



Audio Engineering Society Convention Paper 9431

Presented at the 139th Convention
2015 October 29–November 1 New York, USA

This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library, <http://www.aes.org/e-lib>. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Exploring 3D: A subjective evaluation of surround microphone arrays catered for Auro-3D reproduction system

Alex Ryaboy¹,

¹ New York University, New York, N.Y., 10012, USA
ar1425@nyu.edu

ABSTRACT

As multi-channel systems grow in popularity, audio professionals must make an informed decision when choosing a correct capturing method to deliver their vision. Many of today's microphone arrays that are catered for surround sound with height, employ traditional spaced surround techniques that are aided by an additional array in the upper plane and are widely used to capture a performance in large spaces. This paper uses a perceptual study to evaluate a fully coincident microphone array Double-MSZ and a semi-coincident array Twins Square on Envelopment, Localization and Spatial Impression in a small recording studio environment. The study revealed overall lower widths, better localization, and a more stable vertical imaging for Double-MSZ, while the Twins Square technique exhibited higher ensemble envelopment, and a more spacious perceived environment.

1. INTRODUCTION

1.1. Multi-channel Audio

As home-theater systems grow in popularity, many manufacturers expanded the amount of channels to accommodate seven and even nine channels of separate audio. Several of these multi-channel audio formats started employing height channels. Auro-3D, is a 9.1 multi-channel system introduced by Wilfried Van Baelen, that uses four additional height channels in addition to the standard 5.1 surround sound layout, in order to provide a 3rd (height) dimension to the soundstage and dramatically enhance the listening

experience. The main advantage of this configuration is that the entire upper portion is designated for reproduction of early reflections and for reproduction of spatial diffusion and also allows for creating phantom imagery in the vertical plane [1].

Currently, many microphone arrays that are used for multi-channel audio with height channels use traditional surround sound arrays aided by an additional microphone array placed in the upper plane for capturing height information. OCT-9 is a recommended system for Auro-3D that uses OCT-Surround array with four additional spaced super-cardioid microphones placed directly above the Left, Right, Surround Left, and Right channels. Wide A/B is another system

recommended for Auro-3D, which uses nine omnidirectional microphones, a microphone for each channel, positioned in the same fashion as the Auro-3D speaker configuration [2]. Both of these microphone arrays are using spaced microphone approach. Spaced, or time-of-arrival techniques, are based on the time delay differences caused by the different paths from the sound source to the microphones, while in coincident microphone techniques, the microphones are positioned as close to each other as possible and the stereo image is created by differences in sound pressure levels. Although the aforementioned arrays may be suitable for large concert halls, they may not work in smaller spaces and recording studios that don't have large live rooms. Consequently, the two microphone techniques examined in the study were a fully coincident Double-MSZ array and a semi-coincident array Twins Square.

Double-MSZ technique is compatible with B-Format Ambisonics, and captures four signals WXYZ, also called 'B-Format', which represent omnidirectional pressure and figure of eight pressure gradient in horizontal and vertical planes, creating a sphere that can be manipulated, using similar principles as the MS decoding.

Introduced by Gregor Zielinsky, Twins Square array is based around a unique Sennheiser MKH800 Twin multi-pattern microphone, and combines coincident as well as time-of-arrival approaches. Unlike traditional multi-pattern microphones that combine signals from two transducers into a single output, the Twins' transducers can each be accessed separately, which not only provides the ability to route each output to different channels, but it gives the mixing engineer an ability to decide during post production, which pick-up pattern is better suited for mixing. The microphone array utilizes four Twin microphones arranged in a rectangular fashion, with two microphones mounted at different heights but on the same depth plane at a wide AB distance. Consequently, the front-facing capsules provide signals for the front facing speakers, while the capsules facing the back provide information for the rear channels. An additional cardioid microphone is placed in between of the bottom two microphones for the center channel. What makes this microphone technique unique is that it is coincident from front to rear, and spaced in the vertical and horizontal planes.

