McKenzie, Thomas; Schlecht, Sebastian J.

Source position interpolation of spatial room impulse responses

Published in:
AES Europe 2023

Published: 01/01/2023

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
McKenzie, T., & Schlecht, S. J. (2023). Source position interpolation of spatial room impulse responses. In *AES Europe 2023: 154th Audio Engineering Society Convention* (AES Europe 2023: 154th Audio Engineering Society Convention). Curran Associates Inc.. http://www.aes.org/e-lib/browse.cfm?elib=22044

Source position interpolation of spatial room impulse responses

Thomas McKenzie, Sebastian J. Schlecht

March 6, 2024

Measured spatial room impulse responses (SRIRs) are often used for realistic six degrees-of-freedom (6DoF) virtual reality applications, as they allow for the high quality capture and reproduction of a room's acoustics. Dense sets of SRIR measurements are time consuming to acquire, especially for multiple source and receiver combinations, and so interpolation of sparse measurement sets is required. This paper presents a method for interpolating between higher-order Ambisonic SRIRs with a fixed receiver position but different sound source positions, using a previous methodology. The method is based on linear interpolation with spectral equalisation and RMS compensation, though direct sound, early reflections and late reverberation are processed individually. In a numerical comparison to a basic linear interpolation, the proposed method is shown to more smoothly fade between source positions in root-mean-square amplitude and direction-of-arrival metrics.

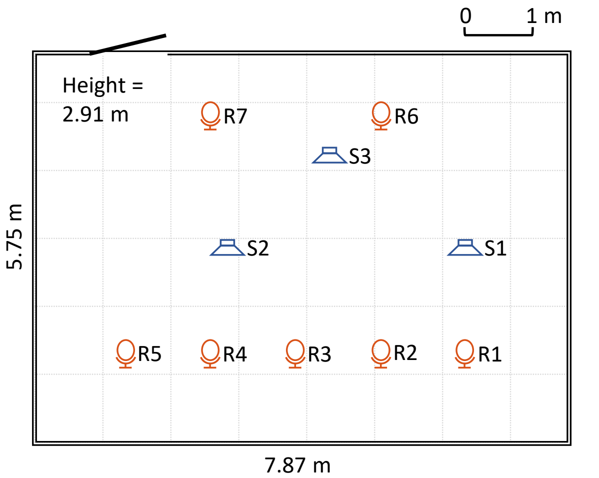


Figure 1: Illustration of the source (S) and receiver (R) positions used in the evaluation. Microphones and loudspeakers are oriented facing north and south, respectively, according to the illustration orientation.

1 Introduction

For virtual and augmented reality experiences with six degrees-of-freedom (6DoF), audio is of paramount importance in producing and maintaining immersion [1], and measurements of room acoustics are still largely considered to give more authentic results than simulations [2]. The room impulse response (RIR) captures the reverberation of a space at a single source and receiver position, and by convolving a dry input signal with a measured RIR, it is possible to auralise the signal to appear to originate inside the measured space.

RIRs measured with spherical microphone arrays, which typically encode the microphone capsule signals into spherical harmonics (SH) [3], are herein referred to as spatial room impulse responses (SRIRs). These can capture the directional elements of a room's reverberation, therefore allowing for a further increase in the accuracy of auralisations. Recent literature has shown how RIRs measured at different receiver positions vary inside a single room. Typically, the direct sound carries the highest perceptual importance and positional changes in the degree of centimetres are perceivable [4]. It therefore requires the most care in reproduction, including reproduction at very high orders of Ambisonics [5]. Next most important are early reflections, which give a sense of the size of the space [6]. Finally, the late reverberation is the least complex, and varies much less with position [4]. It should also be noted that these requirements vary with different stimuli though: sounds with limited frequency bandwidth can forgive larger distances between measurements [7].

Measuring SRIRs at a high spatial density, with multiple source and receiver positions, requires a specific and precise setup and can be highly time-consuming. Therefore, it is desirable to interpolate between sparse sets of measurements.

Previous methods of single channel RIR interpolation include dynamic time warping by stretching the time axes of RIRs until they align [8]; modal interpolation by solving the Helmholtz equation [9]; and using plane-wave decomposition and time-domain equivalent source methods [10]. Attempts at SRIR interpolation include first-order methods [11, 12] which separate specular and diffuse parts to be interpolated separately. Both these methods individually interpolate early reflections using direction-of-arrival (DoA) triangulation estimations.

