



Spatial Room Impulse Response Estimation from a Moving Microphone Array

Downloaded from: <https://research.chalmers.se>, 2026-09-14 17:59 UTC

Citation for the original published paper (version of record):

Deppisch, T., Amengual Gari, S., Calamia, P. et al (2025). Spatial Room Impulse Response Estimation from a Moving Microphone Array. European Signal Processing Conference: 91-95.
<http://dx.doi.org/10.23919/EUSIPCO63237.2025.11226557>

N.B. When citing this work, cite the original published paper.

Spatial Room Impulse Response Estimation from a Moving Microphone Array

Thomas Deppisch*, Sebastià V. Amengual Garí†, Paul Calamia† and Jens Ahrens*

*Chalmers University of Technology, 412 96 Gothenburg, Sweden

{thomas.deppisch, jens.ahrens}@chalmers.se

†Reality Labs Research, Meta, Redmond, WA 98052, USA

{samengual, pcalamia}@meta.com

Abstract—Estimating spatial room impulse responses (SRIRs) is essential for rendering virtual sound sources that realistically blend into a real-world environment in applications like augmented reality. Head-worn devices, such as smart glasses, are pivotal to enabling augmented reality experiences, but their continuous movement poses a unique challenge for the estimation of SRIRs. This study presents a method for estimating an SRIR representing the acoustic environment at a fixed reference point using signals recorded by a moving microphone array near that position. By leveraging the positional data of the microphone array, the proposed recursive least squares algorithm updates a signal model that describes the recorded sound pressure as a combination of the source signal and the SRIR at the reference point, transformed through rotation and translation in the circular harmonic domain. Simulation and measurement results demonstrate that the estimated SRIRs achieve high accuracy in the low and mid-frequency ranges – exceeding the spatial aliasing frequency of the array – while the translation distance and the order of the circular harmonic coefficients limit performance in the high-frequency range.

Index Terms—Augmented Reality, Microphone Array, Movement, Recursive Least Squares, Room Impulse Response

I. INTRODUCTION

Augmented reality (AR) and telepresence applications aim to enhance real-world acoustic scenes with virtual sound sources, typically rendered via headphones. For these virtual sources to appear realistic, they must blend seamlessly with the real acoustic environment by incorporating its spatio-temporal acoustic properties. In most end-user scenarios, room acoustic properties are not readily available and must be estimated from sensor data of head-worn devices like headsets or smart glasses, which typically include microphones and an inertial measurement unit (IMU) for positional tracking [1]. While the estimation of room acoustic properties and room impulse responses (RIRs) from static microphones is well studied [2]–[6], the unique challenges of continuously moving wearables have received limited attention. Exceptions include methods relying on specific excitation signals or movement trajectories to measure room impulse responses at multiple grid positions [7]–[9], and sound field estimation methods aiming to determine the sound pressure at arbitrary points within a region of interest [10]–[12].

This work instead aims to estimate a single spatial room impulse response (SRIR) – a multichannel RIR capturing

temporal and spatial acoustic properties – and builds on recent work utilizing a rotating microphone array [13]. The SRIR can also be interpreted as a set of sound field coefficients, and the proposed approach shares similarities with [12] in leveraging sound field translation theorems of spherical or circular harmonic expansions [14]–[18]. However, unlike [12], which uses a single moving microphone to estimate coefficients and reconstruct sound pressure at multiple positions, this work employs a moving microphone array to reconstruct coefficients at a reference point, aiming to enable accurate binaural rendering of the room acoustics.

The proposed method leverages position-tracking data from the array and assumes knowledge of the source signal. While the source signal is typically unavailable in many application scenarios, it may be estimated blindly from the captured array signals [6], [19]. Microphones in this study are assumed to be arranged on a circle along the equator of a spherical scatterer, which allows for an analytical description of the array transfer functions and provides well-controlled conditions for experiments while being extensible to head-worn arrays due to their similarity in configuration [20]. Limiting the aperture to a circle reduces the number of microphones needed for the same azimuthal resolution compared to a spherical configuration but inherently projects 3D sound fields onto the horizontal plane. Recent studies however suggest that renderings from such arrays offer similar spectral and perceptual properties as fully spherical arrays when sources are located close to the horizon [21], [22].