In concert realism perspective, the aim of the surround microphone array in addition to properly capturing direct sound is to capture correct spatial attributes of the recording space. Direct sound influences the correct

directionality of the source while early reflections are one of the main attributes that determine the perception of depth, distance, and special impression [1]. Although spaced and coincident approaches are aimed at achieving these goals, the resulting sound image is somewhat different. One of the more studied aspects on this topic is the addition of height channels, and correct placement of microphones to capture information feeding those channels. It has been observed in studies researching spatial hearing that the horizontal stereo principles, such as time and amplitude differences may not be applicable in the vertical plane. For example, panning of phantom images in the vertical plane is unstable [3] and localization in the vertical plane relies more on spectral cues rather than inter-aural time and amplitude cues [4]. Similarly, some of the principles guiding horizontal stereo recording techniques may not be applicable when microphones are spaced vertically. A recent study, examining overall preference and spatial impression of a spaced/coincident microphone approaches by comparing four different levels of the vertical spacings, showed that microphones configured in coincidental fashion were not only preferred, but produced greater spatial impression for transient rich material, and there were no significant differences between spaced layer configurations. In his paper 'Channel-Z' [5], Geluso introduces a coincident technique that uses a figure-of-eight microphone oriented upwards combined with traditional stereo arrays. His preliminary research demonstrated that Channel-Z was rated better for 3D imaging criteria compared to recordings where height channel signals contained up-mixed height signal.

1.2. Spatial Audio Evaluation

Subjective evaluation of audio is common practice in the field of music technology where some of characteristics that are evaluated are timbral quality, width and depth of soundstage, image stabilization, and localization. When it comes to multi-channel audio, new characteristics arise, such as envelopment, spaciousness and spatial impression. Evaluating these characteristics is a somewhat challenging task, as these qualities are subtle, and difficult to quantify. Current perceptive studies that investigate these elements tend to conflate several of the aforementioned characteristics into a single rating. For example, in a study that tested effects of vertical microphone spacing subjects evaluated 'Overall Spatial Impression', a term pre-defined by the researchers, 'as encompassing all possible spatial percepts from three dimensions for the

environment as well as the source components' [6]. Although such studies may yield conclusive results about preference and perceived differences of recordings, they produce little information about specific characteristics of perceived audio experience by the subject. For example, when a recording is being evaluated for envelopment, envelopment may be perceived as a single immersive sound source, or an enveloping environment. Similarly, a recording where the perspective is such where one is surrounded by musicians, may also be considered enveloping [7].

Francis Rumsey defines Spatial Impression as 'an auditory perception of the location, dimension and parameters of a sound source and acoustic environment in which the source is located'. In his paper 'Spatial Quality Evaluation for Reproduced Sound' Rumsey, proposes an approach to evaluating spatial impression by dividing the listening experience into hierarchical scenes ranging from individual instrument sources (micro-scenes) to spaces and environments (macro-scenes) [7]. The present study uses Rumsey's scene based recommendations, and focuses on evaluating envelopment, localization and spatial impression based on three dimensions, width, depth and height.

2. EXPERIMENT

2.1. Recording

The recordings took place at NYU's James L. Dolan Studio. The musical program of choice was a five-piece Klezmer band consisting of Violin, Accordion, Clarinet, Voice, Upright Bass, and marching band style Drum. The band was chosen for its distinct instrumentation,

which can be easily identified in a recording. A multi-capsule Josephson C700s microphone with an addition of Sennheiser MKH800 was used for the Double-MSZ array. C700s microphone features three capsules, omnidirectional (W), front-facing figure-of-eight (X), and a side-facing figure-of-eight (Y) each having its own separate output. MKH800 provided the Z capsule by operating in figure-of-eight mode and facing upwards. The array was placed in the center of the room at about 3.5 feet off the floor. Twins Square array was implemented using four MKH800 Twin mics, with both arrays sharing the same center channel. The microphones were spaced about 10 feet apart; the spacing between the lower and the upper plane mics was 5 feet (1.5 meters). The lower plane microphone capsules were on the same plane as the Josephson middle capsule, with Double-MSZ array centered between left and right Twin microphones (Figure 1). The band was arranged in a wide semi-circle, and the arrays were placed about 6 feet in front of the band.

2.2. Reproduction System

The listening test took place at the NYU Music and Audio Research Laboratory. The lab is a dry listening room, equipped with sixteen identical Genelec 8030A active speakers that can be arranged for a number of different speaker configurations. The speaker arrangement of choice was a nine-channel Auro-3D layout (Figure 2). Subwoofer was not used in the evaluation. Playback channels from the Double-MSZ array were derived using the arithmetic in the below table.