A method of higher-order Ambisonic SRIR interpolation was previously published that is robust for interpolation of SRIRs measured in coupled rooms and in the transition between rooms [13, 14], which feature complex and perceivable acoustical features around the coupling aperture [15, 16]. The method was designed to interpolate between receiver positions with a fixed source.

This paper presents a different use-case for the previously published method, which here allows for interpolation between two source positions for a fixed receiver, such as between sources S1 and S2 in Fig. 1. The paper is laid out as follows: Section 2 presents the methodology for interpolating between source positions, describing the separate techniques employed for direct sound, early reflections and late reverberation. Section 3 then evaluates the interpolation technique, comparing the presented method to a basic linear fade method in both root-mean-square level and direction-of-arrival. The results are discussed in Section 4, along with limitations and proposed continued developments, and the paper is concluded in Section 5.

2 Method

This section briefly describes the SRIR interpolation method. The method is designed for 3D, 2D, or 1D (along a line) sets of SRIRs, to be therefore appropriate for 6DoF workflows. It is based on the SRIR interpolation algorithm detailed in previous research [13, 14], to which the reader is directed for a more in-depth description. The method was originally intended for interpolating between receiver positions with a fixed source position. The use-case presented in this paper is to interpolate between source positions for a fixed receiver position. The maximum SH order is denoted in this paper as N, with the order of an individual SH component denoted by n and the degree denoted by m. The Ambisonic Channel Numbering (ACN) and semi-normalised (SN3D) conventions are followed, and all audio is digitised at a sampling rate of 48 kHz.

The direct sound, early reflections and late reverberation are treated separately. Firstly, input SRIRs are time-aligned such that any delay from different distances between the source and receivers is removed. This delay can be reinserted post-interpolation if required. For a given interpolated source position, the Euclidean distance from the nearest two input source positions is calculated, which produces the contributing weight of each source position.

2.1 Direct Sound

The direct sound is taken as the first 200 samples of the input SRIRs (4.17 ms at a sampling rate of 48 kHz). The direction-of-arrival (DoA) of each SRIR's direct sound is first estimated using the pseudointensity vector from its 1st-order components. The result is normalised to produce a single DoA estimation for both input sources.

For each interpolated source position, the direct sounds of the two input SRIRs are steered to the relative new DoA, which is a weighted average of the two input source DoAs. The input SRIR direct sounds are then amplitude weighted based on the relative distance from the input source positions and the interpolated source position, before being summed.

2.2 Early Reflections

For each interpolated SRIR, the transition between early reflections and late reverberation, t_{EL}, is calculated as the time taken for the energy decay curve (EDC) to pass a set threshold energy value [17]. A variable transition time is used to account for different source-dependent reverberation characteristics, such as one source producing a higher reverberation time. The omnidirectional channel of each SRIR is first bandpass filtered at 1 kHz, then normalised to a maximum amplitude of 1. Schroeder integration then produces the EDC, and values of t_{EL} are calculated as the time when the EDC passes below a tenth of the maximum value, rounded to the nearest 1000 samples.

For each interpolated SRIR, the early reflections of the nearest input SRIRs are first windowed from 200 samples (the end of the direct sound) to the transition time t_{EL}. The early reflections are linearly interpolated and then equalised directionally, using spherical harmonic beamforming and reconstruction [18]. For this, the input SH domain responses are evaluated with a set of max-$\mathbf{r}_{\mathrm{E}}$ beams directed to a dense set of uniformly distributed directions on a t-design [19], whereby the t-design with the least number of points that fulfills $t \geq 2N + 1$ is selected. For the 4th-order responses used in this paper for example, the t-design has 48 points. The beamformer output signals for each input source are weighted with the distance weights and summed together.

Next, the summed signals are equalised to match the weighted sum of the input magnitude spectra in each direction. Equalisation is needed to rectify any comb-filtering artefacts that may arise from the summing of correlated signals; this comb-filtering is most apparent when the SRIRs to be interpolated are a greater distance apart. Each beamformed signal is equalised separately, such that colouration is removed in each direction. This is done in equivalent rectangular bandwidth (ERB) frequency bands. For each ERB band, the target RMS is a sum of the RMS of each amplitude-weighted nearest SRIR beam divided by the current RMS of the interpolated beam. An equalisation curve is then calculated by linear interpolation of each ERB band target RMS, between 20 Hz and 20 kHz. After the directionally equalised responses for every directional response are obtained, they are brought back in the SH domain.

