24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27JULY 2017
LONDON CALLING

MICROPHONE ARRAY DESIGN FOR SPATIAL AUDIO OBJECT EARLY REFLECTION PARAMETRISATION FROM ROOM IMPULSE RESPONSES

Miguel Blanco Galindo, Philip J.B. Jackson, Philip Coleman and Luca Remaggi
University of Surrey, Centre for Vision, Speech and Signal Processing, Guildford, UK
email: m.blancogalindo@surrey.ac.uk

Room Impulse Responses (RIRs) measured with microphone arrays capture spatial and non-spatial information, e.g. the early reflections' directions and times of arrival, the size of the room and its absorption properties. The Reverberant Spatial Audio Object (RSAO) was proposed as a method to encode room acoustic parameters from measured array RIRs. As the RSAO is object-based audio compatible, its parameters can be rendered to arbitrary reproduction systems and edited to modify the reverberation characteristics, to improve the user experience. Various microphone array designs have been proposed for sound field and room acoustic analysis, but a comparative performance evaluation is not available. This study assesses the performance of five regular microphone array geometries (linear, rectangular, circular, dual-circular and spherical) to capture RSAO parameters for the direct sound and early reflections of RIRs. The image source method is used to synthesise RIRs at the microphone positions as well as at the centre of the array. From the array RIRs, the RSAO parameters are estimated and compared to the reference parameters at the centre of the array. A performance comparison among the five arrays is established as well as the effect of a rigid spherical baffle for the circular and spherical arrays. The effects of measurement uncertainties, such as microphone misplacement and sensor noise errors, are also studied. The results show that planar arrays achieve the most accurate horizontal localisation whereas the spherical arrays perform best in elevation. Arrays with smaller apertures achieve a higher number of detected reflections, which becomes more significant for the smaller room with higher reflection density.

Keywords: microphone arrays, spatial room impulse responses

1. Introduction

Room Impulse Response (RIR) measurements are one of the most common methods to estimate acoustic properties of a room. A RIR measured with a single microphone provides a representation of the acoustic energy density as a function of time at a specific position in space. From the RIR, objective measures can be estimated (e.g. the reverberation time (T_{60})), some of which may provide physical information (such as the times of arrival (TOAs) of the early reflections), or might relate to perceptual attributes considered useful in concert hall acoustics (for instance, intimacy [1, p. 513]). Measuring RIRs with a microphone array additionally allows spatial information to be extracted, such as the directions of arrival (DOAs) of the early reflections.

The information contained within spatial RIRs can be used to recreate reverberation over a spatial audio system. An overview of approaches to reverberation synthesis can be found in [2]. Traditionally, spatial sound has been consumed in a channel-based format where the content is produced with prior knowledge of the reproduction setup. In this case reverberation is fixed for a particular loudspeaker

layout. Object-based audio, on the other hand, allows a format-agnostic representation, where the audio sources and their metadata form audio objects, which can then be edited before rendering the scene through arbitrary reproduction systems [3]. In this context, reverberation may also be rendered to different loudspeaker layouts and edited in production and reproduction to enhance the user experience, provided it can be parametrised.

A summary of the most relevant parametric reverb methods as well as their advantages and disadvantages with respect to object-based audio are given in [4]. Among those that measure RIRs using microphone arrays, the Spatial Impulse Response Rendering (SIRR) encodes the DOA and the diffuseness coefficient from B-format RIRs for each time-frequency window [5]. However, some of its parameters are not intuitively editable during encoding [4]. The Spatial Decomposition Method (SDM) determines the DOAs of the image sources from array RIRs for each time window along the entire RIR, which is used to spatialise a reference omnidirectional RIR [6]. Plane-Wave Decomposition (PWD) has also been used to encode reflections [7, 8], which are easily editable. However, the low-level parameter description in SDM and PWD requires a large number of plane waves to be encoded and a high-count loudspeaker setup for accurate diffuse rendering of late reverberation [4].