II. SIGNAL MODEL

Inside a solenoidal circular region free of sources, sinks and scatterers, any sound field created by the source $s(k)$ can be described by an infinite number of plane waves with the directional amplitude density $a(k, \varphi)$ [23],

$$p(k, r, \phi) = s(k) \frac{1}{2\pi} \int_0^{2\pi} a(k, \varphi) e^{ikr \cos(\varphi - \phi)} d\varphi, \quad (1)$$

where $p(k, r, \phi)$ is the sound pressure at wavenumber k , radius r and angle ϕ relative to the coordinate origin in the center of the circle, and i is the imaginary unit. The sound field can be equivalently represented using a circular harmonic (CH) expansion with complex exponential basis functions

$$C_m(\phi) = e^{im\phi}, \quad (2)$$

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

separating the radial and angular components while discretizing the sound field coefficients [23],

$$p(k, r, \phi) = s(k) \sum_{m=-\infty}^{\infty} b_m(kr) a_m(k) C_m(\phi). \quad (3)$$

Here, $b_m(kr) = i^m J_m(kr)$ describe the radial dependency, $J_m(kr)$ are the Bessel functions, and $a_m(k)$ are the expansion coefficients. Similar relations supporting 3D sound fields involve spherical harmonics (SHs) and are commonly employed in spherical array processing [14]. This work considers microphones on the equator of a rigid spherical scattering body of radius R , also referred to as equatorial microphone array (EMA). The sound field on the surface of the EMA is described by (3) after exchanging $b_m(kr)$ with

$$b_m^{\text{EMA}}(kR) = \sum_{n=|m|}^{\infty} (N_n^m(\pi/2))^2 b_n(kR). \quad (4)$$

Definitions of the normalizer $N_n^m(\pi/2)$ and the radial terms of a spherical array $b_n(kr)$ are given in [24], [25].

Similar to the SH transform on the sphere, any function on the circle can be described by a weighted sum of CH coefficients. The space-domain and CH-domain plane-wave densities are related via

$$a_m(k) = \frac{1}{2\pi} \int_0^{2\pi} a(k, \phi) C_m^*(\phi) d\phi, \quad (5)$$

$$a(k, \phi) = \sum_{m=-\infty}^{\infty} a_m(k) C_m(\phi), \quad (6)$$

where $(\cdot)^*$ denotes complex conjugation. As shown in (1), a translation in space by (r_t, ϕ_t) can be expressed by applying an angle-dependent phase shift to the plane-wave density,

$$a_t(k, \phi) = a(k, \phi) e^{ikr_t \cos(\phi - \phi_t)}. \quad (7)$$

The CH sound field coefficients with a translated coordinate origin thus are

$$\begin{aligned} a_{m,t}(k, r_t, \phi_t) &\stackrel{(5)}{=} \frac{1}{2\pi} \int_0^{2\pi} a_t(k, \phi) C_m^*(\phi) d\phi \\ &\stackrel{(6)(7)}{=} \frac{1}{2\pi} \int_0^{2\pi} \sum_{\mu=-\infty}^{\infty} a_{\mu}(k) C_{\mu}(\phi) e^{ikr_t \cos(\phi - \phi_t)} C_m^*(\phi) d\phi \\ &\stackrel{(9)}{=} \sum_{\mu=-\infty}^{\infty} a_{\mu}(k) C_{\mu-m}(\phi_t) \\ &\quad \times \frac{1}{2\pi} \int_0^{2\pi} e^{ikr_t \cos(\phi - \phi_t)} C_{\mu-m}(\phi - \phi_t) d\phi \\ &\stackrel{(10)}{=} \sum_{\mu=-\infty}^{\infty} a_{\mu}(k) C_{\mu-m}(\phi_t) i^{\mu-m} J_{\mu-m}(kr_t), \end{aligned} \quad (8)$$

where we exploited the relation

$$C_m(\phi) = C_m(\phi_t) C_m(\phi - \phi_t) \quad (9)$$

and inserted the Hansen-Bessel formula [26]

$$J_m(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ix \cos(\phi)} e^{im(\phi - \pi/2)} d\phi. \quad (10)$$

A similar derivation was performed in [17].