2.3 Late Reverberation

The late reverberation is interpolated in much the same way as for early reflections in Section 2.2, though the directional equalisation is deemed unnecessary due to the higher spectral isotropy and general lower perceptual relevance here. Instead a single direction-independent equalisation filter is employed. Finally, the direct, early and late parts of the SRIRs are combined.

3 Evaluation

The source interpolation algorithm was evaluated numerically, using a publicly available[1] dataset of 4th-order Ambisonics SRIR measurements in a variable acoustics room, intended for 6DoF research [20]. The set of measurements used in this evaluation was the most reverberant *0percent_absorbers* measurements, taken using an MH Acoustics em32 Eigenmike spherical microphone array and encoded to Ambisonics using the *array2sh* Sparta VST plugin [21]. Fig. 1 presents the room geometry and the positions of the receivers and sources.

To assess the proposed source interpolation method, four SRIRs were generated from the measurements between sources S1 and S2 as labelled in Fig. 1. This was done for all 7 receiver positions. The interpolated source positions were equally spaced between the two measured source positions.

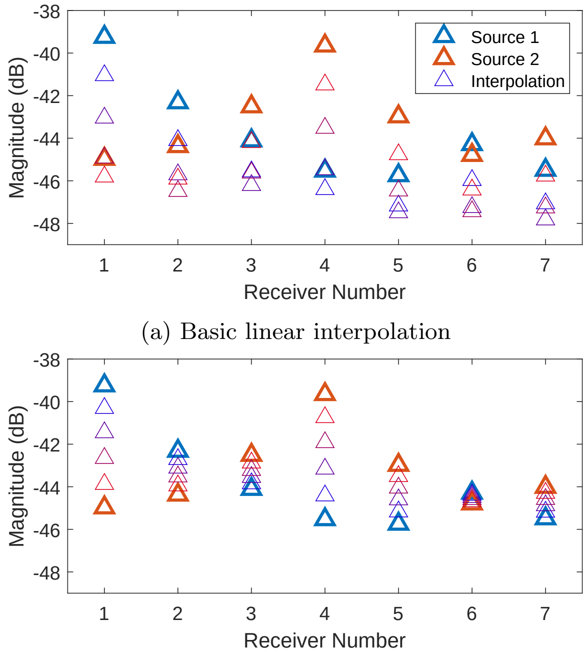
With a lack of ground truth measurements between the source positions available, the proposed source interpolation method was evaluated through comparison to a simple linear interpolation, made using weighted cross fading between measurements.

3.1 RMS Level

Fig. 2 presents the root-mean-square (RMS) level in dB of the interpolated SRIRs (the entire duration of the signal, omnidirectional SH channel), for all 7 receivers, for both the proposed and basic linear methods. For an accurate interpolation, one would expect even intervals in interpolated RMS values between those of the two input SRIRs.

The RMS levels of the SRIRs generated using the proposed method are shown to evenly fade between the two input SRIRs for each receiver position. However, this is not the case for the basic linear method, for which some interpolated positions even produce a lower RMS level than either input source positions.

[1] https://doi.org/10.5281/zenodo.5720723



(a) Basic linear interpolation

(b) Proposed interpolation

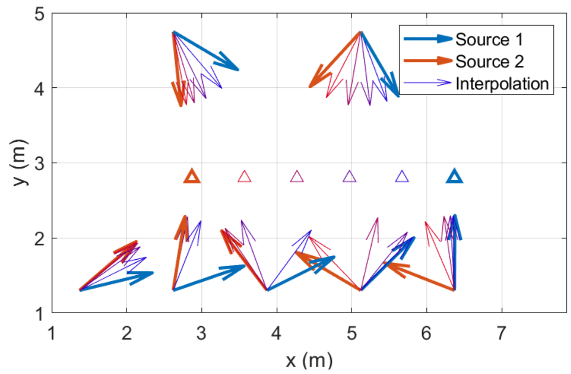
Figure 2: RMS amplitude of SRIRs (omnidirectional SH channel) interpolated between source positions S1 and S2, for all 7 receivers. Receiver numbering is as labelled in Fig. 1.