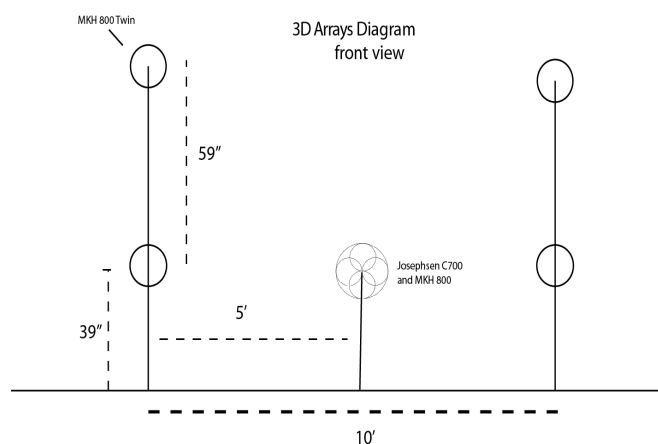


Figure 1

Left	$W+X+Y$
Right	$W+X-Y$
Center	$W+X$
Left-Surround	$W-X+Y$
Right-Surround	$W-X-Y$
Left-Height	$W+X+Y+Z$
Right-Height	$W+X-Y+Z$
Left-Height-Surround	$W-X+Y+Z$
Right-Height-Surround	$W-X-Y+Z$

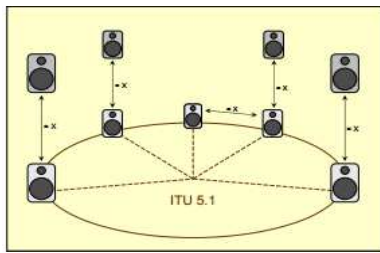


Figure 2 speakers were operating. Subjects were seated in a marked sweet spot of the room in front of a small table, with the questionnaire sheets and a computer mouse to control the graphic user interface. Head movement was not restricted, but the subjects were instructed to not make any drastic changes to the listening position. The gain levels for each speaker were calibrated using pink noise and SPL meter measuring 76db at listening position.

2.3. Test Stimuli

The test stimuli were two tunes from the aforementioned recording session recorded using two techniques, Double-MSZ and Twins Square (Twin). The subjects were presented with a total of 20 clips of music at random that they were asked to evaluate. The clip population consisted of 14 distinct clips (seven 30-second passages extracted from the two tunes recorded using 2 techniques), plus additional 6 of those clips were chosen as 'duplicates' and were randomly placed in the 20-clip population. Duplicates were used to observe how subjects evaluated the same exact piece of audio. Tune 1 consisted of Bass, Violin, Accordion and Voice, and Tune 2 consisted of Bass, Accordion, Violin, Clarinet and Percussion. The two resulting recordings were level-matched, as the Double-MSZ clips were slightly louder. A simple graphic user interface (Figure



Figure 4

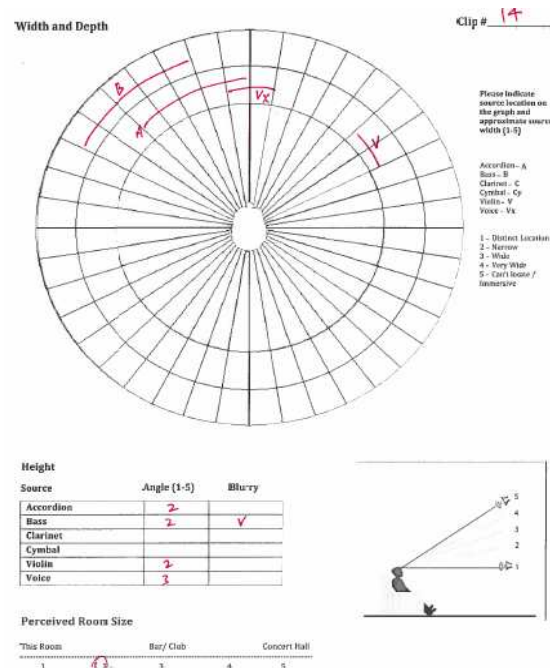
3) was created in MATLAB to accommodate multichannel playback generated by the subject. The GUI consisted of 20 buttons corresponding to each clip, a button to stop audio and buttons for three Demo clips.

2.4. Test Method

17 subjects participated in the study; all came from musical background, most having some experience in

critical listening. The subjects were asked to identify all sources in a short clip of music, and for each source evaluate apparent source width, source location, azimuth angle, and perceived room size. For the experiment, the sources were defined as instruments. While listening to a clip, the subjects were to identify the sources and mark their location on a simple graphical scale (Figure 4), with the appropriate perceived width. The widths were to be marked with notches on the paper indicating widths from 1-5 using the scale given in the instructions. To better control the variables, and to pose boundaries within which the subjects can express the width, a visual Likert scale was created, which ranged from (1-Distinct location) to (5-Immersive), the visual scale also helped align the subjects descriptive definitions with a visual angle. If the source was rated 'immersive' the subjects were instructed to mark (5) and source letter outside of the graph. Three levels of depth were indicated by middle, inner and outer circles of the

