ESTIMATING DIRECTIONAL ROOM IMPULSE RESPONSES AT NEW LISTENING POSITIONS BY INTERPOLATION OF TRIANGULATED DIRECT SOUND AND SPECULAR REFLECTION MEASUREMENTS

J Zhao Institute of Sound and Vibration Research, University of Southampton, UK
X Zheng DJI Technology Co., Ltd., Hong Kong, China
C Ritz University of Wollongong, Wollongong, NSW, Australia
H Ren University of Wollongong, Wollongong, NSW, Australia
D Jang Electronics & Telecommunications Research Institute, Daejeon, Korea

1 INTRODUCTION

Synthesising spatial audio for virtual reality requires generating dynamic 3D sound that adapts to listener movement[1,2]. Realistic reproduction necessitates accurate sound source direction and room acoustics, typically achieved by convolving dry audio with directional room impulse responses (DRIR) for binaural headphone playback.

In streaming VR, bandwidth constraints prevent continuous transmission of new audio for every listener movement. Early solutions used spatial interpolation between multiple Ambisonic recordings[3,4], though these suffer from localisation errors when sources are close[5]. Alternative extrapolation methods[6-11] decompose sound fields into wave components for positional translation, but accuracy decreases with distance from measurement points, which is a problem addressed through higher-order representations requiring impractical hardware. Parametric methods offer an alternative by separating sound fields into directional and diffuse components[12]. This perceptual approach enables position estimation through time-frequency analysis and component reconstruction, showing advantages over linear methods[5]. Similar principles could apply to collaborative source separation techniques[13].

This paper presents a parametric method for synthesising DRIRs at arbitrary listening positions using two DRIR measurements. The core innovation lies in a triangulation-based spatial analysis technique applied to decomposed acoustic components. The methodology begins by separating reference DRIRs into their First-Order Ambisonic constituents: direct sound, specular reflections and diffuse reverberation. Spatial parameters (azimuth and elevation) of direct sound components are extracted from each reference position. Through geometric triangulation of these directional measurements, the physical sound source location is determined. This enables accurate reconstruction of direct sound spatialisation for any target position.

Similarly, early specular reflections undergo directional analysis to identify corresponding image sources via triangulation. Their spatial characteristics are then projected to new listening perspectives. The complete DRIR synthesis combines reconstructed direct sound, spatially-projected specular components and residual diffuse energy. Conceptually aligned with parametric spatial audio techniques[14], our approach builds upon established frameworks for DRIR analysis[15,16]. The principal advancement replaces conventional linear interpolation with geometric triangulation for spatial parameter estimation, significantly extending the valid interpolation range beyond reference position boundaries. To the best of our knowledge, this represents the first implementation of image source triangulation for specular reflection analysis in spatial audio synthesis.

The remainder of this paper is organised as follows: Section 2 provides an overview of the proposed system and detailed introductions of the proposed method, with Section 3 conducting experiments for validation purposes and comparing experimental results with those of representative existing approaches. Contributions are concluded in Section 4.

2 THE PROPOSED SYSTEM AND METHODOLOGY

2.1 System Overview

The proposed system architecture firstly processes B-format microphone recordings through a multi-stage spatial analysis and synthesis pipeline. The input signals initially undergo acoustic preconditioning: a high-pass filter (20 Hz cutoff, aligned with human auditory thresholds) removes DC components, followed by amplitude normalisation and silence truncation. The conditioned W-channel signals are subsequently analysed to identify temporal segments containing direct sound and specular reflections. These isolated components undergo directional analysis to extract their direction-of-arrival (DOA) parameters, as detailed in Section 2.2.