3.2 Direction of Arrival

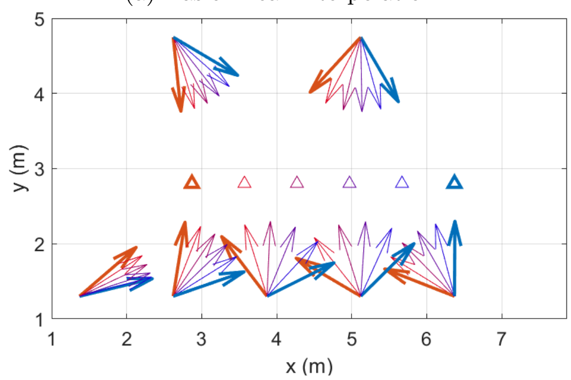
To evaluate the directional aspects of the source interpolation algorithm, Fig. 3 presents the estimated direction-of-arrival (DoA) of the direct sound (first 150 samples) of the interpolated SRIRs between the sources. DoA here was calculated, as in the method for direct sound, using the pseudointensity vector from the 1st-order components. For an accurate interpolation, one would expect even increments in direction between the two input SRIRs for each receiver position.

The basic linear method has a tendency to point closer towards the input source positions for receiver positions R1 to R5, whereas the proposed method produces a significantly more even fade. Interestingly for receiver position R6, both methods appear to produce somewhat similar results with the linear fade appearing to be relatively accurate.

Though Fig. 3 shows a clear directional improvement over the basic linear interpolation method for most tested receiver positions, this is visually somewhat limited in the plot due to only showing a single direction of the direct sound, which is calculated as the dominant direction of the pseudointensity vector. To look in greater detail, the DoA at receiver position R1 (bottom right on Fig. 1) was estimated above 3 kHz in one degree horizontal resolution for all interpolated SRIRs. This was done by steering a 4th-order hyper-cardioid beamformer (also referred to as normalised plane wave decomposition) in all

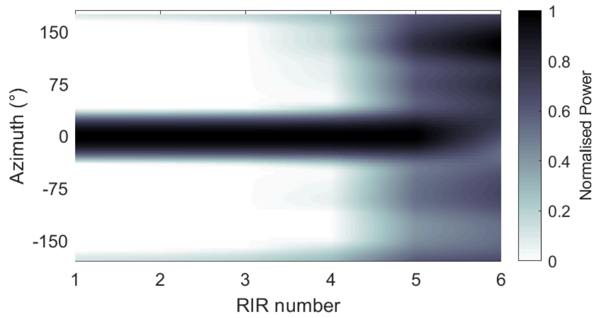


(a) Basic linear interpolation

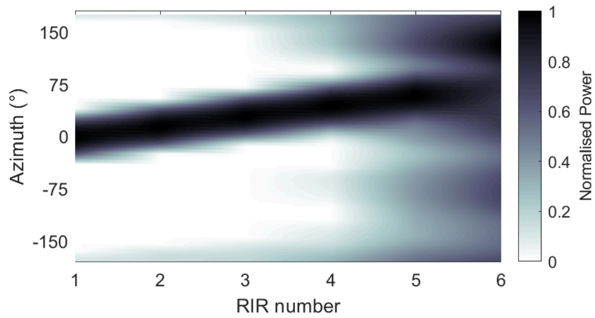


(b) Proposed interpolation

Figure 3: Direction-of-arrival of direct sounds, for SRIRs interpolated between source positions S1 and S2, as labelled in Fig. 1, for all 7 receiver positions. Target source positions denoted by triangles.



(a) Basic linear interpolation



(b) Proposed interpolation

Figure 4: Normalised power response along the horizontal axis of SRIRs interpolated between source positions S1 and S2, for receiver R1 (as labelled in Fig. 1).

horizontal directions [22] and calculating the power, which generally shows the power of the direct sound and loudest early reflections. Fig. 4 presents the normalised power response along the horizontal axis of the interpolated SRIRs for receiver R1.

These plots more clearly demonstrate the directional issues with the basic interpolation method, which simply fades out one SRIR and fades in another. Comparing this to the proposed method, the direct sound (assumed in the plot to be the direction with the highest power) is shown to smoothly move from the first location to the second, whilst retaining the linear fade of the early reflections and late reverberation (shown by the lower power).

