Spatial Room Impulse Response Identification from Rotating Equatorial Microphone Arrays

Citation for the original published paper (version of record):
Deppisch, T., Ahrens, J., Amengual Gari, S. et al (2024). Spatial Room Impulse Response Identification from Rotating Equatorial Microphone Arrays. European Signal Processing Conference: 116-120. http://dx.doi.org/10.23919/EUSIPCO63174.2024.10715285

Spatial Room Impulse Response Identification from Rotating Equatorial Microphone Arrays

Thomas Deppisch, Jens Ahrens
Chalmers University of Technology
412 96 Gothenburg, Sweden
{thomas.deppisch, jens.ahrens}@chalmers.se

Sebastià V. Amengual Garí, Paul Calamia
Reality Labs Research, Meta
Redmond, WA 98052, USA
{samengual, pcalamia}@meta.com

***Abstract*—Spatial room impulse responses (SRIRs) facilitate the rendering of virtual sound sources to realistically augment a real-world acoustic scene, for example in the context of augmented reality. With the integration of microphone arrays into robotic systems and head-worn devices, there arises the need to identify the SRIRs from arrays amidst rotation. This work establishes the groundwork for addressing this issue through the introduction of a recursive least squares filter that adaptively identifies the SRIR from signals captured by a rotating equatorial microphone array and a given reference signal. The method not only compensates for the rotation of the array but exploits the rotation to increase the directional accuracy by updating the SRIR estimates in the space-continuous circular harmonic domain. Results from simulations and measurements show that the accuracy of the SRIR estimates from the method using a rotating array of four irregularly spaced microphones is comparable to measured SRIRs from static arrays with a substantially greater number of equally spaced microphones.**

***Index Terms*—Adaptive System Identification, Augmented Reality, Rotating Microphone Array, Spatial Room Impulse Response**

I. Introduction

Equatorial microphone arrays (EMAs), i.e., circular arrays with microphones arranged along the equator of a rigid sphere such as the one in Fig. 1, have recently been shown to facilitate the expansion of a horizontal projection of the sound field in spherical harmonic (SH) coefficients while requiring significantly fewer microphones than a comparable spherical array for the same azimuthal resolution [1]. EMA signals are thus compatible with established SH rendering pipelines, for example for dynamic binaural headphone rendering. EMA theory further is the basis for rendering signals from arrays with microphones on a circumferential contour around arbitrary scattering bodies such as human heads [2], [3]. The processing of EMA signals is thus highly relevant for practical applications that consider the directionality of the sound field.

Spatial room impulse responses (SRIRs) capture the linear, time-invariant, directional properties of an acoustic environment and hence contain all the required information to create virtual sound sources that realistically blend into a real-world acoustic environment. The authors have recently shown that EMAs can be used to blindly identify SRIRs from a few seconds of recorded speech [4] and that corresponding binaural responses also perform well perceptually [5].

We thank Reality Labs Research at Meta for funding this research.

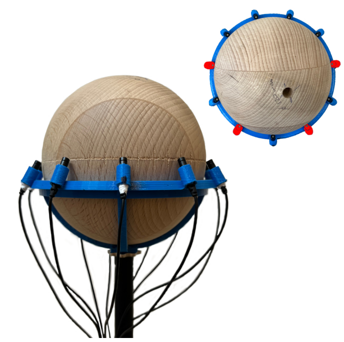

Fig. 1. Front and top view of the equatorial microphone array used in this study. The four microphones used for the SRIR identification are highlighted in red in the top view.

The present work builds on these learnings and proposes a method to adaptively identify SRIRs from rotating EMAs. While the previous works considered static arrays and focused on the blind identification of SRIRs using an estimated pseudo reference signal, this work assumes a known reference signal, or in other words, non-blind conditions, and provides a solution for rotating arrays. Similar to synthetic aperture methods, the proposed method not only compensates for the rotation of the array but exploits the rotation to achieve a higher directional resolution than the static array.

Previous research on rotating microphone arrays considered direction-of-arrival estimation [6], [7], beamforming [8]–[10], and covariance estimation [11], [12]. The processing of signals from moving arrays has been studied in synthetic aperture methods where a phase correction is applied to the array signals to achieve coherent signal summation [13]–[16]. However, the identification of room transfer functions or impulse responses has, to the best knowledge of the authors, not been attempted with any of these approaches.