Spatial intelligence is implemented through DOA measurements from multiple reference positions, which enable geometric triangulation of both primary and image sources. Following source localisation, the system computes spatial parameters for direct and specular components relative to arbitrary listener positions. Synthesis employs a parametric approach, and the W-channel content from reference positions is spatialised using projected directional information, with temporal alignment achieved through calculated delay compensation between specular components and direct sound. The architecture incorporates an adjustable specular-to-diffuse ratio (SDR) when combining reconstructed early reflections with residual reverberation, preserving perceptual characteristics when modelling finite reflection orders. Implementation specifics of the spatial projection module (blue) are elaborated in Section 2.3.

2.2 Component Extraction from FOA DRIRs

This section describes the signal model and method for separating first-order Ambisonics (FOA) signals into direct sound, specular reflections and diffuse components.

2.2.1 Signal Model

It is assumed that the sound recorded at position $L_m(x_m, y_m, z_m)$ in a noise-free environment to be:

$$p_{L_m} = \sum_{j=0}^{J-1} s_{L_j} * h_{L_j \to L_m}, \tag{1}$$

where s_{L_j} is the *j*th source originating at position $L_j = (x_j, y_j, z_j)$, $h_{L_j \to L_m}$ is the corresponding RIR between the source at position L_j and recording position $L_m = (x_m, y_m, z_m)$. The RIR can be modelled as the sum of the direct sound, early specular reflections and late diffuse components as:

$$h_{L_j \to L_m} = h_{L_j \to L_m, dir} + h_{L_j \to L_m, early} + h_{L_j \to L_m, late}. \tag{2}$$

The first objective is to estimate each component of (2) from recordings of the RIR at specific positions. Here it is assumed that $h_{L_j \to L_m}$ of (2) is recorded using a B-format microphone, resulting in a DRIR recording with four channels: an omnidirectional channel (W) and three mutually perpendicular figure-of-eight channels (X, Y, Z). Alternatively, the B-format DRIR can be obtained through simulation of 4 orthogonally arranged microphones as standard A-format and converting these to B-format using standard equations[17,18]. This results in the B-format signals

$$B_{j,m}(n) = \left(X_{j,m}(n), Y_{j,m}(n), Z_{j.m}(n), W_{j,m}(n)\right) \tag{3}$$

for a source j located at L_j and recorded at position L_m.

2.2.2 Identifying the Direct Sound and Specular Reflections

The B-format DRIR are separated into direct sound, specular reflections and diffuse components at each recording position. This is achieved by analysing the W channel of the B-Format DRIR to find short segments of signals containing local peaks that are compared against energy thresholds. As

described in the work[16], this is inspired by approaches to tracking noise levels in audio signal processing[19], which aims to distinguish between short-duration impulsive noise and longer duration increases in the noise level. Here, the approach uses two filters applied to the power of the W channel of the FOA RIR, $P_w(n) = W_{j,m}(n)^2$. The two filters are designed based on window functions that analyse shorter and longer segments and produce signals representing fast tracking and slow tracking of the DRIR power, $P_w^{fast}(n)$ and $P_w^{slow}(n)$. The ratio of the these two power estimates, $R_{dB}(n)$, as given by (4), is analysed to find local peaks at $n = {n_{peak}}^i$, where $R_{dB}(n)$ exceeds a minimum threshold and also where the DRIR power $P_w(n)$ is above a minimum noise floor. Further details can be found in the work[16]. Once the local peaks are identified, an estimated direct sound or specular reflection is determined as the segment of the DRIR surrounding the peak at $n = {n_{peak}}^i$ and where $R_{dB}(n)$ exceeds a threshold γ, chosen based on informal experiments[16].

$$R_{dB}(n) = 10 * log_{10} \frac{P_w^{fast}(n)}{P_w^{slow}(n)}. \tag{4}$$

The approach is applied to the omnidirectoinal (W-channel) of the B-format recorded RIRs, $W_{j,m}(n)$ of (3) This results in a set of *I* specular reflections, which are denoted $W_{,j,m}^{i}(n_i), i = 0\ to\ I, n_i = (n_i - \tau_i/2, n_i + \tau_{i,}/2)$, where *i* = 0 is the direct sound, *I* is chosen based on a desired number of early specular reflections, n_i is the time of the peak of the estimated direct sound or specular reflection, *i*, and τ_i corresponds to the duration of the direct sound or specular reflection segment of $W_{j,m}(n)$. Finally, the B-format representation of these direct sound and specular reflections are formed by extracting the segments from the same positions in the X, Y and Z channels and represented as

$$B_{j,m}^i(n) = \left(X_{j,m}^i(n), Y_{j,m}^i(n), Z_{j,m}^i(n), W_{j,m}^i(n)\right). \tag{5}$$