We obtain the sound pressure captured with an EMA at a translated position by inserting (8) and (4) into (3), and consider an array rotation of angle α by multiplication with $e^{im\alpha}$,

$$p_{r,t}(k, \phi) = s(k) \sum_{m=-\infty}^{\infty} a_{m,t}(k, r_t, \phi_t) b_m^{\text{EMA}}(kR) C_m(\phi) e^{im\alpha}. \quad (11)$$

After limiting the infinite sum in (11) to a maximum CH order $|m|, |\mu| \leq N$, the same can be expressed as a matrix-vector equation

$$\mathbf{p}(k) = \mathbf{C}(\phi_M) \mathbf{B}(k) \mathbf{R}(\alpha) \mathbf{T}(\phi_t, r_t) \mathbf{a}(k) s(k), \quad (12)$$

where $\mathbf{C}(\phi_M) \in \mathbb{C}^{M \times 2N+1}$ contains the CHs evaluated at the M microphone angles ϕ_M , $\mathbf{B}(k) \in \mathbb{C}^{2N+1 \times 2N+1}$ is a diagonal matrix containing the radial terms $b_m^{\text{EMA}}(kR)$, $\mathbf{R}(\alpha) \in \mathbb{C}^{2N+1 \times 2N+1}$ is a diagonal CH rotation matrix containing $e^{im\alpha}$ terms, $\mathbf{T}(\phi_t, r_t) \in \mathbb{C}^{2N+1 \times 2N+1}$ is the CH translation matrix described by (8), and $\mathbf{a}(k) \in \mathbb{C}^{2N+1}$ are the CH sound field coefficients in the coordinate origin which the proposed method aims to estimate.

III. SRIR ESTIMATION FROM A MOVING ARRAY

Room transfer functions can now be identified by minimizing the least squares (LS) error between the measured sound pressure and the sound pressure from estimated CH-domain coefficients according to (12). For notational convenience, the wavenumber and positional arguments are omitted below. An adaptive solution is obtained by applying block-wise processing of B blocks,

$$\min_{\mathbf{w}} \frac{1}{B} \sum_{b=0}^{B-1} \|\mathbf{C} \mathbf{B} \mathbf{R}_b \mathbf{T}_b \mathbf{w} s_b - \hat{\mathbf{p}}_b\|^2 + \delta \|\mathbf{w}\|^2, \quad (13)$$

where the filter $\mathbf{w}$ approximates $\mathbf{a}$, $\hat{\mathbf{p}}_b$ is the measured pressure in block b and δ is a small regularization weight. The translation distance r_t , angle ϕ_t , and the array rotation angle α are averaged within each signal block. The regularized LS-optimal solution,

$$\begin{aligned} \hat{\mathbf{w}} &= \left(\frac{1}{B} \sum_{b=0}^{B-1} s_b^* s_b \mathbf{T}_b^H \mathbf{R}_b^H \mathbf{B}^H \mathbf{C}^H \mathbf{C} \mathbf{B} \mathbf{R}_b \mathbf{T}_b + \delta \mathbf{I} \right)^{-1} \\ &\quad \times \frac{1}{B} \sum_{b=0}^{B-1} s_b^* \mathbf{T}_b^H \mathbf{R}_b^H \mathbf{B}^H \mathbf{C}^H \hat{\mathbf{p}}_b, \end{aligned} \quad (14)$$

can also be obtained recursively, yielding the recursive least squares (RLS) filter

$$\hat{\mathbf{w}}_b = (\mathbf{R}_{\text{ss},b} + \delta \mathbf{I})^{-1} \mathbf{r}_{\text{sp},b}, \quad (15)$$

where a cumulative average is used to estimate

$$\mathbf{R}_{\text{ss},b} = \frac{b-1}{b} \mathbf{R}_{\text{ss},b-1} + \frac{1}{b} s_b^* s_b \mathbf{T}_b^H \mathbf{R}_b^H \mathbf{B}^H \mathbf{C}^H \mathbf{C} \mathbf{B} \mathbf{R}_b \mathbf{T}_b, \quad (16)$$

$$\mathbf{r}_{\text{sp},b} = \frac{b-1}{b} \mathbf{r}_{\text{sp},b-1} + \frac{1}{b} s_b^* \mathbf{T}_b^H \mathbf{R}_b^H \mathbf{B}^H \mathbf{C}^H \hat{\mathbf{p}}_b. \quad (17)$$