The Reverberant Spatial Audio Object (RSAO) encodes low-level parameters of the early reflections (assumed to be specular) based on their DOA, TOA amplitude and frequency shaping whilst modelling the late reverberation as the energy decay for each octave band [9]. Thus, RSAO has the advantage that the early reflections can be edited from the physical parameters of the room measurement but can also be abstracted to higher-level descriptions such as envelopment or intimacy. The late reverberation can potentially be edited based on a physical model (relating to the room volume and surface absorption) or mapping to perceptual attributes or direct-to-reverberant ratio [4].

The above approaches used various microphone array geometries to test and evaluate the proposed parameterisation methods. The objective of this letter is to compare the performance of different microphone array designs to capture early reflection parameters. The RSAO is chosen as the framework to analyse and extract the early reflection parameters due to its format-agnostic approach and intuitive analysis and edition of its parameters.

2. Microphone array designs

The performance of array processing methods is determined by the microphone array design in combination with the signal processing technique. There exist many possible designs for pressure microphone arrays. Uniform linear arrays are commonly used due to their simplicity but are unable to resolve the DOA in three dimensions (due to unavoidable front-back and elevation ambiguities). Horizontal planar arrays also feature up-down confusion. Some microphone array geometries are linked to specific applications or signal processing techniques. For instance, spherical/circular harmonic decomposition (SHD/CHD) are usually applied to uniform spherical/circular microphone arrays, respectively, to map the pressure signals to their equivalent harmonics. The same applies to PWD as the latter is a particular case of the SHD [8, 10] and therefore they share the same challenges in terms of the microphone array design.

In particular, all circular/spherical designs based on SHD and PWD are sensitive to noise at $kr \leq n$ - where k is the wave number, r is the sphere radius and n is the spherical harmonic order - due to the low condition number of the high-order Bessel functions. For the same reason, open spherical and circular microphone arrays are also ill-conditioned at $kr \geq n$ [11]. The latter can be overcome by using solid spheres. However, large solid spheres may not be realisable practically [12], unlike open spheres, and may alter the sound field in room acoustics due to the scattering of the baffle [11].

Alternatives to the solid sphere to overcome the singularities of open spheres at high frequencies are dual- and multiple-radius spheres/circles [12, 13] or a combination of pressure and velocity microphones [14, 15]. However, they require twice as many microphones and pressure-gradient microphones are more sensitive to noise at low frequencies and their directivity pattern is frequency