2.2.3 Deriving DOA for FOA Encoded Direct Sound and Specular Reflections

The DOA of each FOA-encoded direct sound or specular reflection is determined using a time-frequency approach as used for sound source DOA estimation applied to B-format recordings[13] but applied here to the B-format direct sound and specular reflection signals. These B-format signals $B_{j,m}^i(n)$ are first converted to the time-frequency domain using the short-term Fourier transform (STFT), resulting in $B_{j,m}^i(n,k) = \left(X_{j,m}^i(n,k), Y_{j,m}^i(n,k), Z_{j,m}^i(n,k), W_{j,m}^i(n,k)\right)$, where *k* denotes discrete frequency. Here, a 1024-long FFT with 50% overlapping Kaiser-Bessel Derived windows are used. The azimuth $\theta_{j,m}^i(n,k)$ and elevation $\phi_{j,m}^i(n,k)$ of each time-frequency instant, (n,k), is given by

$$\theta_{j,m}^i(n,k) = tan^{-1}\left(\frac{Re\left\{W_{j,m}^i(n,k)\cdot Y_{j,m}^i(n,k)\right\}}{Re\left\{W_{j,m}^i(n,k)^*\cdot X_{j,m}^i(n,k)\right\}}\right), \tag{6}$$

$$\phi_{j,m}^i(n,k) = tan^{-1}\left(\frac{-Re\left\{W_{j,m}^i(n,k)^*\cdot Z_{j,m}^i(n,k)\right\}}{\sqrt{Re\left\{W_{j,m}^i(n,k)^*.X_{j,m}^i(n,k)\right\}^2 + Re\left\{W_{j,m}^i(n,k)^*.Y_{j,m}^i(n,k)\right\}^2}}\right). \tag{7}$$

Following calculation of these time-frequency DOA estimates, a histogram approach is used to find the final DOA estimate for this sound source or image source. For azimuth estimation, a histogram is formed by counting the number of azimuths of (6) within bins of chosen resolution (e.g. 1 degree) and this is then analysed to find peaks corresponding to the final DOA estimate. A similar approach is applied to estimating the elevation. This time-frequency approach is motivated by previous research into DOA estimation applied to reverberant multi-source signals[13]. The analysis is conducted for the direct sound and each specular reflection extracted above, which is calculated up to a desired number of times, *I* + 1.

2.3 Deriving the Interpolated FOA Signals at Virtual Listening Positions

It is assumed that for a given room a set of regularly spaced DRIRs have been obtained through recording or synthesis as B-format. The proposed triangulation approach will be explained in detail here.

It is first assumed that the 3D coordinates positions of measurements $B_{j,1}(n)$ and $B_{j,2}(n)$ for microphones *m*=1 and *m*=2 are known but the source position is unknown. It is also assumed that the estimated azimuth and elevation of the DOA of the direct sound at each of the measurement positions, (α_1, β_1) and (α_2, β_2), have been determined using the DOA estimation methods described in previous reports. The source position $L_0 = (x_0, y_0, z_0)$ can be determined by solving first for the horizontal plane (x_0 and y_0)

$$tan\alpha_1 = \frac{x_1 - x_0}{y_1 - y_0}, \tag{8}$$

$$tan\alpha_2 = \frac{x_2 - x_0}{y_2 - y_0}. \tag{9}$$

Thus,

$$x_0 = \frac{y_2 - y_1 - (tan(\alpha_2)\cdot x_2 - tan(\alpha_1)\cdot x_1)}{tan(\alpha_1) - tan(\alpha_2)}, \tag{10}$$

$$y_0 = tan(\alpha_1)\cdot x_0 - tan(\alpha_1)\cdot x_1 + y_1, \tag{11}$$

$$z_0 = \begin{cases} z_1 + tan(\beta_1)\cdot\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}, & \beta_1 > 0 \\ z_1 - tan(\beta_1)\cdot\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}, & otherwise \end{cases}. \tag{12}$$