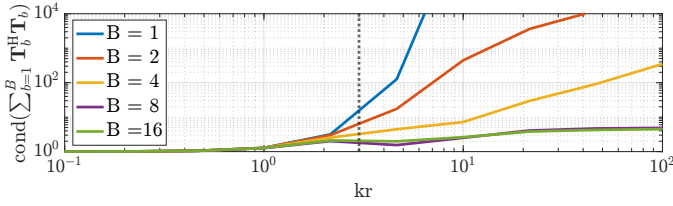


Fig. 1. The condition number of the translation matrix is reduced above $kr = N$ (dotted line) if sufficiently many observations B are available.

An exponential moving average replaces the cumulative average in time-variant scenarios. Estimation is needed only for positive frequencies, with negative-frequency coefficients derived via the conjugate symmetry property of CH coefficients [27]. The SRIR estimate is then obtained from $\hat{\mathbf{w}}$ using the inverse discrete Fourier transform.

IV. IMPACT OF TRANSLATION AND EXPANSION ORDER

The translation operation from (8) is based on the Bessel terms $J_{\mu-m}(kr)$ which only facilitate an error-free translation for $kr < |\mu-m|$ [17]. Zeroth-order, $m = 0$, terms are thus accurately reproduced for $kr < N$ while higher-order terms have lower kr limits. In the proposed algorithm, observations of the sound pressure at translated positions are used to estimate the sound field coefficients at the reference point, effectively inverting the translation operation. To determine the behavior of the translation operator in isolation, we neglect the other terms in the inverse in (14) and analyze the condition number of the matrix $\sum_{b=1}^B \mathbf{T}_b^H \mathbf{T}_b$ for $N = 3$ and different numbers of observations B in Fig. 1. With a single observation, condition numbers are high above $kr = N$ and gradually increase even below that, as expected from the properties of the Bessel terms. However, if more observations are available, their combination facilitates the estimation of the sound field coefficients even beyond $kr = N$ as seen from the lower condition numbers.

Fig. 2 demonstrates the approach for a point source and an EMA with M equally distributed microphones, supporting a CH decomposition of order $N_a = \lfloor (M-1)/2 \rfloor$. By combining observations from multiple array positions outside the reference point (the coordinate origin), the method produces a valid CH representation at the reference point (Fig. 2d), comparable to one derived directly from the array at the reference point (Fig. 2b). With sufficiently many observations, the CH order N of the estimated coefficients can be increased beyond the order N_a the microphone array natively supports (Fig. 2e).

V. SIMULATION STUDY

The following analyzes the RIR estimation performance by converting SRIR estimates to binaural room impulse responses (BRIRs) using the magnitude least squares method [28] and comparing them to the ground truth BRIR in third-octave bands using the normalized projection misalignment (NPM) [29]. This procedure facilitates the comparison of an estimate to a ground truth of higher CH order while focusing on errors that may be perceptually meaningful. For example,

inaccuracies of high-order coefficients at low frequencies only weakly impact the binaural NPM because of the weak directionality of head-related transfer functions (HRTFs) at low frequencies.