Figure 3



graph. The lower-plane front speakers served as a visual anchor and three levels of depth were described as 'at speaker distance', 'behind the speakers' and 'in front of speakers'. Consequently, a single mark on the graph contained four pieces of information, the source, its width, depth, and location. If the subject left out a source, their answers would still be accepted. Height was presented as intervallic Likert scale, however instead of given descriptive terms the measurement

relied on the visual approximation of azimuth angle between the listener and the source. The speakers at the ear-level and upper-plane served as visual anchors, with angle between them (about 35°) divided into 5 levels of height. The height was to be marked on a table with a number (1-5). Since previous studies stated that phantom imagery in the vertical plane is unstable [3], there was an option for subjects to indicate this by including a 'Blurry' checkbox. In order to find out which factors influence the perceived listening environment, and how micro-scenes and macro-scenes are related the subjects were asked evaluate perceived room size, which served as a variable that encompasses all three dimensions. Room-size was measured on a five-level Likert scale. Three distinct room descriptors were chosen from smallest to largest, 1-This Room, 3-Bar/Club, 5-Concert Hall, with 2 and 4 serving as in-between approximations. 1-This Room variable meant the size of the Research Lab which was used to show an extremely intimate space and smaller degree of spaciousness as the listener is experiencing a 'they are here' type of perspective, whereas Concert Hall was used in order to see whether subjects would experience an exaggerated perspective. Although it would be difficult to evaluate the exact size of the room that the listeners perceived, this way of measuring would demonstrate which recording technique provided a level of spaciousness that transforms them to a different listening space.

An instructional sheet was provided emphasizing above-mentioned instructions. When subjects arrived to the test, they were given a short tutorial session and a thorough explanation of what was required of them. 'Demo' clips were used to get subjects acquainted with the interface and the nature of the audio they were about to experience. Two of the demo clips consisted of clarinet playing a short solo passage from different perspectives, while the last demo clip included a short passage of the entire ensemble playing an excerpt from one of the tunes, which was not included in the main study. The first four clips (1-4) were used as training, and were excluded from the study. The subjects had the ability to listen to each clip as many times as they needed, however they were instructed to listen to each clip three times at most and give their initial impressions rather than listening in critically. Once the instructions were given, the subject was left in the room on their own for the remainder of the test. On average, the test took about 40 minutes to complete. After the test, an informal interview was conducted inquiring about the overall experience.

3. RESULTS

A total of 272 clips were evaluated (17 subjects x 16 questionnaires). Once the testing was completed the results were transferred into an OLAP cube multi-dimensional database using Microsoft Excel 2013 and a PowerPivot add-on. This allowed for great flexibility in maneuvering data and viewing results using a pivot-table layout. Several approaches were used in order to identify immediate outliers, visual inspection of the questionnaire to make sure the subjects adhered to the directions of the study, and comparison of the subjects' answers based on identical clips. If subjects' answers varied significantly between three identical pairs of clips they were excluded. Consequently three subjects were excluded from the study. Each dimension was assigned numerical value Width(1-5) Depth(1-3) Height (1-5), Room size(1-5), the height blur factor was assigned 1 if present or 0 if not. Source position was identified based on the center of indicated width, and assigned a value based on how far the position deviated from center (0) on a scale from 90° left (-10) to 90° right (10), with each slice of the circle equaling 1 measure (.5 in between value). A single factor ANOVA was performed on each dimension for two microphone techniques and for all sources combined, showing significant differences for all dimensions except Depth. However, a single factor ANOVA showed significance between the depth and height for all sources combined independent of recording technique used. The data set for source position was unique from the rest, as the source position values were unrestricted to the subjects. The outliers for this data set were identified if they exceeded three standard deviations. A two-factor ANOVA test was performed on this data set to see whether there is a statistical significance between recording techniques and source position specific sources. ANOVA was performed separately for Tune 1 and Tune 2 because the pieces feature slightly different instrumentation. Accordion, Voice and Clarinet were mostly in the center of the soundstage and did not vary positions between techniques ANOVA results for those instruments did not show any significant differences.

