Audio Engineering Society

Convention Paper 10517

Presented at the 151st Convention
2021 October, Online

Comparison of different techniques for recording and postproduction using main-microphone arrays for binaural reproduction.

Josua Dillier[1] and Hanna Järveläinen[2]

[1] *Zurich University of the Arts, Pfingstweidstrasse 96, 8031 Zürich, Switzerland*

[2] *Institute for Computer Music and Sound Technology, Zurich University of the Arts, Zürich, Switzerland*

Correspondence should be addressed to josua@josuadillier.com

ABSTRACT

We present a subjective evaluation of six 3D main-microphone techniques for three-dimensional binaural music production. Forty-seven subjects participated in the survey, listening on headphones. Results show a subjective preference for ESMA-3D, followed by Decca tree with height, of the included 3D arrays. However, the dummy head and a stereo AB microphone performed as well than any of the arrays for the general preference, timbre and envelopment. Though not implemented for this study, our workflow allows the possibility to include individualized HRTF's and head-tracking; their impact will be considered in a future study.

1 Introduction

With a major part of music today listened to on headphones, classical music audiences have become interested in binaural content [1]. Binaural reproduction may play an increasingly important role in the near future, presenting engineers the challenge of recording in a three-dimensional way. For example, Apple Music and Tidal are already incorporating 3D audio.

A dummy head microphone is a simple and effective method to produce binaural recordings for headphones. Its flexibility is very limited due to the fixed microphone type and placement. The only variable is where the dummy head is placed in the recording space. Externalisation can be improved using head-tracking [2,3]. Individual head-related transfer functions (HRTFs) can further improve the listening experience [3]. If these are to be used, a dummy head is no more an option because of the already acoustically processed HRTF of the dummy head itself.

Other options include recording with spot microphones only and adding artificial reverb, or using a 3D main-microphone array [4], or a combination of both. For reproduction of room characteristics, a main- or room-microphone is required. By simple binauralisation of a 2-channel main-microphone, externalisation may be achieved but envelopment might be somewhat lacking due to the low spatial resolution of the two point sources. This may be improved through creating multiple sources by 3D main-microphone arrays.

We investigated, whether equal or better envelopment, timbre and localization could be achieved with a microphone array in comparison to a dummy head. We made comparable recordings using different 3D main-microphone arrays, one 3D room array, a dummy head, and mono spot microphones. In addition, an AB stereo microphone was derived from the lower front microphones of the 2L-Cube inspired 3D main-microphone. The recordings were assessed in a survey. We discuss the results and possible contributing factors.

2 Recordings

“Humoreske” for woodwind-quintet by Alexander von Zemlinsky was recorded using six microphone arrays in a 600-seat, 508 m2 concert hall (RT60[250 Hz-2 kHz] = 2.1 s). The ensemble was arranged in a semi-circle 1m from the end of the stage with an approximate distance of 1.2 m between the players (from left to right): Flute, Clarinet, French Horn, Bassoon, Oboe. The used microphone arrays were:

- ORTF-3D [9]
- 2L-Cube inspired [11]
- Decca Tree with height [14]
- ESMA-3D [15]
- Spaced cardioid 3D-Room array
- Dummy head

2.1 Microphone placement

Side and top views and a concert hall view of the microphone arrays are presented in Figures 1-3. The microphone signals were positioned in 3D-space using Ambisonics, which is based on reproduction on the surface of a sphere [5].
For accurate representation, the microphones of the array must, due to precedence-effect [6], be on the surface of a sphere as well. For maximal decorrelation [7], an even distribution of the microphones on the sphere is required and can be achieved by using the corners of convex regular polyhedra.
For the 2L-Cube and the Decca tree, the shape of a cube was chosen. For the Decca tree setup, a center microphone was added in order not to lose focus on the main sound source with wider spacing [8]. For the 2L-Cube, an optional center microphone was added for observation of this focusing effect.
The ORTF-3D [9] can be arranged vertically coincident [10]: the cube is squashed to a horizontal square. The microphones are still pointing away from the center of where the cube was.