The identification of room impulse responses from rotating microphone arrays has nevertheless been explored in the context of acoustic measurement methods. Approaches from the literature require rotation with constant speed and/or specific measurement signals such as periodic perfect sequences or multiple sinusoids [17]–[20]. The herein-presented method is not intended to be used for acoustic measurements under controlled conditions but for the identification of SRIRs from arbitrary signals during arbitrary azimuthal array rotations so that it may be extended to support the blind identification of SRIRs from head-worn arrays in the future. Constant rotation speed, for instance, is not required. The SRIR identification in this work is performed in the circular harmonic domain as it allows for updating a space-continuous SRIR representation during arbitrary azimuthal array rotations. The efficacy of employing spherical or circular harmonic basis functions for the processing of signals from rotating arrays has also been demonstrated in prior research [6], [11], [19].

II. Signal Model

From the perspective of the receiver, the transfer function from the sound source in a room to a sphere with radius R can be described by a continuum of plane waves. For a microphone array with $\mathcal{M}$ microphones at locations $\Theta_i = (\phi_i, \theta_i)$ on the surface of a sphere defined by the azimuth angle ϕ_i and the zenith angle θ_i, the observed sound pressures are therefore described by the length-$\mathcal{M}$ vector [21, Ch. 2.4]

$$\boldsymbol{p}(k) = \boldsymbol{Y}_{\tilde{N}} \operatorname{diag}\{\boldsymbol{b}_{\tilde{N}}(k,R)\}\, \boldsymbol{h}_{\tilde{N}}(k)\, s(k)\,. \tag{1}$$

The $\mathcal{M} \times (\tilde{N}+1)^2$ matrix $\boldsymbol{Y}_{\tilde{N}}$ contains the $(\tilde{N}+1)^2$ spherical harmonics (SHs) $Y_n^m(\Theta_i)$ of order n and degree m up to order $\tilde{N}$ evaluated at all $\mathcal{M}$ microphone locations. We use the same real-valued definitions of the spherical and circular harmonics as in [22]. The vector $\boldsymbol{h}_{\tilde{N}}(k)$ contains the $(\tilde{N}+1)^2$ SH coefficients of the plane-wave density representing the room transfer function, $s(k)$ is the spectrum of the sound source and $k = 2\pi f/c$ is the wavenumber depending on the frequency f and the speed of sound c. The $(\tilde{N}+1)^2$ terms $\boldsymbol{b}_{\tilde{N}}(k,R)$ describe the radial dependency of the sound pressure and depend on the wavenumber, radius and surface properties of the array. The coefficients do not vary with the SH degree m and thus are equal for all coefficients within the same order n. Analytical expressions for the radial terms are for example available for open and perfectly rigid spherical surfaces [21, Ch. 4]. The operator $\operatorname{diag}\{\cdot\}$ creates a diagonal matrix with $\boldsymbol{b}_{\tilde{N}}(k,R)$ on its main diagonal. In theory, infinitely many SH basis functions, $\tilde{N} \to \infty$, are required to represent the pressure $\boldsymbol{p}(k)$. However, errors due to the limitation of the SH order are rapidly decaying for $\tilde{N} > kR$ so that, for a given array, an upper frequency limit can be found below which the order-$\tilde{N}$ representation is valid [21, Ch. 4.1]. This frequency limit is commonly referred to as the spatial aliasing frequency

$$f_{\mathrm{A}} = \tilde{N}c/(2\pi R)\,. \tag{2}$$

In this contribution, we target equatorial microphone arrays (EMAs) [1]. As such arrays sample the pressure only on a circle, directly decomposing the recorded sound pressure using SH basis functions depending on azimuth and zenith angle is not feasible. The sampled pressure can, however, be expanded as a linear combination of circular harmonic (CH) basis functions $C_m(\phi_i)$ of degree m and corresponding expansion coefficients $\mathring{\boldsymbol{h}}_N(k)$,

$$\boldsymbol{C}_N \mathring{\boldsymbol{h}}_N(k) s(k) = \boldsymbol{p}(k)\,. \tag{3}$$