Results are presented in graphs showing mean (♦) mode (*) and standard deviation (whiskers). Mean values are useful to look at the big picture of responses however, due to the ordinal nature of the data, the mean values were not enough to describe the data accurately, so mode values were also used in order to show which responses appeared most often in the data set. The whiskers on the graphs indicate the Standard Deviation,

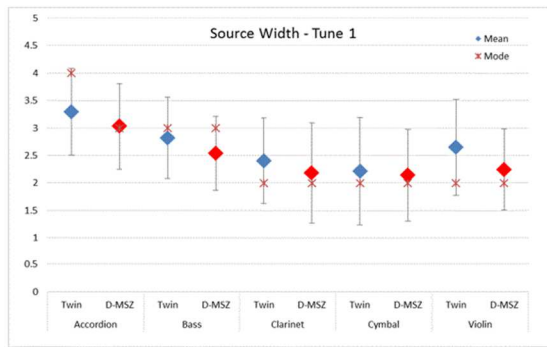


Figure 5

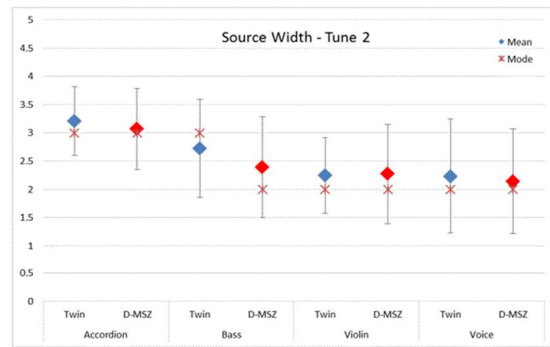


Figure 6

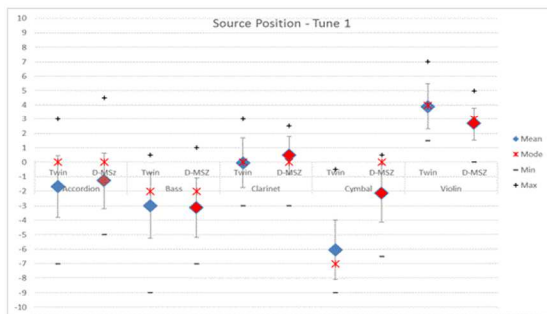


Figure 7

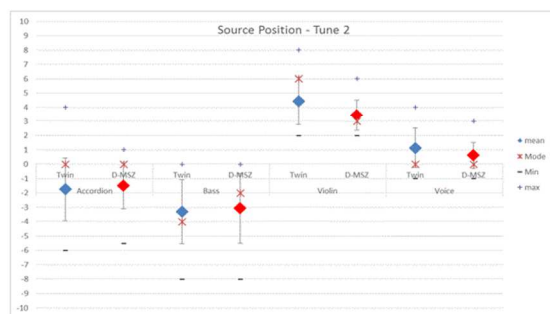


Figure 8

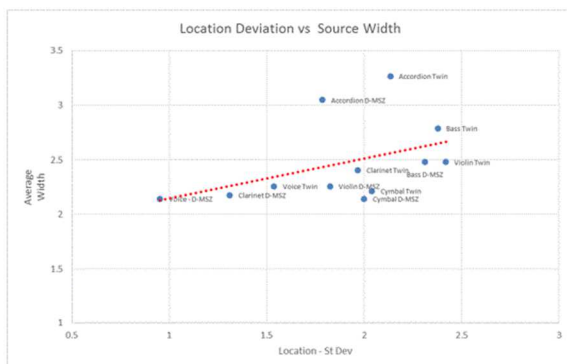


Figure 9

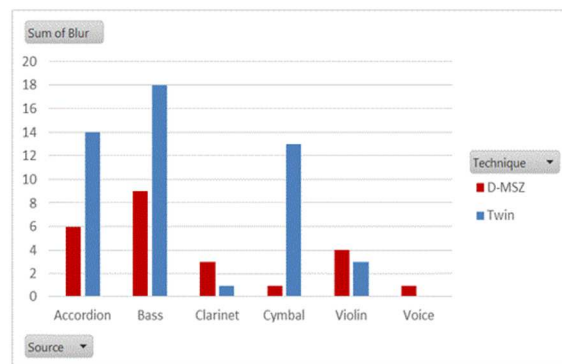


Figure 10

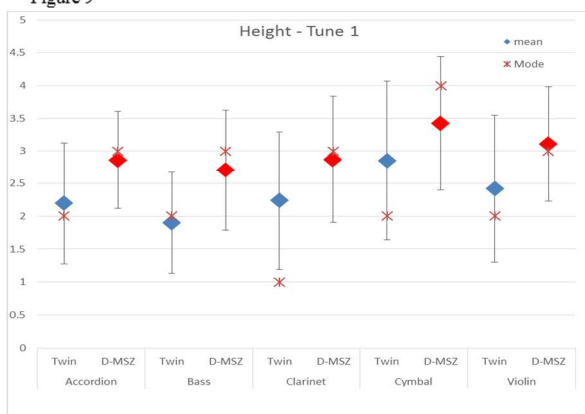


Figure 11

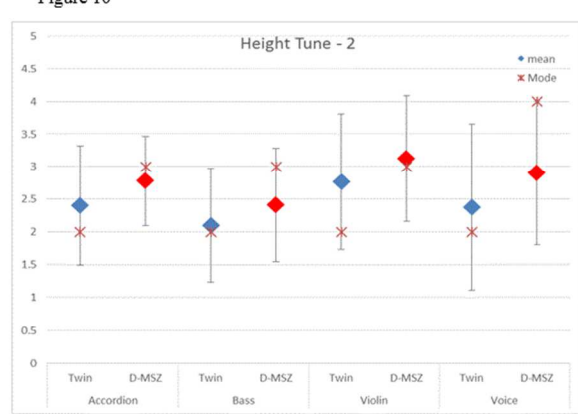


Figure 12

which translates into the deviation of responses from the mean. Larger whiskers equate to higher variances.

4. DISCUSSION

4.1. Width

The average widths for the Twin Square technique were slightly higher than for the Double-MSZ technique, more evident is that some sources have significantly higher width than others (Figures 5/6). The modes show that the subjects mostly identified Accordion and Bass as 3-Wide compared to 2-Narrow for the rest of the sources. Additionally, when looking at the two tunes separately, one can see that the average width for Bass is slightly smaller for both techniques in Tune 2, and the mode for the Double-MSZ technique falls to 2-Narrow. For both recording techniques, Accordion exhibited higher widths, with most subjects identifying the source as 'Very Wide'. Comparing Bass' average width, and the mode between the two Tunes, Tune 2 has slightly smaller width, and for Double-MSZ technique for second Tune most responses describe Bass as 'Narrow'

4.2. Localization

Individual sources were spaced further apart for the spaced 'Twin' technique. Based on variance and Max and Min values it appears that some sources are more easily localized than others (Figures 7/8). For example, when looking at Accordion values for Twin technique although Mean and Mode are not far off, the responses deviated quite significantly, with most answers falling between -4 and 0 which is approximately a 45° angle. Similarly when looking at the Violin for same tune and technique, most responses fall between (4-6) which translates to about 20° angle. Cymbal for Tune 1 is placed around 45° left in the Twins technique, while Double-MSZ technique places it near the center. The graph on Figure 9 maps each source and recording method width, with their location standard deviation values, and shows that there was correlation between consensus on source locations and smaller apparent source widths, which supports previous research that states that low widths are associated with better localization [8].