ESMA is primarily a two-dimensional technique with all microphones on the equator of the sphere. Height is later added with mid/side (MS) decoding. Section 2.2 gives specific details on each setup.

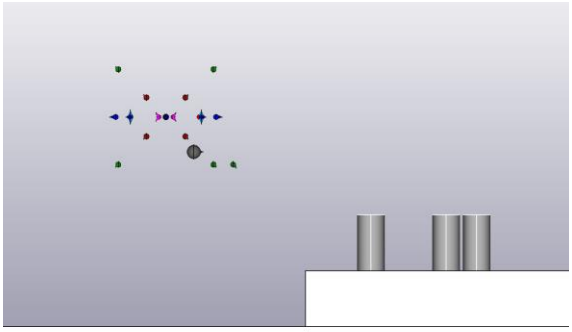


Figure 1. Side view of main microphone arrays and ensemble (right, grey) on stage (white). ORTF-3D (pink), 2L-Cube (red), dummy head (left, grey), ESMA-3D (blue), Decca tree with height (green).

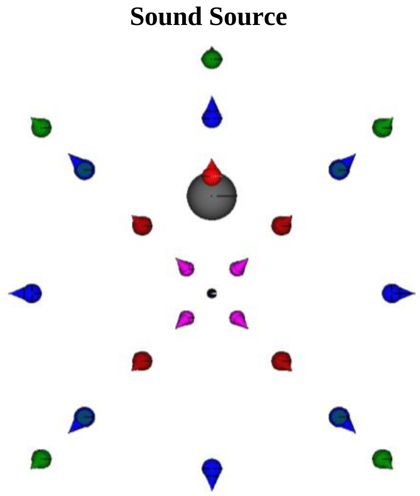


Figure 2. Top view of main microphone arrays; (from center) ORTF-3D (pink), 2L-Cube (red), dummy head (grey), ESMA-3D (blue), Decca tree with height (green).

2.1.1 *Positioning of the arrays*

A common center point for all arrays was defined, located three meters from the ground and three meters from the closest musicians. The directional microphones facing the sound source would record more direct sound compared to the omni microphones; to compensate this and to increase signal separation, the rear directional microphones were faced away from the sound source. All microphones of the main arrays and room array (except figure-of-eights in ESMA-3D) pointed away from the center point to obtain maximal decorrelation, considering that omnidirectional microphones may be somewhat directional at high frequencies.

2.2 Microphone Array setups

Microphone types and spacings are presented in Table 1 and detailed below. The microphone signals were amplified and converted to digital using Lawo Dallis 941/53 cards. The main-arrays and spot microphones were amplified with a gain of 30dB, the room microphone-array with 40dB. The individual microphone signals were recorded using Pyramix digital audio workstation (96kHz, 24bit).

2.2.1 ORTF-3D

The ORTF-3D was configured as described by the developers [9]. Due to the lack of enough available supercardioid microphones, cardioids were used for the upper layer.

2.2.2 2L-Cube inspired

The 2L-Cube [11] inspired setup aims at accurate reproduction of all room reflections through inter channel time delay (ICTD) cues. If the sources are not on the same plane as the microphones, the ICTD angles will be diminished. For calculation purposes, four virtual microphones were placed on the circumference of the sphere. They can be seen as four overlapping AB microphone arrays. Each array has to reproduce an angle of 90°. The correct distance between this configuration according to MARRS [12,13] is 68 cm (sources and microphones on the same height, recording and reproduction angle both 90°).
From these virtual AB arrays, the radius of the sphere was calculated as 48.1 cm. The spacing of the adjacent microphones was 55.5 cm.
The optional center microphone (reference is unclear about this, but presumably it is meant to be on the lower layer) was lifted to the height of the center point and positioned at 0° on the surface of the sphere.

2.2.3 Decca tree with height

For the Decca tree with height [14], a different approach was chosen. Decca tree setups are usually wide-spaced omni microphones. This configuration produces an open, spacious sound with a solid center image, but is not suitable for accurate stereophonic directional imaging [8]. Thanks to a narrower spacing in this setup, correct directional imaging is maintained for direct sound even though it is ignored for reflections. The lower front microphones L-C-R are responsible for correct imaging of the sources. L-C and C-R both only reproduce 45°. The spacing for an AB microphone with a reproduction angle of 90° is therefore doubled to 136 for the width L-R, resulting in a sphere radius of 117.8 cm.