The $\mathcal{M} \times 2N+1$ matrix $\boldsymbol{C}_N$ contains the CHs evaluated at the microphone azimuth angles ϕ_i up to the maximum order N that is limited by the number of microphones in the array and is typically much smaller than the order of the sound field, $N \ll \tilde{N}$. SHs and CHs are closely related, with the CHs describing the azimuthal dependency of the SHs and only depending on the degree m. Nevertheless, we also assign the CHs a matching maximum expansion order N for the conversion between CHs and SHs and we use $\mathring{(\cdot)}$ to denote CH expansion coefficients. In contrast to the quadratically increasing $(N+1)^2$ coefficients in a set of SHs, the set of $2N+1$ CH coefficients only grows linearly with increasing order N. Thus, significantly fewer microphones are needed in EMA processing using CHs to create a uniquely or over-determined system of equations compared to SMA processing using SHs.

To consider microphone arrays under azimuthal rotation α, the real-valued rotation matrix $\boldsymbol{R}_N(\alpha)$ [23, Ch. 5.2.2] is introduced so that the expansion with $\boldsymbol{C}_N$ is performed with the rotated microphone angles,

$$\boldsymbol{C}_N \boldsymbol{R}_N(\alpha) \mathring{\boldsymbol{h}}_N(k) s(k) = \boldsymbol{p}(k)\,. \tag{4}$$

Note that (3) and (4) do not explicitly model the radial terms $\boldsymbol{b}_{\tilde{N}}(k,R)$ so that their influence is implicitly captured in the coefficients $\mathring{\boldsymbol{h}}_N(k)$. As shown in [1], [22], the CH coefficients $\mathring{\boldsymbol{h}}_N(k)$ can be converted to the SH coefficients of an order-limited horizontal projection of the original plane-wave composition $\boldsymbol{h}_{\tilde{N}}(k)$, including a compensation for the radial terms which is often referred to as radial filtering. The obtained SH coefficients of the horizontally-projected sound field then allow for rotations in three degrees of freedom (3DoF) and binaural rendering using established approaches such as the magnitude-least-squares method [24], [25].

III. SRIR Identification

We propose a frequency-domain recursive least squares (RLS) filter to adaptively estimate the CH representation $\mathring{\boldsymbol{h}}_N(k)$ of the room transfer function from recorded microphone signals $\boldsymbol{p}(k)$ and a reference signal. An RLS filter targeting static spherical arrays was formulated for SH domain signals in [26]. We assume arbitrary azimuthal rotations of the EMA and knowledge of the rotation angle α at any given time. In this contribution, we also assume knowledge of the reference signal, i.e., we directly employ the anechoic source signal $s(k)$ as reference. An estimate of the reference can, however, be obtained from the microphone signals by dereverberation and beamforming [4], [5]. The dependency on the wavenumber k is omitted for notational brevity in the following.

The RLS filter exploits (4) to find the LS-optimal CH coefficients $\mathring{\boldsymbol{w}}_N$ of the room transfer function (approximating $\mathring{\boldsymbol{h}}_N(k)$) by relating the measured sound pressures in $\boldsymbol{p}$ to the filtered source signal s while accounting for azimuthal array rotations α,

$$\min_{\mathring{\boldsymbol{w}}_N} \sum_{b=0}^{B-1} \lambda^{B-b} \left\| \boldsymbol{C}_N \boldsymbol{R}_N(\bar{\alpha}_b) \mathring{\boldsymbol{w}}_N s_b - \boldsymbol{p}_b \right\|^2 . \tag{5}$$