The distances between the source and the two FOA microphones can be determined as

$$d_1{}^2 = (x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2, \tag{13}$$

$$d_2{}^2 = (x_2 - x_0)^2 + (y_2 - y_0)^2 + (z_2 - z_0)^2. \tag{14}$$

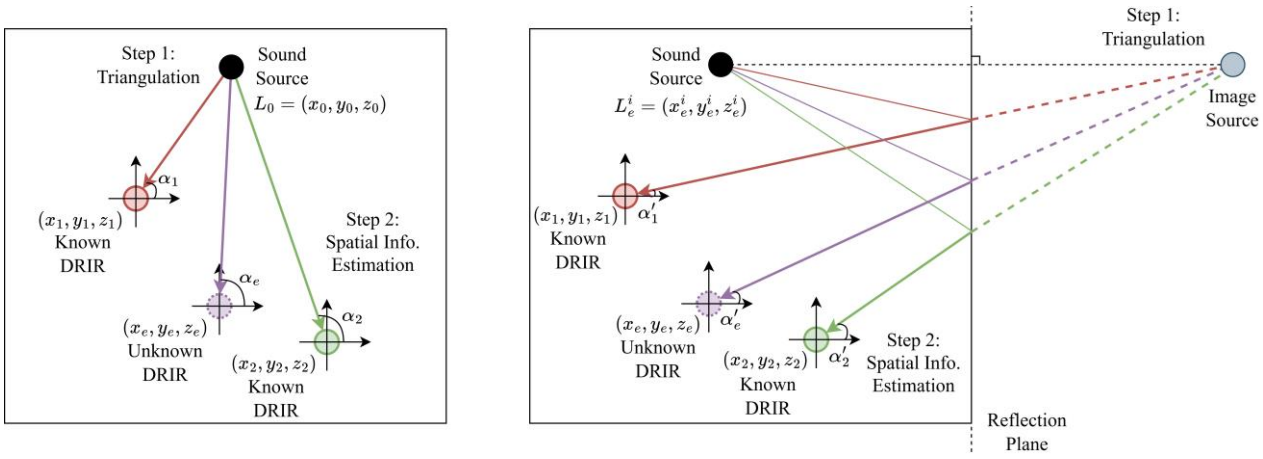
This process can be repeated to find the estimated virtual position of the image sources, as illustrated in Figure 1 (b), but using the estimated DOAs of the corresponding image sources at each measured position. To ensure the same image source is being found at each measured position, these are chosen based on assumed knowledge of the room size, which allows for estimating the position of each image source, and assuming that the image sources corresponding to the early reflections arrive at each FOA recording position at similar time. Future work may explore more sophisticated approaches based on correlation or other techniques used to confirm the similarity of extracted image sources at each position.

2.3.1 Finding Positions of the Direct Sound and Specular Reflections

The process above estimates the positions of the sound source and virtual image sources relative to the new listening position. This in turn allows for estimating the DOA of both the direct sound and extracted specular reflections relative to the new listening position. For the direct sound, once the sound source position is known, the interpolated azimuth and elevation, (α_e, β_e) can be found using the new listening position $L_e = (x_e, y_e, z_e)$ as

$$\alpha_e = arctan(\frac{y_e - y_0}{x_e - x_0}), \tag{15}$$

$$\beta_e = arctan(\frac{z_e - z_0}{\sqrt{(x_e - x_0)^2 + (y_e - y_0)^2 + (z_e - z_0)^2}}). \tag{16}$$



(a) (b)

Figure 1. Using triangulation to find sound source (a) and image sources (b) at a new position

This process is then repeated for each of the extracted image sources to obtain the set of interpolated DOAs $\left(\alpha_{j,e}^i, \beta_{j,e}^i\right), i = 1\ to\ I$ where *I* is the desired number of early specular reflections. In Figure 1 (b), this is illustrated for one specular reflection off the right wall and with an estimated azimuth of α_e'.