2.2.4 ESMA-3D

ESMA-3D was configured with 8 cardioid microphones in the horizontal plane, as proposed [15]. As recommended [15,16], the spacing between the adjacent microphones was 55 cm [12,13]. For every second cardioid, a vertically oriented figure-of-eight microphone was arranged coincidentally. Directional microphones pointing up and down can be derived with MS decoding [16].

2.2.5 Spaced Cardioid 3D-Room Array

An additional array was set up for capturing room reflections and reverberation. It had a different center point than the other arrays and was not intended to represent the sound sources on its own, but to be mixed on a low level to add spaciousness to the main arrays or spot microphones.
The center point was set in the diffuse field at reasonable distance to any boundary surfaces.
It was on the same horizontal axis with the source and main array and maintained symmetry to this axis.
The height of the array was set at three meters to avoid comb filtering, and the distance to the sound source was ca 12 m. A cube layout was chosen, with a side length of 200 cm and cardioids pointing away from the center point, offering optimal decorrelation

for high envelopment [7] while maintaining some relation of the signals for consistent reproduction.

2.2.6 Dummy Head and Spot Microphones

Due to physical limitations, the dummy head was placed 50 cm lower and 40 cm closer to the ensemble than the center point of the main-microphone arrays.

Array	**Polar pattern**	**Microphone**	**Radius/ side length**
ORTF-3D	Supercardioid, upper layer: cardioid	Schoeps MK 42, upper layer Schoeps MK 4	17.3cm (virtual sphere) / 20cm
2L-Cube	Omni	DPA 4006	48.1cm / 55.5cm
Decca tree with height	Omni	L-C-R: Neumann TLM50, LS-RS: Neumann KM131, height channels: Schoeps MK 2S	117.8cm / 136cm
ESMA-3D	Cardioid, height channels: figure-of-eight	Neumann KM 184, height channels: Schoeps CCM 8	71.9cm / 55cm (adjacent)
Spaced Cardioid 3D-Room Array	Cardioid	Neumann KM 184	173cm / 200cm
Dummy head		Neumann KU 100	
Spot microphones	Cardioid	Neumann KM 184	

Table 1. Microphone types, polar patterns and spacing of all setup microphones.

3 Postproduction Processing

The editing was done using Pyramix [17]. The recorded microphone signals were grouped and edited the same way. For monitoring, the integrated 3D Panner with a 2nd order Ambisonics bus and the IEM binaural decoder [18] were used. The 2L-Cube and spot microphones were listened to while editing. After editing, all channels corresponding to the microphone signals were exported to individual mono files without any gain changes.

Figure 3. Main-microphone arrays and spot microphones in the concert hall.

For convenience, the files were imported to Reaper digital audio workstation and individual tracks were created for each file.
The dummy head channels were hard panned and sent directly to the output.
A stereo version was made, copying the lower front microphones of the 2L-Cube and hard panning them as an AB setup. No binauralisation was applied.
Within the microphone arrays, the levels of the microphone signals were maintained the same for all channels to preserve the interaural level difference (ILD) cues. Some adjustments were made in the Decca tree setup, compensating the output levels of the different microphone types. Balance between the different spot microphones was also maintained.

3.1 Higher Order Ambisonics Encoding

To maintain the possibility of using headtracking and individualized HRTFs, encoding was made to higher order Ambisonics using the IEM MultiEncoder [18].

3.2 Positioning of Signals in 3D-Space

For every array, a separate 64-channel bus was created, to allow 7th order Ambisonics encoding. The mono signals from the arrays and spot microphones were sent to different input channels of the corresponding bus and thereby to the input channels of the IEM MultiEncoder on the bus.