The filter minimizes the LS error in a block-processing framework of B signal blocks and the forgetting factor $\lambda \in (0, 1]$ may be used to give older signal blocks lower importance to adapt the filter in time-varying scenarios. As the array rotation angle may change within a signal block, the circular mean of the rotation angles within a block is employed as block rotation angle $\bar{\alpha}_b$. The LS-optimal estimate is derived by expanding (5) and setting its gradient with respect to $\mathring{\boldsymbol{w}}_N$ to zero, yielding

$$\hat{\mathring{\boldsymbol{w}}}_N = \left(\sum_{b=0}^{B-1} \lambda^{B-b} s_b^* s_b \boldsymbol{R}_N^\top(\bar{\alpha}_b) \boldsymbol{C}_N^\top \boldsymbol{C}_N \boldsymbol{R}_N(\bar{\alpha}_b) \right)^{-1} \times \sum_{b=0}^{B-1} \lambda^{B-b} s_b^* \boldsymbol{R}_N^\top(\bar{\alpha}_b) \boldsymbol{C}_N^\top \boldsymbol{p}_b \,, \tag{6}$$

where the superscript $(\cdot)^*$ denotes complex conjugation and $(\cdot)^\top$ denotes the transpose operation. The estimate at signal block b is thus recursively obtained as

$$\hat{\mathring{\boldsymbol{w}}}_{N,b} = \boldsymbol{R}_{\mathrm{ss},b}^{-1} \boldsymbol{r}_{\mathrm{sp},b} \,, \tag{7}$$

where

$$\boldsymbol{R}_{\mathrm{ss},b} = \lambda \boldsymbol{R}_{\mathrm{ss},b-1} + s_b^* s_b \boldsymbol{R}_N^\top(\bar{\alpha}_b) \boldsymbol{C}_N^\top \boldsymbol{C}_N \boldsymbol{R}_N(\bar{\alpha}_b) \,, \tag{8}$$

$$\boldsymbol{r}_{\mathrm{sp},b} = \lambda \boldsymbol{r}_{\mathrm{sp},b-1} + s_b^* \boldsymbol{R}_N^\top(\bar{\alpha}_b) \boldsymbol{C}_N^\top \boldsymbol{p}_b \,. \tag{9}$$

The time-domain SRIR estimate is then obtained from the frequency-domain room transfer function estimate $\hat{\mathring{\boldsymbol{w}}}_N$ using the inverse discrete Fourier transform.

IV. Simulation Study

A simulation study analyzing the normalized projection misalignment (NPM) [27] is performed to validate the proposed method and investigate its convergence behavior. The NPM is a scale-independent distance metric for the comparison of an impulse response estimate to its ground truth. As it considers all channels of the estimate, which in the present study represent the CH coefficients of the SRIR, it acts as a spatio-temporal error measure.

We employ a dataset of 10 simulated SRIRs of $200\,\mathrm{ms}$ length using an EMA of $4\,\mathrm{cm}$ radius with 60 equally spaced microphones to serve as ground truth and to facilitate the simulation of the array rotation. The simulated array allows for a maximum CH order of $\tilde{N} = 29$ and has an aliasing frequency of $f_\mathrm{A} = 39.6\,\mathrm{kHz}$ which is far beyond the upper frequency limit in this work given by the Nyquist frequency of $24\,\mathrm{kHz}$. The SRIRs were simulated using the image source method with the implementation from [28]. For each SRIR, a shoebox-shaped room with random dimensions between $4 \times 4 \times 2\,\mathrm{m}$ and $10 \times 8 \times 5\,\mathrm{m}$ drawn from a uniform distribution was generated. The omnidirectional source and the EMA were placed at random positions in the rooms while ensuring a minimum distance of $1\,\mathrm{m}$ to the walls and $2\,\mathrm{m}$ to each other. The absorption coefficient of the walls was generated randomly with values between 0.2 and 0.7, resulting in reverberation times between $239\,\mathrm{ms}$ and $476\,\mathrm{ms}$. The sampling frequency was set to $48\,\mathrm{kHz}$ in all simulations and measurements (Sec. V).

From the simulated SRIRs, array signals were obtained by convolution with $60\,\mathrm{s}$ of Gaussian white noise, and these signals were transformed to the CH domain using $\tilde{N} = 29$. An EMA rotating at a constant speed with four microphones at azimuth angles $-98\,°$, $-33\,°$, $33\,°$, and $98\,°$ was then simulated by evaluating the high-order CH signals at four time-variant microphone positions, i.e. by varying the microphone positions in $\boldsymbol{Y}_{\tilde{N}}$ in (1) over time. The microphone arrangement was chosen to correspond to the one in the measurement-based evaluation and is shown in Fig. 1. Although constant-speed rotations are not a requirement for the method, they are beneficial to investigate the influence of different rotation speeds; rotations with non-constant speed are tested in Sec. V.