Array RIRs of 160 ms length were simulated for an array with 60 equally distributed microphones on the equator of a rigid sphere with 6 cm radius in 18 shoebox-shaped rooms using the image source method from [30]. The room dimensions were drawn from a uniform random distribution between $4 \times 4 \times 2$ m and $10 \times 8 \times 5$ m. Source and initial array positions were generated randomly while ensuring a minimum distance of 1 m to the walls and 1.5 m between source and array. This initial array position was used exclusively to compute reference SRIRs supporting an order of 12, which allows for an aliasing-free CH decomposition up to a frequency of 10 kHz. The reference SRIR was computed as the regularized least-squares solution for $\mathbf{a}_m$ of (3). Additionally, 20 array positions were simulated at uniformly distributed random positions within a circle of radius $\{0.2 \text{ m}, 0.4 \text{ m}, 1 \text{ m}\}$ around the initial position. The array RIRs at these positions were convolved with 2 s of white noise, and array rotations at $40^\circ/\text{s}$ were simulated using a high-order CH representation as in [13] to generate signals of a four-channel rotating array with nominal microphone angles of $\{-100^\circ, -30^\circ, 30^\circ, 100^\circ\}$, supporting a CH decomposition of $N_a = 1$. The signals of all positions were concatenated resulting in a length of 40 s. From the concatenated signals, SRIR estimation was performed using (15) with a block length of 320 ms, a hop size of 80 ms, 48 kHz sample rate, and $\delta = 10$ determined by a grid search.

Fig. 3 shows the binaural NPM in third-octave bands for various SRIR orders N . Solid lines indicate the results of the proposed method, while dashed lines show a baseline using the method from [13] where only the array rotation but not the translation is taken into account, i.e., $\mathbf{T}_b$ is the identity matrix. The proposed method consistently outperforms the baseline. Best performance occurs at low frequencies, with NPMs around -20 dB for translation distances of up to 0.2 m and 0.4 m, and slightly higher (-17 dB) for translations of up to 1 m. At mid frequencies, the method's benefits are most pronounced, with smaller translation distances and higher orders N reducing NPMs across a broader frequency range. High frequencies from around 4 kHz exhibit NPMs near 0 dB, reflecting the limitation of the method. Consistent with Fig. 2, combining observations at multiple translations and rotations allows the method to exceed limits defined by $kr = N_a$. Since the array translation is smaller than the maximum translation, $\max(r_t)$, for most observations, NPMs increase gradually rather than sharply for $kr > N$. NPMs may also increase for $kr < N$ due to the required regularization of the radial terms and the unequal contributions of CH coefficients to NPM at specific frequencies.

VI. MEASUREMENTS

While the simulation study investigated the method's performance for a large number of rooms where signals from

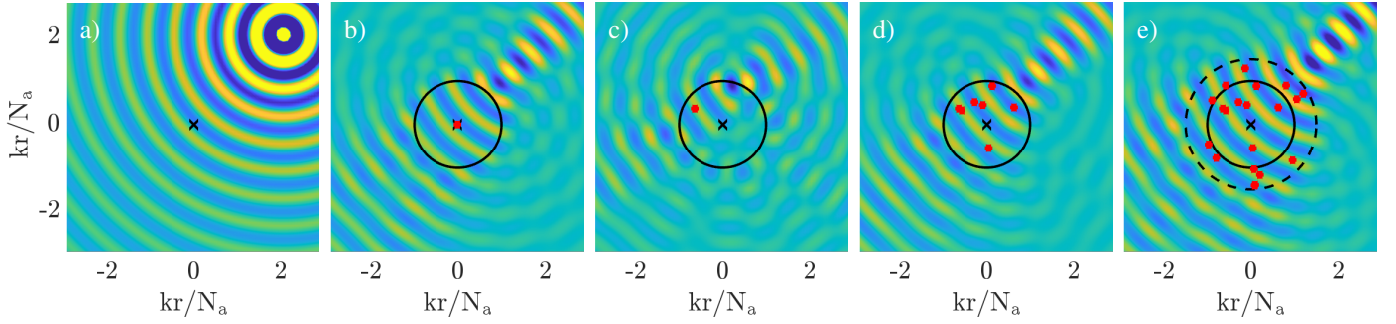


Fig. 2. The real part of the sound pressure from a point source in a) is accurately reproduced within the black circle marking $kr = N_a$ in b) from CH coefficients of order N_a from a microphone array (red dot) at the origin (black cross). c) CH coefficients from an off-origin array are translated to the origin via (8) but the resulting sound field is not accurate within $kr < N_a$. d) Combining multiple N_a th-order off-origin observations via (14), yields accurate origin coefficients within $kr < N_a$. e) With sufficient observations of order N_a , the order of the obtained CH coefficients may be increased to $N > N_a$, reproducing the sound field accurately within $kr < N$ (dashed circle).