3.2.1 Cubical arrays

The positions in the Ambisonics encoder should match the positions of the actual microphones [9]. However, after subjective comparisons, we placed the ORTF-3D, 2L-Cube and Decca tree at +-40° instead of +-45° elevation. The positionings are presented in Table 2.

Apart from the center microphone, the positioning was the same for ORTF-3D, 2L-Cube and Decca tree. The center microphone of 2L-Cube was placed at 0° azimuth, 0° elevation and that of Decca tree at 0° azimuth, -45° elevation. For the Spaced Cardioid 3D-Room Array, the positions of the azimuth were the same as in the other arrays. Elevation was different depending on the purpose of the array. When used with one of the main arrays, +-55° elevation was chosen for the sources to not be on the same spot as the main array.

A mix was also produced with only the room array and spot microphones. In this configuration the elevation was left at +-45.

Microphone Position	Azimuth	Elevation
L	45°	-45° (-40°)
R	-45°	-45° (-40°)
LS	135°	-45° (-40°)
RS	-135°	-45° (-40°)
Lh	45°	45° (40°)
Rh	-45°	45° (40°)
LSh	135°	45° (40°)
RSh	-135°	45° (40°)

Table 2. Theoretical (and actual) positioning of a cube layout in Ambisonics.

3.2.2 ESMA-3D

For the ESMA-3D, the positions of the cardioid microphones were all on 0° elevation and in 45° intervals on the azimuth, beginning at 0°. The height and low channels were generated with MS decoding of the figure-of-eight microphones with their coincident cardioid microphone. These four height and four low channels were placed at the azimuth of the coincident cardioids they were derived from. Elevation was chosen by comparison by the authors and set to +-45°.

3.2.3 Spot Microphones

As usually done in the main-microphone technique, spot microphones were added to all the arrays, except for the dummy head. For the stimuli they were only used in the mix of Room array and spot microphones but included for interested readers.

The positioning of the spot microphones was array-dependent. The positions are presented in Table 3. Azimuth was defined through listening to where the different instruments were localized on the main microphone. Elevation, more difficult to localize [19], was set by estimating the actual angle in the recording setup and refined by comparisons of minor position changes by the authors.

Array	Azimuth	Elevation
ORTF-3D	18°, 9°, 0°, -9°, -18°	-50°, -40°, -30°, -40°, -50°
2L-Cube	22°, 11°, 0°, -11°, -22°	-50°, -40°, -30°, -40°, -50°
Decca tree with height	24°, 12°, 0°, -12°, -24°	-50°, -40°, -30°, -40°, -50°
ESMA-3D	26°, 13°, 0°, -13°, -26°	-40°, -30°, -20°, -30°, -40°
Spaced cardioid 3D-Room array	20°, 10°, 0°, -10°, -20°	-40°, -30°, -20°, -30°, -40°

Table 3. Final positions of the spot microphones from left to right (flute to oboe).

3.3 Binaural Decoding

IEM BinauralDecoder [18] was used in the mixing process. The implemented HRTF filters were measured on a KU 100, for comparison with the recorded dummy head of the same type.

4 Survey

4.1 Stimuli

Measures 1-71 (duration 1’43’’) of Zemlinsky’s “Humoreske” were used as stimuli, mixed as follows and normalized to the same loudness (-23 LUFS integrated):

- ORTF-3D
- 2L-Cube without center microphone
- Decca tree with height
- ESMA-3D
- mix of Room array and spot microphones
- Dummy head
- AB main microphone (lower front microphones from 2L-Cube)

For better comparison and for no apparent audible benefit, spot- or room-microphones or artificial reverberation were not used, except for the mix of spot microphones and Room array. The balance between Room array and spot microphones was adjusted by the authors to produce a comparable mix with the main microphone arrays. The center microphone of the 2L-Cube was muted for a more symmetrical sound field and better localization in the authors’ perspective. Ambisonics encoding and binaural decoding was applied as described in section 3.