The SRIR identification was performed using the pressure signals from the rotating array and the generated noise as the reference signal. The short-time Fourier transform (STFT) block processing used a block length of $400\,\mathrm{ms}$, a hop size of $100\,\mathrm{ms}$, and a square-root Hann window. The block length was set to twice the length of the ground truth SRIR to ensure a valid approximation of the convolutive transfer function as a multiplicative one. The obtained SRIRs were truncated to a length of $200\,\mathrm{ms}$ for comparison with the ground truth. The forgetting factor was set to $\lambda = 1$ as the transfer paths from the source to the rotating array surface were time-invariant.

The ground truth SRIR for a target CH order of N was obtained by truncating the simulated order-$\tilde{N}$ SRIR to the first $2N + 1$ channels. As the NPM only determines a relative misalignment with respect to the ground truth, modifications of the ground truth SRIR were added to the comparison to support the interpretation of the results. The first type of comparison was obtained by adding white noise to all channels of the ground truth SRIR to achieve a specific ratio of the direct-sound peak in the omnidirectional, zeroth-order CH coefficient, and the RMS of the noise floor. This ratio is referred to as the peak-to-noise ratio (PNR) in the following. The second type of comparison was obtained by simulating static EMAs with different numbers of equally spaced microphones. The corresponding SRIRs exhibit errors due to spatial aliasing for frequencies above their spatial aliasing frequency.

Fig. 2 shows the median NPM of the proposed method as solid lines and the median absolute deviation of the NPM as shadow surfaces for different rotation speeds and SRIR orders N. The different comparison NPMs are illustrated by horizontal lines with a text label denoting the PNR or the number of microphones $\mathcal{M}$ of the static comparison array. Although the SRIR identification with the proposed method only has access to the signals of four irregularly spaced microphones, see Fig. 1, it achieves NPMs that are comparable to measurements from static arrays with $\mathcal{M} = 19$ (for $N = 3$), $\mathcal{M} = 23$ ($N = 5$), and $\mathcal{M} = 25$ ($N = 7$) equally spaced microphones after 30 to $60\,\mathrm{s}$ of convergence. The NPMs of the proposed method are also comparable to the ground truth with a noise floor corresponding to a PNR of 50 to $55\,\mathrm{dB}$.

We further observe an influence of the rotation speed. Although a higher rotation speed leads to a faster initial convergence, the final NPM can be increased as seen in the

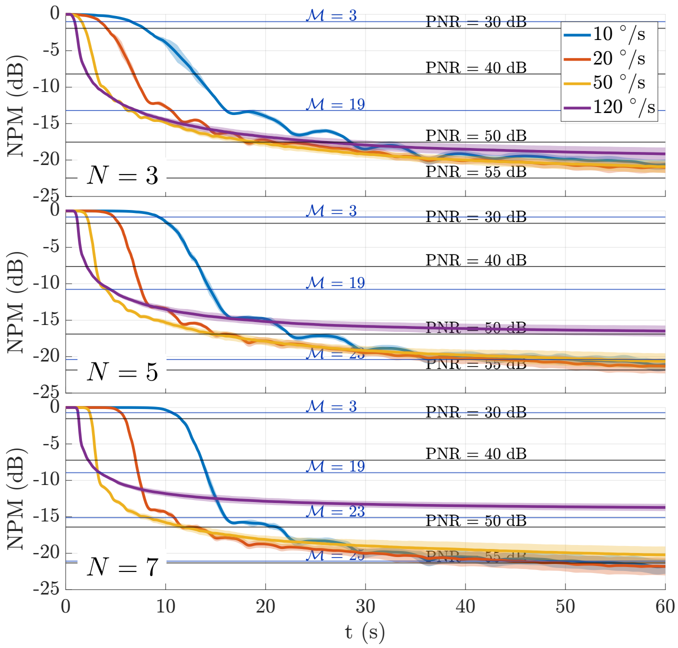

Fig. 2. Convergence behavior of the proposed method using simulations of different SRIR orders N and different rotation speeds. The horizontal lines show NPMs from modified ground-truth SRIRs with varying PNRs and varying numbers of microphones $\mathcal{M}$ for comparison.