4.3. Height & Depth

The graphs for source height (Figures 11/12) show that average height for the Double-MSZ technique was higher than Twin for all sources and both tunes. The

modes also support that observation, showing that most subjects rated Twin technique slightly above lower speaker level while the angle for the Double-MSZ technique mostly fell to level 3 and 4 which is approximately 17° higher. Although the sources heights for the Twin technique were lower, the graph shows higher response variances for that technique for most of the sources.

The sums of Blur values (Figure 10) show that Blur was significantly higher for the Square technique, with Accordion, Bass and Cymbal reporting the highest values. The Blur was reported much less for Double-MSZ technique and was mostly equally distributed with Bass and Accordion reporting most Blur values. It must be noted that cumulatively for both techniques the blur was only reported 73 times for all subjects and all clips, with most subjects choosing to leave the box blank.

The results for Depth were the most inconclusive. The sources did not show any correlation between depth and position, nor between depth and microphone techniques. While individual measures for Depth did not provide useful results, when looking at both techniques combined and all sources summed together, there's a correlation, as depth increases, azimuth angle increases as well.

4.4. Perceived Room Size

Perceived room size was higher for the Twin technique. Most subjects identified the perceived room size for the Twin technique as '3-Bar/Club' while the Double - MSZ technique mostly elicited 'This Room' responses. The graph on Figure 16 shows how the room sized varied as the test progressed. There are noticeable sharp drops and peaks in three areas. The first one occurs when two back-to-back Twin clips are followed by a Double-MSZ clip, a sharp peak occurs when Twin is played in between two Double-MSZ clips, and the last peak occurs when a Twin technique is followed by two back-to-back Z clips. The variance shows the answers varied widely across all clips, and there's higher variance for Double-MSZ technique. One can see that the subjects heard significant differences between the clips, and rated the Twin technique as having larger perceived environment, though for five consecutive clips which contained both tunes, most subjects