4.2 Design and participants

The survey was carried out online using PsyToolkit [20,21]. Participants downloaded the stimuli and listened on headphones. Data was first collected for the two headphone predictors HF1 (In Ear / On Ear / Over Ear / not specified) and HF2 (closed-back / semi-open / open / n.s.). The seven audio samples were then rated in random order. Participants could listen to the sample as many times as they wished and then give their ratings on the four attributes, described as follows (translated from German):
Localization = How clearly can you localize in which direction the different instruments play?
Envelopment = How much do you feel surrounded by the sound?
Timbre = How do you perceive the timbre of the instruments?
Overall = How much do you like it in general?

Participants gave responses by setting a continuous slider for each attribute. The values were mapped to the range [0,1]. The end points were described as follows (translated from German):
Localization: very diffuse - very clear
Envelopment: very weak - like in a concert hall
Timbre: very coloured, - very natural
Overall: don‘t like at all - like very much

The participants were sound engineering and music students, musicians, and consumers, and each person could participate once. N=47 participants completed the survey and were considered in the analysis.

5 Results and Discussion

Statistical analyses were made using Bayesian inference [24], a statistical method where the probability of a hypothesis is updated as more information becomes available. The computations were carried out in R[22] using the brms package [23]. We fit a multivariate regression model, predicting the four response variables from microphone setup and the headphone predictors HF1 and HF2. Since all participants rated all seven microphone arrays, intercept was allowed to vary over participant. The additive model was specified as follows (using the syntax of brms):

```
mvbind(Overall, Localization, Timbre, Envelopment) ~
setup + HF1 + HF2 + (1|p|subject),
```

where the function mvbind() defines separate response variables and (1|p|subject) sets the individual intercepts modelled as correlated. The model parameters were estimated by Bayesian inference [24].
Considering the limited response range [0,1], we fit two models: one assuming normal response distributions and another assuming beta-distributions. In the latter case, values were mapped into the open interval (0,1)[25]. According to a leave-one-out comparison, the Gaussian family model fit the data better; therefore, we report estimates based thereon.

5.1 Results

The estimated effects are seen in Figures 4-7. Notably, AB main microphone and dummy head were rated the highest for all four attributes. In

Overall ratings as well as for Timbre and Envelopment, AB main microphone, dummy head, ESMA-3D and Decca tree were rated credibly higher than 2L-Cube, ORTF-3D and Room array. For Localization, dummy head was rated credibly higher than any of the other setups, followed by ESMA-3D and AB main microphone. The rest were rated credibly lower than the first three. In Figures 4-7, these groupings are marked by dashed shapes.
Neither of the headphone predictors had a significant impact on the ratings, with the exception that Localization was rated credibly lower with On Ear headphones. Residual variances were positively correlated between all response variables, ranging from 0.44 to 0.67.
Interval estimates from posterior Markov chain Monte Carlo (MCMC) draws are presented in Figure 8. All effects are relative to the stereo AB microphone setup. Parameters whose uncertainty interval does not contain zero are considered having a credible effect; a negative estimate signifies a negative effect on ratings compared to the stereo AB setup, and vice versa. The stimuli, dataset, and R code are available in our repository[1].

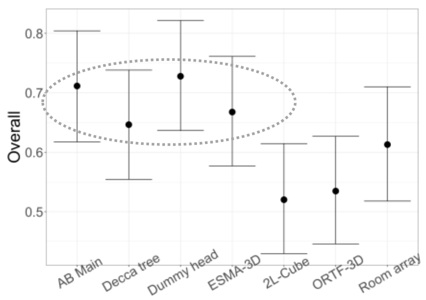
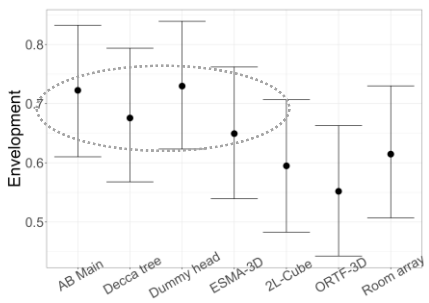


Figure 4. Conditional effects of setup on "Overall".



[1]http://doi.org/10.5281/zenodo.5005884

Figure 5. Effects of setup on "Envelopment".