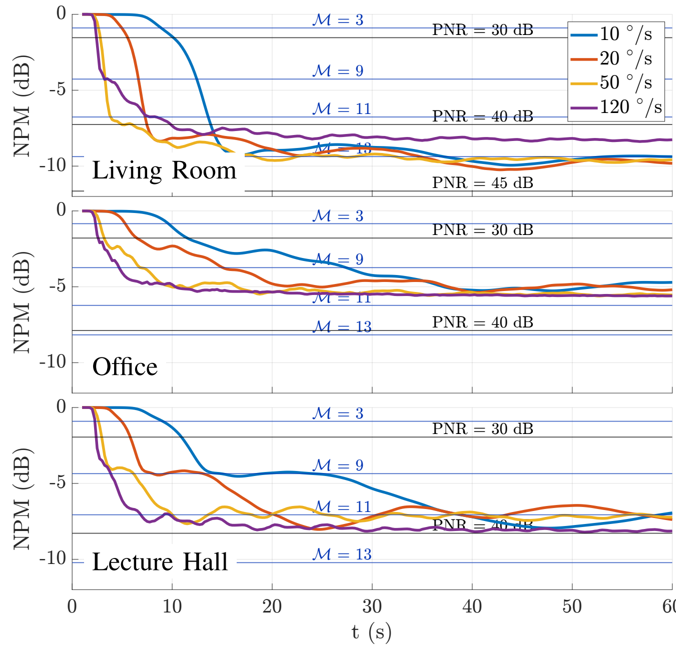

Fig. 3. Convergence behavior of the proposed method using measurements with the array from Fig. 1 in three different rooms, $N = 5$, and different maximum rotation speeds. The horizontal lines show NPMs from modified ground-truth SRIRs with varying PNRs and varying numbers of microphones $\mathcal{M}$ for comparison.

examples using the highest tested rotation speed of $120\,^\circ/\mathrm{s}$. The faster initial convergence is explained by the fact that the array covers a larger azimuthal range during a faster rotation. If insufficient directional information is available for a target order N due to slow rotation, the matrix $\boldsymbol{R}_{\mathrm{ss},b}$ is singular, resulting in a high NPM of $0\,\mathrm{dB}$ during the initial part of the identification. The increased final NPMs after fast rotation may be explained by the use of the mean rotation angle in each signal block. During faster rotations, the rotation angle changes considerably within each signal block, and approximation by the mean creates a larger error. This seems to be specifically important when high CH orders are estimated; these cannot be resolved accurately when the array is rotating quickly.

V. Measurements

To further test the proposed method under realistic conditions, we conducted measurements in 3 different rooms: a lab room furnished as a living room with dimensions of $4.8 \times 3.7 \times 2.5\,\mathrm{m}$ and a reverberation time of $270\,\mathrm{ms}$, an office of dimensions $6.4 \times 4.0 \times 2.3\,\mathrm{m}$ and a reverberation time of $300\,\mathrm{ms}$, and a lecture hall of dimensions $9.6 \times 9.7 \times 3.5\,\mathrm{m}$ and a reverberation time of $450\,\mathrm{ms}$. The utilized microphone array is shown in Fig. 1 and comprises 11 Røde Lavalier GO microphones on the equator of a wooden sphere of $6\,\mathrm{cm}$ radius. A ground truth SRIR was obtained by measuring SRIRs at two sequential array rotations to obtain a virtual array of 22 equidistant microphones, facilitating a maximum CH order of $\tilde{N} = 10$ and resulting in a spatial aliasing frequency of $f_{\mathrm{A}} = 9.1\,\mathrm{kHz}$. To obtain the ground truth, the corresponding CH-domain SRIRs were thus lowpass filtered at $9\,\mathrm{kHz}$, and all following investigations are limited to frequencies below that. For the SRIR identification, measurements with the array under rotation were performed. The array was thus placed on the turntable of the VariSphear measurement system [29] which supports rotations at a variable speed and captures the rotation trajectory. To investigate the method with variable-speed rotations, we used a cyclical rotation procedure, rotating the array from $0\,^\circ$ azimuth to $360\,^\circ$ and back repeatedly, which resulted in significant acceleration and deceleration phases during every turn. The maximum rotation speed was varied as before between $10\,^\circ/\mathrm{s}$ and $120\,^\circ/\mathrm{s}$. White Gaussian noise was used as the reference signal.