2.3.2 Forming the B-format DRIR at New Listening Positions

The B-format DRIR that includes the estimated direct sound and each specular reflection at a new listening position can be formed from the W channel of their FOA representations

$$\begin{cases} W_{j,e}^i = \frac{W_{1,e}^i}{\sqrt{2}} \\ X_{j,e}^i = W_{j,e}^i \cdot cos\left(\alpha_{j,e}^i\right) \cdot cos\left(\beta_{j,e}^i\right) \\ Y_{j,e}^i = W_{j,e}^i \cdot sin\left(\alpha_{j,e}^i\right) \cdot cos\left(\beta_{j,e}^i{}_e\right) \\ Z_{j,e}^i = W_{j,e}^i \cdot sin\left(\beta_{j,e}^i\right) \end{cases}. \quad (17)$$

The time delay of arrival (in samples) of each specular reflection with respect to the direct sound at the new listening position can be estimated using the distance difference between the propagation path of the direct sound and that of the specular reflection:

$$n_{delay}^i = \left(\sqrt{(x_e - x_e^i)^2 + (y_e - y_e^i)^2 + (z_e - z_e^i)^2} - \sqrt{(x_e - x_0)^2 + (y_e - y_0)^2 + (z_e - z_0)^2}\right) \cdot \frac{fs}{c}, \quad (18)$$

where the positions of the sound source, image source and the listening position are $L_0 = (x_0, y_0, z_0)$, $L_e^i = (x_e^i, y_e^i, z_e^i)$ and $L_e = (x_e, y_e, z_e)$, repectively. fs is the sample rate and c is the speed of sound propagation in air. The final estimated B-format DRIR at the new listening position that includes the direct sound and specular reflections is formed by concatenating these time delay compensated specular B-format DRIRs and represented as $B_{j,m}^s(n) = \left(X_{j,m}^s(n), Y_{j,m}^s(n), Z_{j,m}^s(n), W_{j,m}^s(n)\right)$.

The final step is merging the interpolated specular B-format DRIR with the diffuse component. Since the diffuseness is generally regarded as non-directional, we employed the diffuse component of one of the known DRIRs used in the interpolation. An important step is to ensure that the SDR for the interpolated DRIR matches a target SDR, which is defined as the ratio of the direct sound and specular reflections to the diffuse component of the DRIR. Here, it is assumed that the SDR remains relatively constant within the region where the interpolated and known DRIRs are located and the target SDR is determined as the average SDR estimated from the known DRIRs. This is calculated as

$$\overline{SDR} = \frac{1}{2}\sum_{m=1}^{2} \frac{\sum W_{j,m}^{S}(n) \cdot W_{j,m}^{S}(n)}{\sum\left(\left(W_{j,m}^{D}(n)\right) \cdot \left(W_{j,m}^{D}(n)\right)\right)}, \tag{19}$$

where $W_{j,m}^{D}(n) = W_{j,m}(n) - W_{j,m}^{S}(n)$ is the estimated diffuse component.