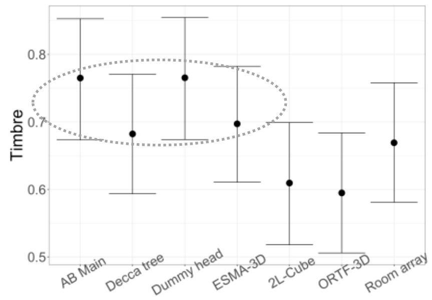


Figure 6. Effects of setup on "Timbre".

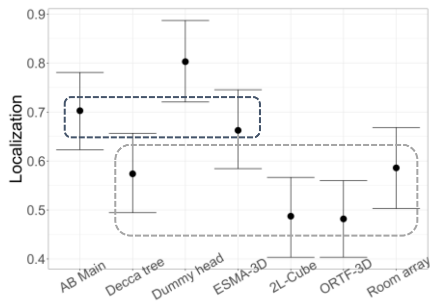


Figure 7. Effects of setup on "Localization".

5.2 Discussion

In all four attributes, the results show a small but credible difference between the four more preferred setups (AB main microphone, dummy head, ESMA-3D and Decca tree with height) and the less preferred setups (2L-Cube, ORTF-3D, and Room array).
This outcome somewhat surprised authors, who preferred the 2L-Cube inspired and the Decca tree with height. The results suggest, however, that a stereo AB microphone may be preferred over a number of 3D microphone array techniques.
Inter-rater differences were large, as is typical for subjective rating tasks. This may be partly due to the different amounts of correlation between the listeners' individual HRTFs and the used HRTF.

It remains unknown how much familiarity may have contributed to the result. If neither head-tracking nor individualized HRTFs are used, dummy head recording still seems to be a better choice than 3D

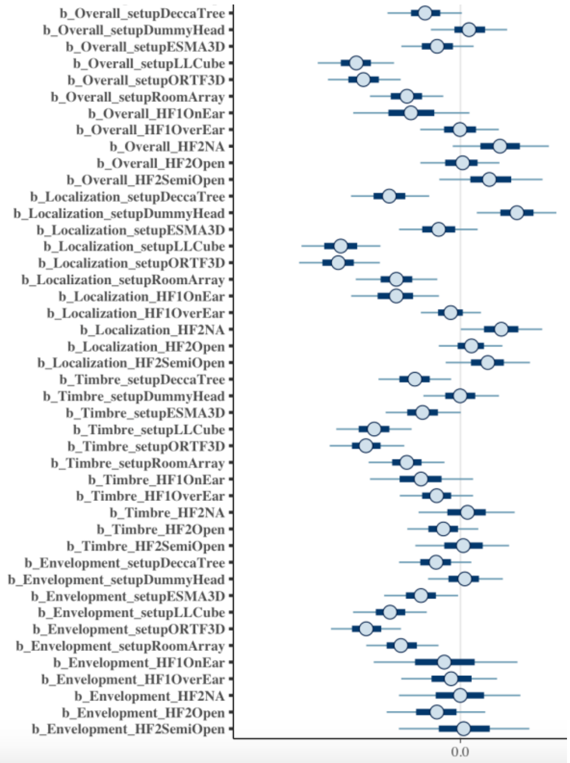


Figure 8. Posterior interval estimates. Positive of negative effects are relative to the stereo AB main microphone setup.

microphone arrays. The slightly closer positioning to the sound source of the dummy head may have impaired the comparison.

It also remains unclear whether the low performance of ORTF-3D was influenced by the usage of cardioid microphones for the upper layer. Furthermore, the different microphone types in Decca tree with height may have affected its performance.

Localization could be improved by adding spot microphones to the arrays and for 2L-Cube by increasing directionality at high frequencies using a large diaphragm capsule or spherical attachment[26].

Different encoding and decoding with individualized HRTFs and head-tracking may further improve the performance of the arrays.

Conclusions

A subjective evaluation of recordings and mixing approaches, made using the main-microphone technique, was carried out. Listening on headphones, participants generally preferred the AB main microphone and dummy head recordings, along with the 3D microphone techniques ESMA-3D and Decca tree with height. The described technical setup includes the possibility of using individualized HRTFs and head-tracking. Their impact will be investigated in future studies.