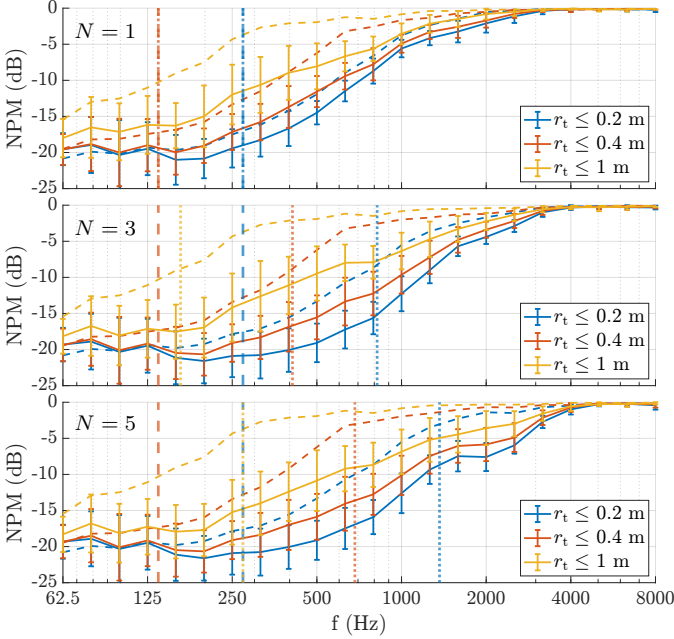


Fig. 3. Mean and standard deviation of the binaural NPM from SRIR estimates of different orders N of the proposed method (solid lines) and mean NPM of the baseline (dashed lines). Dashed vertical lines indicate $k \max(r_t) = N_a$, dotted vertical lines $k \max(r_t) = N$.

different static array positions were concatenated, the following considers measurements with a moving microphone array in three different rooms: a lab room furnished as a living room (reverberation time 300 ms, reference-position-to-source distance 2.8 m), an office (reverberation time 350 ms, distance 3.1 m), and a lecture room (reverberation time 430 ms, distance 3.3 m). In each room, a reference SRIR was measured using 22 microphones on the equator of a wooden sphere of 6 cm radius. The measurements were converted to the CH domain using order 10, yielding a spatial aliasing frequency of 9.1 kHz. Then, the microphone array was mounted on an Edlkrone Pan Pro turntable attached to an Edlkrone SliderPlus V5 Pro Long slider, enabling 360° rotation and translations of up to ± 45 cm from the reference point. In each

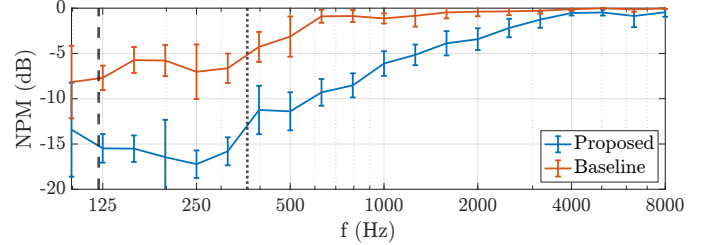


Fig. 4. Mean and standard deviation of binaural NPM based on measurements with a moving microphone array in three different rooms.

room, measurements were conducted with the array moving toward and away from the source as well as orthogonal to the source-array axis. The array moved within a ± 45 cm range at a speed of 4.3 cm/s while rotating within $\pm 90^\circ$ at $8.6^\circ/\text{s}$. During these movements, 60-second pink noise measurements were performed. Signals from four microphones at nominal angles $\{-98^\circ, -33^\circ, 33^\circ, 98^\circ\}$ were selected for the SRIR estimation, supporting $N_a = 1$. Block lengths were 400 ms (lab room), 500 ms (office), and 600 ms (lecture room), and the hop size was one quarter the block length. Regularization was set to $\delta = 1$. Both the reference and estimated SRIRs were converted to BRIRs as described in Sec. V, and binaural NPMs were calculated in third-octave bands.