Fig. 3 shows the results for the three rooms and a maximum CH order of $N = 5$. The results of the proposed method with only four irregularly spaced microphones are comparable to measurements with static arrays of 11 or 13 equally spaced microphones and to modified ground truth RIRs with added noise corresponding to a PNR of around $40\,\mathrm{dB}$. The achieved NPMs are significantly higher than in the simulation study which may be due to the back-and-forth rotation scheme and due to inaccuracies in the synchronization of rotation data and audio signals, deviations from a perfectly rigid, spherical scattering body, and measurement noise. The proposed method however still offers a considerable benefit when compared to measurements from static EMAs with up to 11 microphones.

We observe the same slower initial convergence at slower rotation speeds as in the simulations but do not see a generally lower performance of the SRIR estimates from the fastest

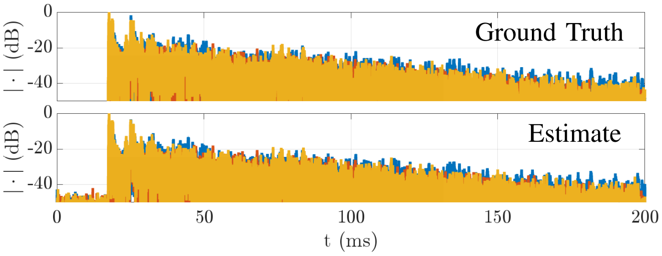

Fig. 4. Magnitude of the first three channels of the time-domain ground-truth SRIR (top) and the SRIR estimate (bottom) for the lecture hall.

rotation speed. This may be due to the mentioned measurement inaccuracies, limiting the overall achievable NPM.

Fig. 4 shows the magnitudes of the first three channels, i.e., the zeroth- and first-order coefficients, of the ground truth and of the identified SRIR of the lecture hall after $60\,\mathrm{s}$ of adaptation during cyclic rotation with a maximum speed of $50\,^{\circ}/\mathrm{s}$. The estimate is very similar to the ground truth regarding decay slope, early reflections, and the magnitude relations of the three channels but exhibits an increased noise floor with a PNR of $52\,\mathrm{dB}$.

VI. Conclusion

We proposed a method for the adaptive identification of SRIRs from rotating equatorial microphone arrays. The method not only compensates for but exploits the rotation to achieve a higher directional resolution than a corresponding static array. The method was tested with an array of four irregularly spaced microphones, different rotation speeds, and white noise, using simulations and measurements. The obtained SRIR estimates from the proposed method were found to be comparable to measured SRIRs from static arrays with a significantly greater number of equally spaced microphones. Future research should extend the method to support the blind identification of SRIRs from speech signals using head-worn arrays during natural head rotations to facilitate the virtual source rendering for augmented reality use cases.

References

[1] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. Amengual Garí, "Spherical harmonic decomposition of a sound field based on observations along the equator of a rigid spherical scatterer," *The Journal of the Acoustical Society of America*, vol. 150, no. 2, pp. 805–815, 2021.

[2] ——, "Spherical Harmonic Decomposition of a Sound Field Using Microphones on a Circumferential Contour Around a Non-Spherical Baffle," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 30, pp. 3110–3119, 2022.

[3] ——, "A Head-Mounted Microphone Array for Binaural Rendering," in *Int. Conf. on Immersive and 3D Audio*, 2021, pp. 1–7.

[4] T. Deppisch, J. Ahrens, S. V. Amengual Garí, and P. Calamia, "Blind Estimation of Spatial Room Impulse Responses Using a Pseudo Reference Signal," in *Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2024, p. 1–5.

[5] T. Deppisch, N. Meyer-Kahlen, and S. V. Amengual Garí, "Blind Identification of Binaural Room Impulse Responses from Smart Glasses," *preprint, arXiv:2403.19217*, 2024.