References

[1] T. Parnell, and C. Pike, "An Efficient Method for Producing Binaural Mixes of Classical Music from a Primary Stereo Mix," Audio Engineering Society, Engineering Brief 430, 2018.

[2] E. Hendrickx, P. Stitt, J. Messonnier, J. Lyzwa, BR. F.. Katz, and C. de Boishéraud, "Improvement of Externalization by Listener and Source Movement Using a “Binauralized” Microphone Array," J. Audio Eng. Soc., vol. 65, no. 7/8, pp. 589-599, 2017.

[3] Dickreiter, M., Dittel, V., Hoeg, W., and Wöhr, M.: “Handbuch der Tonstudiotechnik”, vol. 1, chapter 5.6, W. de Gruyter GmbH Berlin/Boston, 2014.

[4] H. Lee, "Multichannel 3D Microphone Arrays: A Review," J. Audio Eng. Soc., vol. 69, no. 1/2, pp. 5-26, 2021.

[5] M. Neukom, "Ambisonic Panning," Audio Engineering Society, Paper 7297, 2007.

[6] Haas, H. (1949). “The influence of a single echo on the audibility of speech,” J. Audio Eng. Soc. 20, 145-159, English translation, 1972.

[7] G. Theile, and H. Wittek, "Principles in Surround Recordings with Height," Audio Engineering Society, Paper 8403, 2011.

[8] G. Theile, "Natural 5.1 Music Recording Based on Psychoacoustic Principals," Audio Engineering Society, Paper 1904, 2001.

[9] H. Wittek, and G. Theile, "Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR," Audio Engineering Society, Paper 9869, 2017.

[10] H. Lee, and C. Gribben, "Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array," J. Audio Eng. Soc., vol. 62, no. 12, pp. 870-884, 2014.

[11] M. Lindberg, "3D Recording with the '2L-Cube,'" http://www.2l.no/artikler/2L-VDT.pdf

[12] H. Lee, D. Johnson, and M. Mironovs, "An Interactive and Intelligent Tool for Microphone Array Design," Audio Engineering Society, Engineering Brief 390, 2017.

[13] Applied Psychoacoustics Lab (APL) University of Huddersfield, Microphone Array Recording and Reproduction Simulator: https://marrsweb.hud.ac.uk/

[14] H. Lee, and D. Johnson, "An Open-Access Database of 3D Microphone Array Recordings," Audio Engineering Society, Engineering Brief 543, 2019.

[15] M. Williams, "Migration of 5.0 Multichannel Microphone Array Design to Higher Order MMAD (6.0, 7.0 & 8.0) with or Without the Inter-Format Compatibility Criteria.," Audio Engineering Society, Paper 7480, 2008.

[16] H. Lee, "Capturing 360° Audio Using an Equal Segment Microphone Array (ESMA)," J. Audio Eng. Soc., vol. 67, no. 1/2, pp. 13-26, 2019.

[17] Merging Technologies, Pyramix digital audio workstation: https://www.merging.com/products/pyramix

[18] Institute of Electronic Music and Acoustics, IEM Plug-in Suite: https://plugins.iem.at/

[19] J. Blauert, Spatial Hearing: The Psychoacoustics of Human Sound Localization, The MIT Press, 1997.

[20] Stoet, G. PsyToolkit - A software package for programming psychological experiments using Linux. Behavior Research Methods, 42(4), 1096-1104, 2010.

[21] Stoet, G. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Teaching of Psychology, 44(1),24-31, 2017.

[22] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[23] Bürkner P. "brms: An R Package for Bayesian Multilevel Models Using Stan." Journal of Statistical Software, 80(1), 1–28, 2017.

[24] Kruschke, J. Doing Bayesian Data Analysis (2nd Ed.), Boston: Academic Press 2015.

[25] Smithson, M. & Verkuilen, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol. Methods 11, 54–71 (2006).

[26] M. Schneider, "Omnis and Spheres - Revisited," Audio Engineering Society, Paper 5338, 2001.