4 Discussion

The proposed method for SRIR source interpolation has been evaluated through RMS level and DoA analysis. It shows clear improvements over a basic linear interpolation method, with an even fade between the RMS of the two input source positions. Compared to the RMS level when using basic linear interpolation, where some in-between SRIRs had a lower RMS than both input SRIRs, the proposed method is more appropriate and in line with expectations. This mirrors the results seen for the receiver interpolation study [13].

Similarly, the DoA analysis shows that the proposed method evenly rotates from one source position to the other in direct sound, whilst linearly fading the directional elements of the rest of the SRIRs. It would be possible to use a more advanced interpolation method for early reflections as in [10, 11, 12] in future developments. However, further perceptual evaluation would be needed to determine the perceptual benefits of more physically accurate early reflection interpolation, as the testing in [13] found that a linear interpolation of the early reflections and late reverberation, with amplitude normalisation and equalisation, is largely perceptually acceptable.

Fig. 3 illustrates the small inaccuracies in DoA estimation, as the arrows do not all point exactly to the source positions. This is due to the fact that the direction of the source is estimated based on the input signals, and not known by the algorithm. One reason for this is that, in the event of occlusion between the source and receiver, the current method will estimate the dominant direction of the sound source. This is in contrast to a method that renders towards the target source location, whereby occluding objects could then have no influence on the interpolated source to receiver path. However, some inaccuracies in DoA estimation may have been due to the use of the the pseudoinverse vector as the DoA estimation method. A higher-order DoA estimation method could be employed instead, such as EigenBeam Estimation of Signal Parameters via Rotational Invariance Techniques (EB-ESPRIT) [23], which has been implemented in a related study [24].

5 Summary

This paper has presented a new use-case for a previously published method of higher-order Ambisonic spatial room impulse response (SRIR) interpolation, whereby it is here used for interpolating between two source positions in SRIR measurements instead of between receiver positions. The method treats direct sound, early reflections and late reverberation separately, whereby the direct sound of the input signals is rotated prior to summing and equalisation, whereas the early reflections and late reverberation are not rotated.

The proposed method has been compared numerically to a basic linear interpolation method which simply applies a weighted gain to the input spatial room impulse responses (SRIRs), and the proposed method is shown to more smoothly fade from one source to the other.

The findings of this study suggest that it is possible to make fewer measurements and still obtain a set of SRIRs with many source positions.

Further evaluation using ground-truth measurements would be beneficial, as would perceptual testing. Extending the methodology to extrapolate other positions not directly between measurements is a logical next step.

A Matlab implementation of the source interpolation technique, `interpolate_SRIRs_source.m`, along with a demonstration script which reproduces the figures presented in this paper, `demo_interpolate_SRIRs_source.m`, is available to download at `https://github.com/thomas-mckenzie/srir_interpolation`.

References

[1] Potter, T., Cvetković, Z., and De Sena, E., "On the relative importance of visual and spatial audio rendering on VR immersion," *Frontiers in Signal Process.*, 2(September), pp. 1–10, 2022, doi:10.3389/frsip.2022.904866.

[2] Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., and Weinzierl, S., "A round robin on room acoustical simulation and auralization," *J. Acoust. Soc. Am.*, 145(4), pp. 2746–2760, 2019, doi:10.1121/1.5096178.

[3] Gerzon, M. A., "Periphony: with-height sound reproduction," *J. Audio Eng. Soc.*, 21(1), pp. 2–10, 1973.

[4] Neidhardt, A., Tommy, A. I., and Pereppadan, A. D., "Plausibility of an interactive approaching motion towards a virtual sound source based on simplified BRIR sets," in *AES 144th Conv.*, pp. 1–11, Milan, 2018.

[5] Ahrens, J. and Andersson, C., "Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre," *J. Acoust. Soc. Am.*, 145(4), pp. 2783–2794, 2019, doi:10.1121/1.5096164.

[6] Hameed, S., Pakarinen, J., Valde, K., and Pulkki, V., "Psychoacoustic cues in room size perception," in *AES 116th Conv.*, Berlin, 2004.