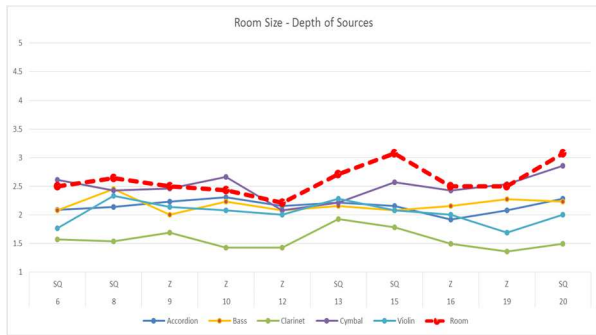


Figure 13



Figure 14

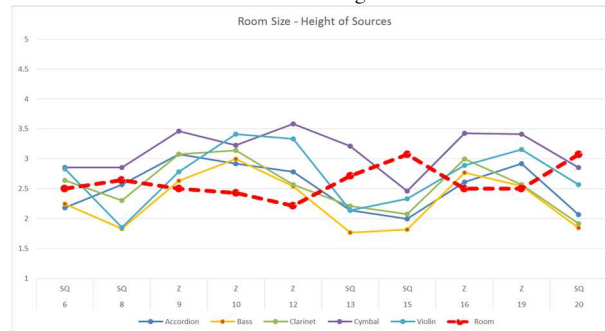


Figure 15

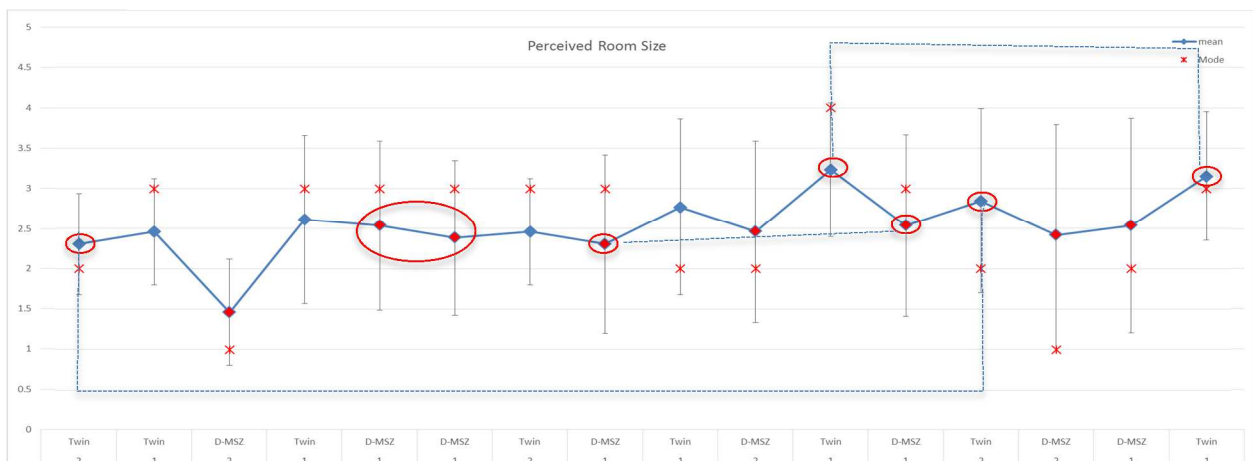


Figure 16

identified both techniques as 3 – Bar/Club. It is also interesting to note that although the Double-MSZ technique showed higher overall elevation of sources, these results did not translate into larger room size. The circled pairs of clips identify the duplicates in the clip set, and show that subjects answered consistently for identical piece of audio played consecutively and over o. Lastly, graphs on Figures 13/14/15 show how perceived room size changes compared to single dimensions, one can see that source width and depth slightly correlate with perceived room size.

5. CONCLUSION

The present study investigated two different 3D recording techniques, in a small studio environment, for Auro-3D speaker configuration. Using a perception test, the study evaluated the recordings on three-dimensional qualities, and whether the recording method influenced the end result. The study found that there were significant differences between techniques as

well as sources themselves in terms of localization in the horizontal as well as vertical planes, and perceived room size.

The Twins Square technique showed a wider soundstage and wider overall ensemble width, which corresponds with previous research suggesting that spaced techniques exhibit greater spatial impression due to larger microphone spacing which result in lower degree of correlation between signals [4]. Both techniques placed all sources correctly to where the players were positioned during the recording except, Double-MSZ technique incorrectly placed the Cymbal closer to the center. The Bass and Accordion were rated as being the widest sources. The higher Bass width may be due to the fact that lower frequencies are less directional. The width of Bass was slightly narrower for Tune 2 for both techniques, the reason for this may be that the bass in Tune 1 was bowed, while it was plucked in Tune 2, which may give it a slightly percussive sound and perhaps make it more easily locatable. Accordion may have been picked as the widest simply due to its size, and a similarly large frequency range which in the case of both tunes provided lower and mid-register harmony. Also, both the Bass and Accordion did not have an easily distinguishable melody, which could catch the listener's ear, while more melodic sources were localized easier and were rated narrower.