The same SDR is ensured in the interpolated B-format DRIR by adjusting the contribution of the diffuse component as

$$\begin{cases} W_{j,e}(n) = W_{j,e}^{S}(n) + W_{j,1}^{D}(n) \cdot g \\ X_{j,e}(n) = X_{j,e}^{S}(n) + X_{j,1}^{D}(n) \cdot g \\ Y_{j,e}(n) = Y_{j,e}^{S}(n) + Y_{j,1}^{D}(n) \cdot g \\ Z_{j,e}(n) = Z_{j,e}^{S}(n) + Z_{j,1}^{D}(n) \cdot g \end{cases}, \tag{20}$$

where $g = \sqrt{\frac{W_{j,e}^{S}(n) \cdot W_{j,e}^{S}(n)}{W_{j,1}^{D}(n) \cdot W_{j,1}^{D}(n) \cdot \overline{SDR}}}$.

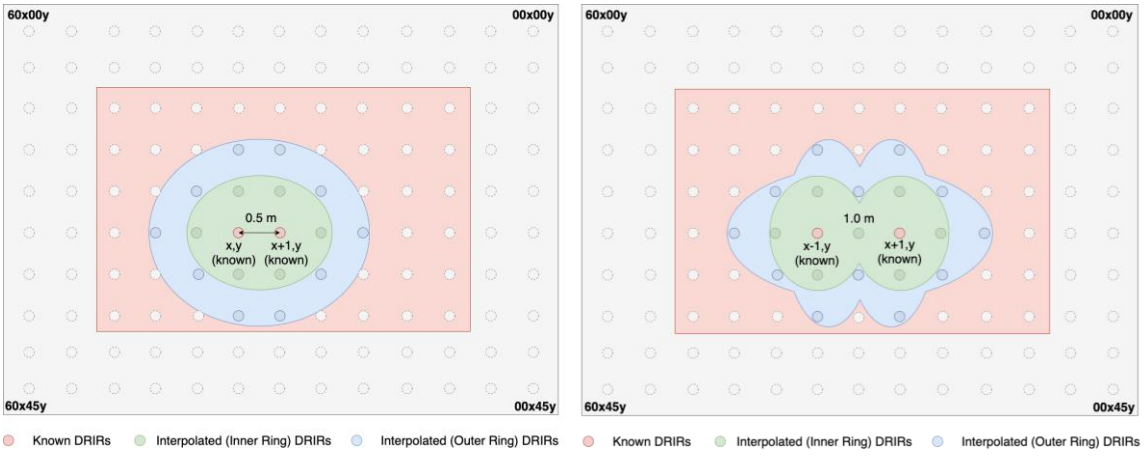
3 EXPERIMENTAL SETUP AND RESULTS

3.1 Datasets and Experimental Setup

Evaluation employs both measured and simulated B-format DRIRs. Real measurements[20] were captured in a 236 m³ classroom (9 m × 7.5 m × 3.5 m) using a Soundfield SPS422B microphone array at 96 kHz. A total of 130 positions were arranged in a 6 m × 6 m grid with 50 cm spacing. Simulated DRIRs were generated using RIR Generator[21] with comparable room dimensions and a configured reverberation time of 1.8 seconds. B-format signals were derived from four hypercardioid microphone recordings at each position via equation (3).

Two configurations were designed to experiment the proposed triangulation method, which are

- Condition A: Interpolation using adjacent DRIR pairs (red positions in Figure 2a) to estimate positions in inner (green) and outer (blue) regions.
- Condition B: DRIR pairs spaced 1 m apart (Figure 2b) for more challenging extrapolation tasks within similarly defined regions.



(a) (b)

Figure 2. Experiment setup: (a) Condition A: interpolation using adjacent DRIRs (0.5m apart) considering two regions, the inner ring (green) and outer ring (blue); (b) Condition B: interpolation using distant DRIRs (1.0m apart) considering two regions, the inner ring (green) and outer ring (blue).