[7] Neidhardt, A. and Reif, B., "Minimum BRIR grid resolution for interactive position changes in dynamic binaural synthesis," in *AES 148th Conv.*, pp. 1–10, Online, 2020.

[8] Masterson, C., Kearney, G., and Boland, F., "Acoustic impulse response interpolation for multichannel systems using Dynamic Time Warping," in *AES 35th Int. Conf.*, pp. 1–10, London, 2009.

[9] Das, O., Calamia, P., and Gari, S. V. A., "Room impulse response interpolation from a sparse set of measurements using a modal architecture," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, pp. 960–964, Online, 2021, ISBN 9781728176055.

[10] Antonello, N., De Sena, E., Moonen, M., Naylor, P. A., and Van Waterschoot, T., "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Trans. on Audio, Speech and Lang. Process.*, 25(10), pp. 1929–1941, 2017, doi:10.1109/TASLP.2017.2730284.

[11] Müller, K. and Zotter, F., "Auralization based on multi-perspective Ambisonic room impulse responses," *Acta Acustica*, 6(25), pp. 1–18, 2020.

[12] Zhao, J., Zheng, X., Ritz, C., and Jang, D., "Interpolating the directional room impulse response for dynamic spatial audio reproduction," *Applied Sciences*, 12(4), 2022, doi:10.3390/app12042061.

[13] McKenzie, T., Meyer-Kahlen, N., Hold, C., Schlecht, S. J., and Pulkki, V., "Auralisation of measured room transitions in virtual reality," *J. Audio Eng. Soc.*, pp. 1–12, 2023, doi:10.17743/jaes.2022.0084.

[14] McKenzie, T., Meyer-Kahlen, N., Daugintis, R., McCormack, L., Schlecht, S. J., and Pulkki, V., "Perceptually informed interpolation and

rendering of spatial room impulse responses for room transitions," in *Int. Congress on Acoustics*, Gyeongju, Korea, 2022.

[15] Xiang, N., Jing, Y., and Bockman, A. C., "Investigation of acoustically coupled enclosures using a diffusion-equation model," *J. Acoust. Soc. Am.*, 126(3), pp. 1187–1198, 2009, doi: 10.1121/1.3168507.

[16] McKenzie, T., Schlecht, S. J., and Pulkki, V., "The auditory perceived aperture position of the transition between rooms," *J. Acoust. Soc. Am.*, 152(3), pp. 1871–1882, 2022, doi: 10.1121/10.0014178.

[17] Hidaka, T., Yamada, Y., and Nakagawa, T., "A new definition of boundary point between early reflections and late reverberation in room impulse responses," *J. Acoust. Soc. Am.*, 122(326), 2007, doi:10.1121/1.2743161.

[18] Hold, C., McKenzie, T., Götz, G., Schlecht, S. J., and Pulkki, V., "Resynthesis of spatial room impulse response tails with anisotropic multi-slope decays," *J. Audio Eng. Soc.*, 70(6), pp. 526–538, 2022, doi:10.17743/jaes.2022.0017.

[19] Hardin, R. H. and Sloane, N. J., "New spherical designs in three and four dimensions," in *IEEE Int. Symposium on Information Theory - Proceedings*, volume 441, p. 181, 1995, doi: 10.1109/isit.1995.531530.

[20] McKenzie, T., McCormack, L., and Hold, C., "Dataset of spatial room impulse responses in a variable acoustics room for six degrees-of-freedom rendering and analysis," in *arXiv preprint*, pp. 1–3, 2021, doi:10.48550/arXiv.2111.11882.

[21] McCormack, L., Delikaris-Manias, S., Farina, A., Pinardi, D., and Pulkki, V., "Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localized sub-band sound-field analysis," in *AES 144th Conv.*, 2018.

[22] Politis, A., *Microphone array processing for parametric spatial audio techniques*, Phd thesis, Aalto University, 2016.

[23] Jo, B. and Choi, J.-W., "Parametric direction-of-arrival estimation with three recurrence relations of spherical harmonics," *J. Acoust. Soc. Am*, 145(1), pp. 480–488, 2019.

[24] McCormack, L., Politis, A., McKenzie, T., Hold, C., and Pulkki, V., "Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple Ambisonic receivers," *J. Audio Eng. Soc.*, 70(5), pp. 355–372, 2022, doi:10.17743/jaes.2022.0010.