Double-MSZ showed higher elevation for all sources. The explanation for this may be the position of the array. Since it was positioned about 3.5 feet off the floor, at about musicians' waist level, it is possible that the resulting angle elevated sources, due to direct sound entering the Z channel. During post-test interviews several subjects revealed that they thought they heard different perspectives, one of sitting in the front row with musicians on stage, and another being on stage with musicians. Another conclusion that could be drawn is that Double-MSZ showed stronger phantom imagery in the vertical plane as the technique is able to maintain a stable elevation angle. This is also supported by the lower Blur values for Double-MSZ. It could be concluded that Double-MSZ is much more sensitive to the position than a spaced array but has more stable imagery. Bass and Accordion showed highest 'Blur' values, which may indicate that these sources may also be wide in the vertical plane. Height graphs also revealed that clarinet values for the Twin array were evenly dispersed between first three levels and may be caused by high directionality of the instrument, since many clarinet players move their instrument up and down as they play.

The study did not find any strong links between depth and microphone arrays. One possibility for this may be simply that the musicians were all positioned relatively on the same plane with only slight deviations in depth, which may not have been picked up by the arrays or the listener. Correlation between height and depth is definitely something that should be looked into further as there were no immediate reasons for this phenomenon. Many subjects expressed difficulties gauging depth in the post-test interviews.

The perceived room size values show that there were noticeable differences between techniques, with Twin technique showing much higher values. These results don't necessarily show that one technique is better than another at reproducing the room size or that one technique has better spaciousness than the other, however, it clearly shows that there's a difference in the way the arrays capture the ambience of the room. Since Dolan Studio is bigger than the research lab where the test took place and a bit smaller than a club it would be a 2 on the given scale. Therefore Twin Squares produced a slightly exaggerated room size, while Double-MSZ a smaller, more intimate room size. The high standard deviation shows that subjects did not have a clear consensus on the perceived room size for either technique. Lastly the results show that elevated sources for Double-MSZ did not have an impact on the room size. However, widths and depths of sources correlated slightly with room size for both techniques, possibly showing that subjects interpreted room size in terms of source or ensemble envelopment, and depth.

In summary, The Twins Square technique demonstrated a more enveloping soundstage, wider ensemble width, and a more spacious environment compared to the Double-MSZ technique, which showed better localization in vertical and horizontal planes, stable imagery, and a more intimate perspective. The test did not find any strong linking between microphone technique and perception of depth; however, both techniques showed a correlation between depth and height. Additionally, the test showed that the source type has a significant influence on its localization, where sources with lower frequency range demonstrated poorer localization than those in the higher frequency spectrum; however, sources that exhibit stronger directionality are more sensitive to source movement. While these findings solely describe this specific recording situation, I believe they could possibly be observed with other material and different recording environments, which is certainly something to explore in the future.

6. ACKNOWLEDGEMENTS

The author would like to express his sincere thanks to his advisor, Prof. Paul Geluso, for providing instrumental guidance, advice and support throughout this endeavor, and during my time at NYU and would also like to thank Prof. Agnieszka Roginska for her assistance and feedback. The author would also like to thank everyone who participated in the listening test; this project would not have been possible without your help and feedback.

7. REFERENCES

- [1] G. Theile and H. Wittek, "Principles in Surround Recordings with Height," *Audio Engineering Society*, vol. Convention Paper 8403, 2011.
- [2] H. Wittek, "Systems and Techniques for 3D Recording," Schoeps Mikrofone.
- [3] J. L. Barbour, "Elevation Perception: Phantom Images In The Vertical Hemi-Sphere," *Audio Engineering Society*, Vols. 24th International Conference: Multichannel Audio, The New Reality, 2003.
- [4] H. Lee and C. Gribben, "Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array," *Audio Engineering Society*, vol. 62, no. 12, 2014.
- [5] P. Geluso, "Capturing Height: The Addition of Z Microphones to Stereo and Surround Microphone Arrays," *Audio Engineering Society*, vol. Convention Paper 8595, 2012.
- [6] H. Lee and C. Gribben, "Audio Engineering Society," *On the optimum microphone array configuration for height channels*, Vols. Convention e-Brief 93, 2013.
- [7] F. Rumsey, "Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm," *Audio Engineering Society*, vol. 50, no. 9, 2002.
- [8] J. M. Victoria and T. Gorne, "Apparent Source Width in ITU Surround," *Audio Engineering Society*, vol. Convention Paper 7809, 2009.