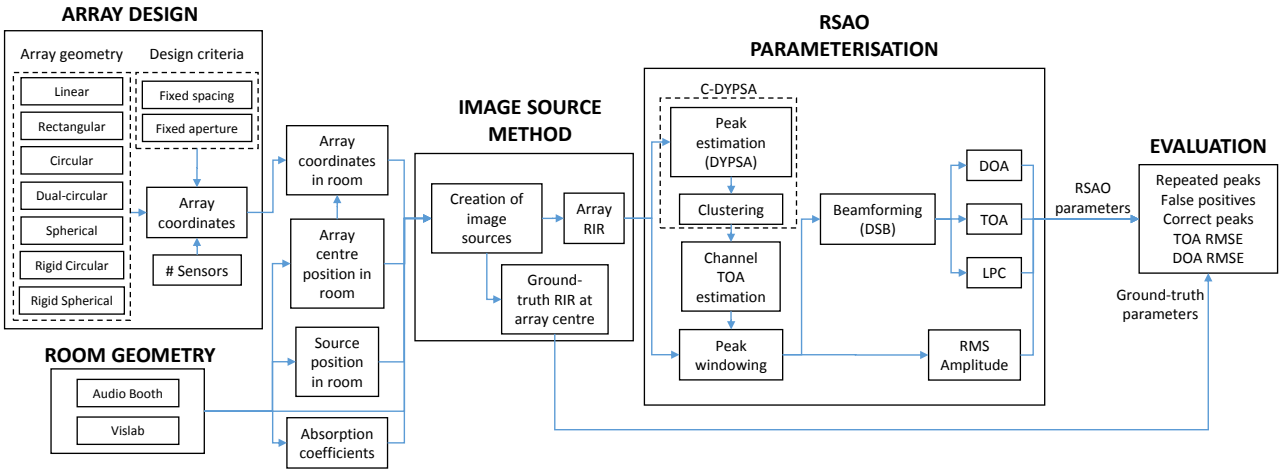

Figure 1: Block diagram of the simulation tool built

dependent. The spherical shell was proposed by Rafaely where microphones are placed within the volume of two spheres with equal sensor spacing [16]. For spherical arrays with non-uniformly distributed microphones, an optimisation method was proposed to approximate the beampattern of an equivalent modal beamformer [17]. To the authors' knowledge, the only study comparing array geometries showed that the angular resolution of spherical and rectangular arrays are similar using PWD when the rectangular side equates the sphere diameter [10].

From the above it is seen that first, SHD imposes certain limitations to the array design. Secondly, it is unclear how different array geometries would perform to capture reverberation parameters as there are no comparative studies. This is because signal processing methods tend to be prioritised over microphone array designs. Hence, this study evaluates the performance of five different open array geometries (linear, rectangular, circular, dual-circular and spherical) and two rigid baffled arrays (circular and spherical) to capture early reflection RSAO parameters.

3. Methodology

A simulation tool was implemented to undertake the comparative analysis of different microphone arrays capturing RSAO parameters. The first part is the initialisation stage where the microphone array design and room are input variables. The specific array geometry, number of sensors and spacing can be selected as well as the position of the centre of the array in the room. These parameters are used to generate RIRs up to second-order reflections using the Image Source Method (ISM) both at the array microphone positions as well as at the centre of the array. The array RIRs are passed on to the RSAO parametrisation module which extracts the reverb parameters. Finally, RSAO parameters are compared against the ground truth at the centre of the array by converting them to measurable metrics before calculating the errors. A block diagram of the simulation tool is shown in Figure 1.

The ISM, originally presented in [18], is a widely used method to simulate RIRs by mirroring the sound source through each boundary, creating equivalent "free field" image sources. Petersen proposed a method to model reflections arriving at inter-sample time instants [19], which is important for microphone array applications. Jarret et al. proposed a method to model both open array RIRs in the frequency domain and rigid-baffled array RIRs in the frequency domain using the SHD [20], which is adopted in this study. Reflections up to second order only are modelled - since the focus of this work is on the early reflections - giving a total of 24 image sources plus the direct sound. As the direct sound and the early reflections are encoded in the same way, hereafter the term early reflections will also include the direct sound. Octave band absorption coefficients were obtained by selecting appropriate materials from [21] to match the boundary finishes of the actual rooms being simulated.

3.1 Reverberant Spatial Audio Object (RSAO)

The RSAO is a method to parametrise measured RIRs. The original implementation was presented by Remaggi et al. [9], consisting of the following steps: First, a peak detection for each microphone channel is performed using the dynamic programming phase-slope algorithm (DYPSA) [22], which is then refined by an amplitude threshold of 25 dB below the maximum peak, a minimum separation between peaks of 2 ms and a group delay slope of 0.2. Then the TOA estimates are clustered (C-DYPSA [23]) for all microphone channels deleting also potential outliers. Only the first two reflections apart from the direct sound are encoded. For each reflection, the average TOA across all channels is the encoded TOA, which is also used as the reference point to window the RIRs [23] using a fixed-length window, before being sent to the delay-and-sum beamformer (DSB) [24] to obtain the DOA. The frequency shaping of the reflections is approximated from the windowed RIRs using 17 linear predicitive coding (LPC) coefficients [25]. The encoded frequency response and energy are the average LPC and windowed RIRs across all channels, respectively. The late reverberation is encoded as the energy decay per octave band.