Fig. 4 presents the mean and standard deviation of the NPMs for the method and the baseline in the three rooms, using $N = 3$. A dashed line indicates the frequency where $k \max(r_t) = N_a$, and a dotted line $k \max(r_t) = N$. Consistent with the simulations, the method benefits significantly from translation data, achieving NPMs around -15 dB at low frequencies and rising toward 0 dB near 4 kHz.

Fig. 5 compares the first 200 ms of the time-domain reference SRIR, the SRIR estimate, and the baseline for the lecture room and movement toward the source. The estimate from the proposed method closely matches the reference in terms of early reflections and energy decay, with only minor differences in CH coefficient magnitudes. In contrast, the baseline estimator generates less distinct direct sound and early reflections, and greater channel magnitude mismatches.

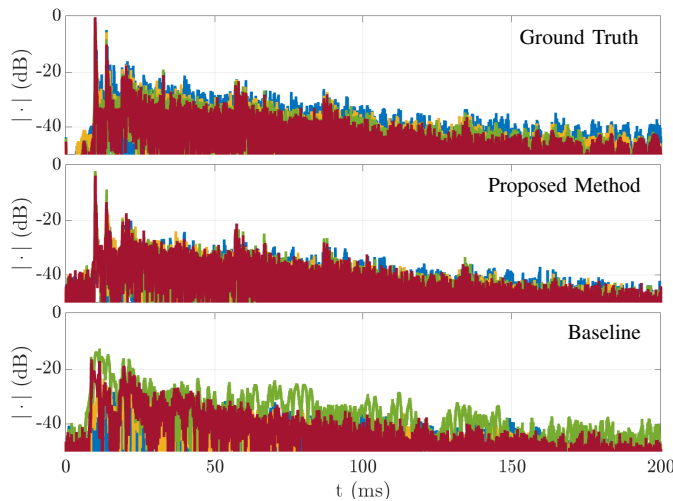


Fig. 5. Magnitude of the ground truth SRIR in the time domain (top), the estimate from the proposed method (center), and the baseline (bottom).

VII. CONCLUSION

This work introduced a novel method for estimating an SRIR at a fixed reference position from a moving microphone array. The method demonstrated strong performance at low and mid frequencies but is limited at high frequencies by the array's displacement from the reference point and the CH order. Notably, it has been shown that the CH order can be extended beyond the microphone's native capability if a sufficient number of observations are available. Future work should assess the perceptual impact of inaccuracies at high frequencies and test the method with head-worn arrays.

REFERENCES

- [1] K. Zmolikova, S. Merello, K. Kalgaonkar, J. Lin, N. Moritz, P. Ma, M. Sun, H. Chen, A. Saliou, S. Petridis, C. Fuegen, M. Mandel, and M. Ai, "The CHiME-8 MMCSG Challenge : Multi-modal conversations in smart glasses," in *8th International Workshop on Speech Processing in Everyday Environments (CHiME)*, 2024, pp. 7–12.
- [2] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of Room Acoustic Parameters: The ACE Challenge," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 24, no. 10, pp. 1681–1693, 2016.
- [3] C. J. Steinmetz, V. K. Ithapu, and P. Calamia, "Filtered Noise Shaping for Time Domain Room Impulse Response Estimation from Reverberant Speech," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2021, pp. 221–225.
- [4] A. Ratnarajah, I. Ananthabhotla, V. K. Ithapu, P. Hoffmann, D. Manocha, and P. Calamia, "Towards Improved Room Impulse Response Estimation for Speech Recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.
- [5] S. Lee, H. S. Choi, and K. Lee, "Yet Another Generative Model for Room Impulse Response Estimation," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2023, pp. 1–5.
- [6] T. Deppisch, N. Meyer-Kahlen, and S. V. Amengual Garí, "Blind Identification of Binaural Room Impulse Responses from Smart Glasses," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, p. 4052–4065, 2024.
- [7] 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.
- [8] N. Hahn and S. Spors, "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.
- [9] K. MacWilliam, T. