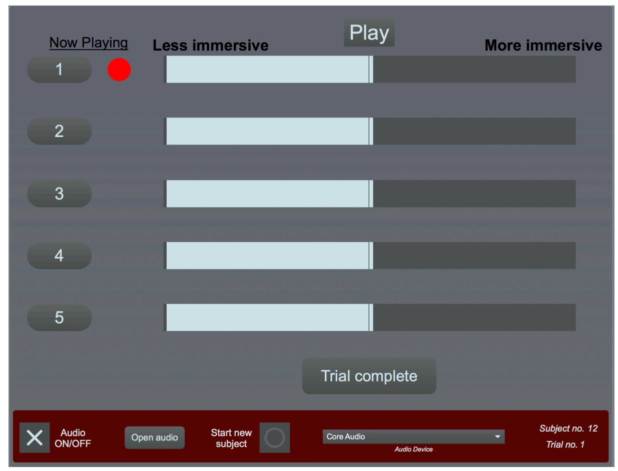


Fig. 7: Graphical User interface.

### 3.2.1. Subjective Preference Question

In addition to rating the treatments for immersiveness in the user interface, the subjects were asked to select a personal preference from each of the five treatments presented in each trial. These preferences were noted on a questionnaire provided to each subject.

## 3.3. Subjects

Thirty test subjects were drawn from the students and staff of the graduate program in Sound Recording at McGill University. All subjects had significant musical training, averaging more than 14 years; and averaging over 9 years of experience in music recording and production. The subject pool was composed of individuals specializing in recording, production, and mixing.

# 4. RESULTS AND ANALYSIS

## 4.1. Immersion Ratings

An analysis of variance was performed on the immersion ratings elicited by each of the five height channel levels. Prior to the analysis, the normality of each group was verified using a one-sample Kolmogorov-Smirnov test. All five showed as normal. Differences were found between the group means ($p < 0.05$). The means for each height level were 0 dB: 76; -6 dB: 62; -16dB: 46; -22dB: 45; -144 dB: 45. Tukey's HSD test revealed a significant difference between the immersion of the 0 dB and -6 dB height channel levels. The three lower levels (-16, -22, -144) were significantly different from the two higher levels, but were not different from each other (Fig. 8).

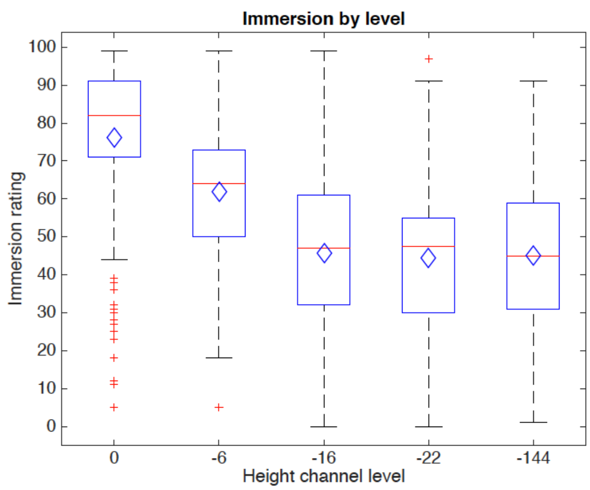


Fig. 8: Immersion ratings grouped by height channel levels.

## 4.2. Preference Question Data

To analyze preference choices, the five height channel playback levels were split into two groups. The first group was deemed "non-immersive" and consisted of the three height levels that produced no differences in immersion ratings (-16, -22, and -144dB). The second group was deemed "immersive", and consisted of the two choices linked with high immersion (0 and -6 dB).

Immersive stimuli were preferred significantly more often than non-immersive stimuli ($p < 0.05$, binomial test, Fig. 10). Data were excluded from one subject who forgot to complete the questionnaire

## 4.3. Subject Consistency Scores

Subjects varied considerably in the consistency of their preferences. Some subjects chose the same immersion level repeatedly from trial to trial, while others shifted their preferences over the course of the test.

The consistency of each subject was gauged by the variance in his or her preferences. To measure consistency, each preference choice was assigned to one of three groups, and each group was associated with a numeric immersion level. Immersion level 0 contained the 0 dB height channel choice; immersion level -1 the -6 dB choice; and immersion level -2 the -16 dB, -22 and -144 dB choices. Variance in height-choice levels was then calculated for each subject. This variance, multiplied by -1, was referred to as the subject's consistency score (Fig. 9).

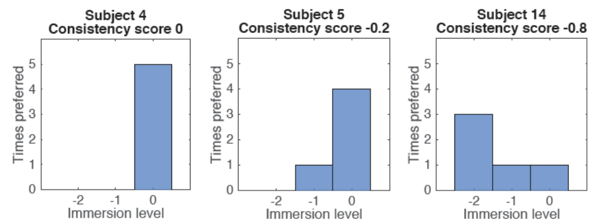


Fig. 9: Consistency Scores: Examples of three subjects exhibiting high, medium and low consistency scores.

## 4.4. Preferences for Consistent Subjects

Consistency scores were used to divide the subjects into consistent and inconsistent groups. A score of -0.3 was used as a cutoff. Nineteen subjects were at or above this cutoff and deemed consistent. Ten subjects were below cutoff, and deemed inconsistent.

In examining the consistent group, a trend toward preference for immersion became clearer. Among consistent subjects immersion was preferred in 77% of trials, versus 68% in both groups combined (Fig. 10). Both results were significant ($p < 0.05$, binomial test).

The statistical results of the Immersion Ratings test provided significant results for the perception of immersive content, and provides a baseline for the minimum level at which height channels can be perceived in this particular test scenario.

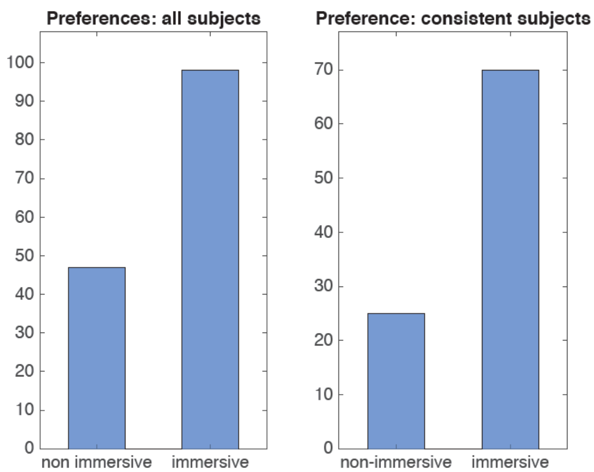


Fig. 10: Preference choices for all subjects: Immersive stimuli were preferred in 98 out of 145 trials.

## 5. CONCLUSION

The findings of this study were:

1) There is a minimum level of height information below which subjects could not differentiate added height content. These levels, -16dB, -22dB, provided the same perceived immersion as the mid eight-channel loudspeakers with no additional immersive content (-144dB).

2) The subjects could perceive three distinct content levels during testing: 0dB (immersive), -6dB (immersive), and the “-16, -22, and -144dB” group (little or no-immersive-content).

3) The level of the immersive content needs to be substantially louder to be perceived, ≥ 10dB.

4) The Preference Question results suggest that subjects preferred a more immersive environment than the more subtle levels of immersion, when given the choice.

## 6. POSSIBILITIES FOR FUTURE WORK

The next step in this research will be to provide users with a level control to set their preferred level of height channel content. Research will also continue in developing microphone arrays and recording techniques that provide stable 3D images for popular music mixing and reproduction.

## 7. ACKNOWLEDGEMENTS

This work was supported by McGill University, Schulich School of Music Department of Sound Recording, Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT), and Fonds de recherché Société et culture Québec (FRQSC).

## 8. REFERENCES

[1] Corey, Jason; Woszczyk, Wieslaw. "Localization of Lateral Phantom Images in a 5-Channel System with and without Simulated Early Reflections." In *113 AES Convention*. Los Angeles, CA: Audio Engineering Society, 2002.

[2] Gebhardt, Mario; Kuhn, Clemens; Pellegrini, Renato. "Headphones Technology for Surround Sound Monitoring – a Virtual 5.1 Listening Room." In *122 AES Convention*. Vienna, Austria: Audio Engineering Society, 2007.

[3] Geluso, Paul. "Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays." In *Audio Engineering Society 132nd Convention*. Budapest, Hungary: Audio Engineering Society, 2012.

[4] Hamasaki, Kimio; Hiyama, Koichiro; Okumura, Reiko. "The 22.2 Multichannel Sound System and Its Application." In *AES Convention:118*. Barcelona, Spain: Audio Engineering Society, 2005.

[5] Hamasaki, Kimio; Shinmura, Toru; Akita, Shoji; Hiyama, Koichiro. "Multichannel Recording Techniques for Reproducing Adequate Spatial Impression." In *AES Conference:24th International Conference: Multichannel Audio, The New Reality*. Banff, Alberta, Canada: Audio Engineering Society, 2003.

[6] ———. "Approach and Mixing Technique for Natural Sound Recording of Multichannel Audio." In *AES Conference:19th International Conference: Surround Sound - Techniques, Technology, and Perception*. Schloss Elmau, Germany: Audio Engineering Society, 2001.

[7] Hinata, Tsuyoshi; Ootakeyama, Yuichi; Sueishi, Hiromi. "Live Production of 22.2 Multichannel Sound for Sports Programs." In *AES 40th*

*International Conference: Spatial Audio: Sense the Sound of Space*. Tokyo, Japan: Audio Engineering Society, 2010.

[8] International Telecommunications Union,. "Itu-R Bs.2159-4 Multichannel Sound Technology in Home and Broadcasting Applications." Geneva, Switzerland, 2012.

[9] Ono, Kazuho; Nishiguchi, Toshiyuki; Matsui, Kentaro; Hamasaki, Kimio. "Portable Spherical Microphone for Super Hi-Vision 22.2 Multichannel Audio." In *AES Convention:135*. New York, NY, USA: Audio Engineering Society, 2013.

[10] Society of Motion Picture & Television Engineers,. "Ultra High Definition Television — Audio Characteristics and Audio Channel Mapping for Program Production." White Plains, NY: Society of Motion Picture & Television Engineers, 2008.

[11] Stenzel, Hanne; Scuda, Ulli. "Producing Interactive Immersive Sound for Mpeg-H: A Field Test for Sports Broadcasting." In *137 AES Conference*. Berlin, Germany: Audio Engineering Society, 2014.

[12] Theile, Günther; Wittek, Helmut. "Principles in Surround Recordings with Height." In *130 AES Convention*. London, UK: Audio Engineering Society, 2011.

[13] Williams, Michael. "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 *Audio Engineering Society 134th Convention*. Rome, Italy: Audio Engineering Society, 2013.

[14] ———. "Microphone Array Design for Localisation with Elevation Cues." In *Audio Engineering Society 132nd Convention*. Budapest, Hungary: Audio Engineering Society, 2012.

[15] Woram, John. *The Recording Studio Handbook*. Plainview, NY: ELAR Publishing, 1982.