3.2 Comparative Results

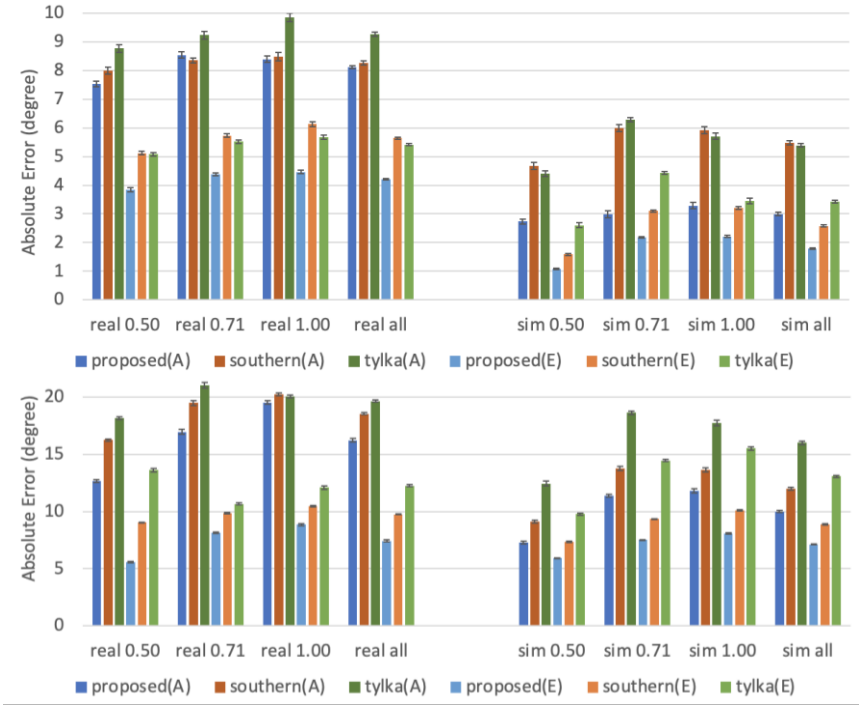


Figure 3. Azimuth and elevation average errors across all situations for different inter-microphone distance with a 95% confidence interval (CI) for: (a) the direct sound; (b) the five earliest reflections.

The experimental results are summarised into Figure 3 where the positions with the same distance to the known positions are grouped together, which indicates that the proposed system outperforms the existing systems by having the lowest errors in most real-world cases and all simulated cases for the direct sound. In Figure 3 (b), the absolute errors are increased compared to the direct sound. In terms of the grouped accuracy for each position with the same distance to the known positions, the proposed system achieved relatively lower absolute errors for all positions compared to the existing two methods. In general, the “Outer Ring” results show higher errors than “Inner Ring” results.

4 CONCLUSIONS

This paper proposes a parametric approach for dynamic spatial audio reproduction utilising sound field measurements. DRIRs are firstly decomposed into direct sound, a series of early specular reflections and diffuse components. Subsequently, the proposed method triangulates sound sources and image sources to estimate the interpolated direct sound and early specular reflections at a new listening position. Experiments compare the proposed algorithm with two existing representative methods for interpolation, showing that the proposed triangulation method obtains more accurate interpolation of DOAs of the direct sound and specular reflections in most scenarios. Future work will consider dynamic spatial audio reproduction at new virtual listening positions using the proposed DRIR interpolation approach and knowledge of listener head orientation and source positions.

5 ACKNOWLEDGMENT

This research was supported by the Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government, grant number 22ZH1200 (The research of the basic media contents technologies).