Modifications of the state-of-the-art implementation have been made for this study. First, since the interest of this study relies on the early reflections, the late reverberation is not parametrised but the total number of peaks to be detected is increased to 25. Similarly, the minimum separation to detect adjacent peaks was reduced to 20 samples (i.e. 0.4 ms at 48 kHz) in order to enhance the peak detection time resolution. The amplitude threshold was set to -35 dB of the direct sound at the centre of the array to improve the dynamic range while limiting noise peaks being detected when modelling sensor noise. The clustering of C-DYPSA is modified to allow several reflections of a given channel to be shifted at once if all their differences with respect to the median of their previous or following reflections are smaller than their own median. This is required as DOA for consecutive reflections may not be the same, leading to some channels being able to detect both reflections when they are more than 20 samples apart, but not for other channels, therefore requiring realignment. Positions of elements after being moved are left empty unless being occupied by another element. Subsequently, reflections are checked to be detected by a minimum of 12 microphones to improve its robustness.

The resulting TOA matrix is used to calculate a range within which the RIRs will be windowed. This is given by $TOA_{range} = [TOA_{mid} - 3/4T_{aperture}, \quad TOA_{mid} + 3/4T_{aperture}]$, where $TOA_{mid} = (TOA_{max} - TOA_{min})/2$, TOA_{min} and TOA_{max} are the minimum and maximum TOAs among all microphone channels for a given detected reflection, and $T_{aperture}$ is the time it takes for the sound to travel between the two furthest microphones of the array. The selection of this TOA_{range} improves the likelihood that the reflection will be included in all channels even when the TOA_{max} and TOA_{min} do not correspond exactly to the nearest and furthest microphones for that reflection. Using TOA_{range} the RIRs are windowed around each reflection and beamformed to extract the TOA of the array manifold in addition to the DOA, which has proven to be a more robust prediction than the average TOA among all channels. LPC coefficients are encoded from the beamformed signal.

4. Simulations

A 48-microphone dual-circular array was used by Remaggi et al. [9] to capture RSAO parameters. Here the aim is to compare the performance of five open array configurations (linear, rectangular, circular, dual-circular and spherical) and two arrays (circular and spherical) mounted on a spherical rigid baffle, capturing RSAO parameters, using the dual-circular array as the benchmark. The latter features 48 microphones uniformly spaced over two circumferences of radii 0.085 and 0.107 m. Two design criteria were chosen to compare the different array geometries: fixing the number of microphones and the sensor spacing (distance between two closest microphones) and fixing the number of microphones and the array aperture (maximum size of the array). The equivalent array aperture or microphone spacing for each array geometry depending on the design criterion is given in Table 1. The rectangular array was the only design to use 49 microphones for it to be an even square.

Table 1: Varying aperture and spacing of array geometries depending on design criteria

Array design criterion	**Linear**	**Rectangular**	**Circular / Rigid circular**	**Dual-circular**	**Spherical / Rigid spherical**
Fixed spacing (varying aperture)	1.37 m	0.24 m (0.17 m in x and y)	0.43 m	0.21 m	0.11 m
Fixed aperture (varying spacing)	0.005 m	0.036 m	0.014 m	0.029 m	0.056 m

Two rooms with different sizes and absorption characteristics were modelled. The smaller room (Audio Booth) of dimensions 4.12 x 4.98 x 2.1 m and absorption characteristics similar to those of a listening room has a mid-frequency T_{60} (T_{mf60}) below 0.2 s. Vislab is a larger audio-visual lab with dimensions 7.79 x 11.84 x 4.02 m and a reported $T_{mf60} < 0.4$ s although this was measured with a large curtain covering the back half of the room, reducing its effective volume. The full empty room is modelled here to obtain the reflections from the back. Equivalent positions from those taken during recordings in [23] were chosen for the centre of the microphone arrays and the source which was positioned at ($r = 1.68$ m, $\phi = -30°, \theta = 30°$) with respect to the array centre. The effect of sensor noise was also modelled for some scenarios as uncorrelated normally-distributed random noise with a variance σ^2_{noise} such that $10\log_{10}\left(\sigma^2_{noise}/x^2_{ref}\right) = -50$ dB where x_{ref} is the amplitude of the direct sound at the array centre. Deviation of the ideal microphone positions was also considered using a uniform distribution with maximum offset of ± 2 mm for each x, y and z dimensions.

4.1 Evaluation

The reference DOA and TOA are obtained directly from the ISM. Their parametrised counterparts are given by the RSAO encoding. The other two parameters are the RMS amplitude and the LPC coefficients which are used alongside the TOA to synthesise the parametrised RIR. RIRs are windowed around the TOA to evaluate each peak's energy and frequency response.

Since the number of detected peaks is often smaller than the number of total early reflections, in order to correctly assign each detected reflection to the equivalent reference reflection, the TOA_{ref} which is closest to each TOA_{param} is found. If two or more parameters have the same TOA error, the one whose DOA error is smallest is selected and the others are considered as repeated peaks or false positives, depending on whether they are below the TOA and DOA thresholds (see below). As mentioned in section 2, planar arrays feature up-down indetermination whereas the linear array is unable to resolve around a cone of confusion. This means that they could exhibit any value from the indetermination, leading to very large errors in DOA when compared to the spherical array. Thus, the planar arrays are evaluated on the upper hemisphere only. The linear array is evaluated along the front half of the equator by setting the elevation to zero and converting the azimuth to a lateral angle.

Before evaluating the overall error of each parameter for a given array design, thresholds for the DOAs and TOAs are imposed to prevent very large errors (false positives) from compromising the final root mean square error (RMSE) values. The extraction of the TOA from the DSB makes it very accurate and a threshold of $TOA_{error} \leq 0.1$ ms was found to ensure the correct reflection is captured as larger errors led to different reflections (or artefacts) being encoded instead. For the DOA thresholds, an initial simulation was carried out to assess the DOA errors for all array designs with both design criteria under ideal conditions for a single sound source in free field at 10 m from the array manifold. Azimuth and elevation errors were calculated for a sound source at every azimuth ϕ in the horizontal plane and every elevation θ in the median plane, respectively. The maximum DOA errors were 5° in azimuth for the linear array and 20° in elevation for the rectangular array, which represent errors due to the DSB only as no reflections were considered. The analysis of the equivalent array designs in both rooms under study show that the distribution of the probability density function (PDF) of the directional errors is clustered within $\pm 5°$ in azimuth and $\pm 15°$ in elevation, with

single-occurrence events at much larger angle errors. From both analyses it was determined that the former represent fine errors due to the DSB resolution and the latter are false positives which will be discarded for the RMSE evaluation. The higher elevation error for the free field case is because the room simulation does not generate reflections at every possible DOA, and thus the worst-case scenario was not found. However, for compatibility with other source and microphone positions the maximum free field elevation error was set as the threshold $DOA_{el\ error} \leq 20°$, without altering the RMSEs, and with $DOA_{az\ error} \leq 5°$. Nevertheless, there is certain relaxation in the selection of the DOA thresholds due to the PDF distribution being clustered around 0 as mentioned above and because the TOA threshold also needs to be complied with, which is often more onerous for false positive rejections.

4.2 Results

Table 2 shows the results for the Audio Booth under ideal conditions with fixed spacing and aperture designs. Fixing the spacing, the number of correct peaks is smallest for the linear array (6), followed by the circular arrays and finally the rectangular, dual-circular and spherical (16) out of a total of 25 reflections being modelled in all scenarios. It is seen that the smaller the aperture of the array the more correct reflections are detected. This is because the DSB windows the RIRs around the detected peaks based on the array aperture. Thus, large arrays are more likely to have two reflections within a given window and therefore will detect fewer reflections overall. Conversely, if the arrays are compared fixing their aperture, the number of correct peaks is very similar across all geometries.

Table 3 shows the equivalent results obtained for Vislab. The number of correct peaks increases compared to those in the Audio Booth. This is because the reflections are sparser in time due to its larger volume and more reflections can be detected independently. The baffled arrays in some cases detected fewer reflections than their open counterparts due to the additional signal-to-noise ratio (SNR) required to detect peaks by all microphones due to the shadowing of the baffle. The SNR for the peak detection was set to -35 dB of the direct sound in order to minimise noise detection when present, limiting the detection of later reflections captured by the occluded microphones of the baffled arrays. This is also seen for the fixed spacing design in the Audio Booth (Table 2).

Table 4 shows the results in the Audio Booth in presence of sensor noise or sensor offset. Under sensor noise the number of correct peaks reduced by 1 for the rectangular and rigid circular and by 2 for the dual-circular, remaining the same for the other arrays. Similarly, the effect of microphone misplacement reduces the number of correct peaks by 1 or 2, due to an increase in false positives, except for the linear and circular arrays which remained unaltered. It is worth noting that the effects of sensor noise and array offset are shown for the Audio Booth as it represents a more challenging environment due to the higher reflection density. However, smaller variations in correct peaks were seen in Vislab when modelling sensor noise or offset compared to the ideal case.

The TOA and DOA errors are very consistent for all the array designs and scenarios under study, with $TOA_{RMSE} < 0.05$ ms and $DOA_{azRMSE} < 2°$. For the Audio Booth and a fixed aperture criterion, DOA_{azRMSE} is smaller than 0.5° under ideal conditions, with the linear array having the largest error and the circular, dual-circular and spherical arrays showing the smallest errors. $DOA_{azRMSEs}$ increase very marginally when including sensor offset, proving that the DSB is a robust method against positioning errors. In terms of elevation both spherical arrays outperform all the other arrays with errors below 1° even with positioning offset. This shows the near-uniform resolution of this geometry both in azimuth and elevation. The largest DOA_{elRMSE} was given by the rigid circular array with a value close to 6°. This was found to be because of the small amplitude of the shadowed microphones to the overall maximum beamformed output signal which results in larger ambiguity near the equator. A filter-and-sum or other weighted beamforming method is expected to improve the performance of the rigid circular array. Finally, the DOA errors increase slightly for all arrays in Vislab.

Table 2: RSAO reflection accuracy for Audio Booth, ideal conditions: fixed spacing vs fixed aperture

Array geometry	Fixed spacing				Fixed aperture			
	Correct peaks	TOA RMSE [ms]	DOA az RMSE [°]	DOA el RMSE [°]	Correct peaks	TOA RMSE [ms]	DOA az RMSE [°]	DOA el RMSE [°]
Linear	6	0.03	1.68	- -	15	0.02	0.63	- -
Rectangular	14	0.02	0.60	4.04	14	0.02	0.48	2.28
Circular	11	0.03	0.30	2.02	15	0.02	0.26	3.16
Dual-circular	15	0.02	0.25	3.03	15	0.02	0.25	3.03
Spherical	16	0.02	1.16	0.94	14	0.02	0.30	0.53
Rigid circular	12	0.04	0.26	4.18	14	0.03	0.53	5.81
Rigid spherical	14	0.02	1.65	1.00	14	0.03	0.31	0.76

Table 3: RSAO reflection accuracy for Vislab, ideal conditions: fixed spacing vs fixed aperture

Array geometry	Fixed spacing				Fixed aperture			
	Correct peaks	TOA RMSE [ms]	DOA az RMSE [°]	DOA el RMSE [°]	Correct peaks	TOA RMSE [ms]	DOA az RMSE [°]	DOA el RMSE [°]
Linear	14	0.03	2.02	- -	18	0.03	1.68	- -
Rectangular	18	0.03	0.48	3.38	18	0.03	1.04	3.06
Circular	17	0.03	0.32	2.29	18	0.03	0.39	3.37
Dual-circular	18	0.03	0.42	3.83	18	0.03	0.42	3.83
Spherical	18	0.02	1.67	1.20	19	0.03	0.91	0.65
Rigid circular	14	0.04	0.24	4.83	16	0.03	0.68	5.88
Rigid spherical	18	0.02	1.67	1.33	15	0.03	0.57	0.63

Table 4: RSAO reflection accuracy for Audio Booth with fixed aperture: sensor noise vs sensor offset

Array geometry	Sensor noise				Sensor offset			
	Correct peaks	TOA RMSE [ms]	DOA az RMSE [°]	DOA el RMSE [°]	Correct peaks	TOA RMSE [ms]	DOA az RMSE [°]	DOA el RMSE [°]
Linear	15	0.02	1.00	- -	15	0.02	1.39	- -
Rectangular	13	0.02	0.42	2.34	13	0.02	0.81	2.99
Circular	15	0.02	0.00	3.04	15	0.03	0.41	3.15
Dual-circular	13	0.02	0.00	3.31	13	0.02	0.49	3.37
Spherical	14	0.03	0.32	0.60	13	0.02	0.31	0.55
Rigid circular	13	0.03	0.28	5.68	12	0.02	0.67	6.61
Rigid spherical	14	0.03	0.53	0.53	13	0.03	0.44	0.73

5. Conclusion

A method to evaluate the performance of different array designs to capture early reflection RSAO parameters has been presented. RIRs up to second-order reflection were generated using the ISM for five open arrays and two rigid-baffled arrays and their parameters were extracted and compared against a ground-truth reference. The results show that the larger the array the fewer number of correct reflections are detected, since the beamformer needs to window the RIRs based on the array aperture. Similarly, more correct reflections were detected in the larger room Vislab as their reflections are sparser in time, reducing the likelihood of several reflections falling within a given time window and therefore maximising the total number of reflections being detected. To resemble some measurement conditions, sensor noise and positioning offset were also modelled. The method seemed fairly robust to these tolerances as the number of correct reflections remained unaltered or reduced by one or two, becoming more negligible for the larger room. The smallest DOA errors in azimuth were given by the circular and dual-circular arrays of approximately 0.5° (even under non-ideal conditions), highlighting the good horizontal localisation of planar arrays, with the linear array featuring the largest and yet small error below 2°. The spherical array showed the smallest elevation error, being this very similar to that in azimuth, re-emphasising the uniform 3D localisation resolution inherent in its geometry.