[6] V. Tourbabin, H. Barfuss, B. Rafaely, and W. Kellermann, "Enhanced robot audition by dynamic acoustic sensing in moving humanoids," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2015, pp. 5625–5629.

[7] J. Lawrence, J. Ahrens, and N. Peters, "Direction of Arrival Estimation Using the Rotating Equatorial Microphone," *Frontiers in Signal Processing*, vol. 4, p. 1–17, 2024.

[8] M. Zohourian, A. Archer-Boyd, and R. Martin, "Multi-channel speaker localization and separation using a model-based GSC and an inertial measurement unit," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2015, pp. 5615–5619.

[9] Y. Wakabayashi, K. Yamaoka, and N. Ono, "Sound Field Interpolation for Rotation-Invariant Multichannel Array Signal Processing," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 31, pp. 2286–2298, 2023.

[10] S. Hafezi, A. H. Moore, P. H. Guiraud, P. A. Naylor, J. Donley, V. Tourbabin, and T. Lunner, "Subspace Hybrid MVDR Beamforming for Augmented Hearing," *arXiv:2311.18689*, pp. 1–14, 2023.

[11] A. H. Moore, W. Xue, P. A. Naylor, and M. Brookes, "Noise covariance matrix estimation for rotating microphone arrays," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 27, no. 3, pp. 519–530, 2019.

[12] J. Casebeer, J. Donley, D. Wong, B. Xu, and A. Kumar, "NICE-Beam: Neural Integrated Covariance Estimators for Time-Varying Beamformers," *arXiv:2112.04613*, 2021.

[13] N.-C. Yen and W. Carey, "Application of synthetic-aperture processing to towed-array data," *Journal of the Acoustical Society of America*, vol. 86, no. 2, pp. 754–765, 1989.

[14] M. H. Johnson, "Synthetic elements for moving line arrays," in *OCEANS*, 2009, pp. 1–6.

[15] J. Ramirez, J. Odom, and J. Krolik, "Exploiting array motion for augmentation of co-prime arrays," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2014, pp. 525–528.

[16] C. Ming, H. Niu, Z. Li, and Y. Wang, "Passive synthetic aperture for direction-of-arrival estimation using sparse Bayesian learning," *The Journal of the Acoustical Society of America*, vol. 153, no. 4, p. 2061, 2023.

[17] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1636–1645, 2007.

[18] C. Antweiler, A. Telle, and P. Vary, "NLMS-Type System Identification of MISO Systems with Shifted Perfect Sequences," in *Int. Workshop on Acoustic Echo and Noise Control*, 2008, pp. 1–4.

[19] N. Hahn and S. Spors, "Continuous measurement of impulse responses on a circle using a uniformly moving microphone," in *23rd European Signal Processing Conference*, 2015, pp. 2581–2585.

[20] ——, "Continuous measurement of spatial room impulse responses using a non-uniformly moving microphone," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2017, pp. 205–208.

[21] B. Rafaely, *Fundamentals of Spherical Array Processing*, 2nd ed. Springer, 2019.

[22] J. Ahrens, "Ambisonic Encoding of Signals From Equatorial Microphone Arrays," Chalmers University of Technology, Tech. Rep., 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2211.00584

[23] F. Zotter and M. Frank, *Ambisonics, A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer, 2019.

[24] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Proc. of the German Annual Conference on Acoustics (DAGA)*, 2018, pp. 339–342.

[25] T. Deppisch, H. Helmholz, and J. Ahrens, "End-to-End Magnitude Least Squares Binaural Rendering of Spherical Microphone Array Signals," in *Int. Conf. on Immersive and 3D Audio*, 2021, pp. 1–8.

[26] N. Meyer-Kahlen and S. J. Schlecht, "Blind Directional Room Impulse Response Parameterization from Relative Transfer Functions," in *IEEE Int. Workshop on Acoustic Signal Enhancement*, 2022, p. 1–5.

[27] D. R. Morgan, J. Benesty, and M. Mohan Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 174–176, 1998.

[28] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: Algorithm and applications," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1462–1472, 2012.

[29] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio," in *Proc. of the German Annual Conference on Acoustics (DAGA)*, Berlin, 2010, pp. 717–718.