6 REFERENCES

1. T. Lentz, D. Schröder, M. Vorländer and I. Assenmacher, “Virtual Reality System with Integrated Sound Field Simulation and Reproduction,” *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, Art. no. 1, Dec. 2007, doi: 10.1155/2007/70540.
2. F. Zotter, M. Frank, C. Schörkhuber and R. Höldrich, “Signal-independent approach to variable-perspective (6DoF) audio rendering from simultaneous surround recordings taken at multiple perspectives,” presented at the Fortschritte der Akustik - DAGA, Mar. 2020.
3. N. Mariette and B. Katz, “SoundDelta - largescale, multi-user audio augmented reality,” in *EAA Symp. on Auralization*, Jan. 2009, pp. 1–6.
4. A. Southern, J. Wells and D. Murphy, “Rendering walk-through auralisations using wave-based acoustical models,” in *2009 17th European Signal Processing Conference*, Aug. 2009, pp. 715–719.
5. J. G. Tylka and E. Y. Choueiri, “Fundamentals of a parametric method for virtual navigation within an array of ambisonics microphones,” *J. Audio Eng. Soc.*, vol. 68, no. 3, pp. 120–137, Mar. 2020, doi: https://doi.org/10.17743/jaes.2019.0055.
6. E. Fernandez-Grande, “Sound field reconstruction using a spherical microphone array,” *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1168–1178, Mar. 2016, doi: 10.1121/1.4943545.
7. D. Menzies and M. Al-Akaidi, “Nearfield binaural synthesis and ambisonics,” *J. Acoust. Soc. Am.*, vol. 121, no. 3, pp. 1559–1563, Mar. 2007, doi: 10.1121/1.2434761.
8. “Analysis and synthesis of sound-radiation with spherical arrays.” https://iem.kug.ac.at/en/projects/workspace/projekte-bis-2008/dsp/analysis-and-synthesis-of-sound-radiation-with-spherical-arrays.html (accessed Dec. 16, 2021).
9. Dylan Menzies and Marwan Al-akaidi, “Ambisonic synthesis of complex sources,” *J. Audio Eng. Soc.*, vol. 55, no. 10, pp. 864–876, Oct. 2007.
10. Y. Wang and K. Chen, “Translations of spherical harmonics expansion coefficients for a sound field using plane wave expansions,” *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3474–3478, Jun. 2018, doi: 10.1121/1.5041742.
11. J. G. Tylka and E. Y. Choueiri, “Performance of linear extrapolation methods for virtual sound field navigation,” *J. Audio Eng. Soc.*, vol. 68, no. 3, pp. 138–156, Mar. 2020.
12. O. Thiergart, G. Del Galdo, M. Taseska and E. A. P. Habets, “Geometry-based spatial sound acquisition using distributed microphone arrays,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 12, pp. 2583–2594, Dec. 2013, doi: 10.1109/TASL.2013.2280210.
13. X. Zheng, C. Ritz and J. Xi, “Encoding and communicating navigable speech soundfields,” *Multimed. Tools Appl.*, vol. 75, no. 9, pp. 5183–5204, May 2016.
14. O. Puomio, N. Meyer-Kahlen and T. Lokki, “Locating image sources from multiple spatial room impulse responses,” Appl. Sci., vol. 11, no. 6, Art. no. 6, Jan. 2021.
15. K. Müller and F. Zotter, “Auralization based on multi-perspective ambisonic room impulse responses,” *Acta Acust.*, vol. 4, no. 6, Art. no. 6, 2020, doi: 10.1051/aacus/2020024.
16. J. Zhao, X. Zheng, C. Ritz and D. Jang, “Interpolating the directional room impulse response for dynamic spatial audio reproduction,” *Appl. Sci.*, vol. 12, Jan. 2022.
17. “SoundField | Microphones and processors with unique surround sound capabilities.” http://www.soundfield.com (accessed Jun. 27, 2021).
18. M. Dabin, C. Ritz and M. Shujau, “Design and analysis of miniature and three tiered B-format microphones manufactured using 3D printing,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 2674–2678. doi: 10.1109/ICASSP.2015.7178456.
19. G. Ma and C. P. Brown, “Noise level estimation,” WO2015191470A1, Dec. 17, 2015 Accessed: Dec. 16, 2021. [Online]. Available: https://patents.google.com/patent/WO2015191470A1/en?inventor=guilin+ma&oq=guilin+ma+&page=2.
20. R. Stewart and M. Sandler, “Database of omnidirectional and B-format impulse responses,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010)*, Dallas, TX, USA, 14–19 March 2010.
21. RIR-Generator. [Online]. Available: https://github.com/ehabets/RIR-Generator (accessed on 25 July 2022).