REFERENCES

1. Beranek, L. L., *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, Springer Science & Business Media, New York, berlin, Heidelberg, Hong Kong, London, Milan Paris, Tokyo, second edi edn. (2004).
2. Välimäki, V., Parker, J. D., Savioja, L., Smith, J. O. and Abel, J. S. Fifty Years of Artificial Reverberation, *IEEE Transactions on Audio, Speech, and Language Processing*, **20** (5), 1421–1448, (2012).
3. Spors, S., Wierstorf, H., Raake, A., Melchior, F., Frank, M. and Zotter, F. Spatial sound with loudspeakers and its perception: A review of the current state, *Proceedings of the IEEE*, **101** (9), 1920–1938, (2013).
4. Coleman, P., Franck, A., Jackson, P. J., Remaggi, L. and Melchior, F. Object-Based Reverberation for Spatial Audio, *Journal of the Audio Engineering Society*, **65** (1/2), (2017).
5. Merimaa, J. and Pulkki, V. Spatial impulse response rendering I: Analysis and synthesis, *AES: Journal of the Audio Engineering Society*, **53** (12), 1115–1127, (2005).
6. Tervo, S. and Politis, A. Direction of Arrival Estimation of Reflections from Room Impulse Responses Using a Spherical Microphone Array, *Audio, Speech, and Language Processing, IEEE Transactions on*, **23** (10), 1539–1551, (2015).
7. Melchior, F., Sladeczek, C., Partzsch, A. and Brix, S. Design and Implementation of an Interactive Room Simulation for Wave Field Synthesis, *AES 40th International Conference: Spatial Sound*, pp. 1–8, (2010).
8. Rafaely, B., Balmages, I. and Eger, L. High-resolution plane-wave decomposition in an auditorium using a dual-radius scanning spherical microphone array., *The Journal of the Acoustical Society of America*, **122** (May), 2661–2668, (2007).
9. Remaggi, L., Jackson, P. J. B. and Coleman, P. Estimation of room reflection parameters for a reverberant spatial audio object, *Audio Engineering Society 138th Convention*, pp. 1–10, (2015).
10. Rafaely, B. Plane wave decomposition of the sound field on a sphere by spherical convolution, *The Journal of the Acoustical Society of America*, **116** (4), 2149–2157, (2004).
11. Rafaely, B. Analysis and design of spherical microphone arrays, *IEEE Transactions on Speech and Audio Processing*, **13** (1), 135–143, (2005).
12. Balmages, I. and Rafaely, B. Open-sphere designs for spherical microphone arrays, *IEEE Transactions on Audio, Speech and Language Processing*, **15** (2), 727–732, (2007).
13. Kuntz, A. and Rabenstein, R. Wave field analysis using multiple radii measurements, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 317–320, (2009).
14. Kuntz, A. and Rabenstein, R. Cardioid Pattern Optimization for a Virtual Circular Microphone Array, *EAA Symposium on Auralization*, pp. 1–4, (2009).
15. Chen, H., Abhayapala, T. D. and Zhang, W. Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis, *The Journal of the Acoustical Society of America*, **138** (5), 3081–3092, (2015).
16. Rafaely, B. The spherical-shell microphone array, *IEEE Transactions on Audio, Speech and Language Processing*, **16** (4), 740–747, (2008).
17. Li, Z. and Duraiswami, R. Flexible and optimal design of spherical microphone arrays for beamforming, *IEEE Transactions on Audio, Speech and Language Processing*, **15** (2), 702–714, (2007).
18. Allen, J. B. Image method for efficiently simulating small-room acoustics, *The Journal of the Acoustical Society of America*, **65** (4), 943, (1979).
19. Peterson, P. M. Simulating the response of multiple microphones to a single acoustic source in a reverberant room., *The Journal of the Acoustical Society of America*, **80** (5), 1527–1529, (1986).
20. Jarrett, D. P., Habets, E. A. P., Thomas, M. R. P. and Naylor, P. A. Rigid sphere room impulse response simulation: Algorithm and applications, *Journal of the Acoustical Society of America*, **132** (3), 1462–1472, (2012).
21. Vorländer, M., *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*, Springer Science & Business Media, Berlin, rwth editi edn. (2008).
22. Naylor, P. A., Kounoudes, A., Gudnason, J. and Brookes, M. Estimation of glottal closure instants in voiced speech using the DYPSA algorithm, *IEEE Transactions on Audio, Speech and Language Processing*, **15** (1), 34–43, (2007).
23. Remaggi, L., Jackson, P. J. B., Coleman, P. and Wang, W. Acoustic Reflector Localization : Novel Image Source Reversion and Direct Localization Methods, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, **25** (2), 296–309, (2017).
24. Van Veen, B. D. and Buckley, K. M. Beamforming: A versatile approach to spatial filtering, *IEEE assp magazine*, **5** (2), 4–24, (1988).
25. Makhoul, J. Linear Prediction: A Tutorial Review, *Proceedings of the IEEE*, **63** (4), 561–580, (1975).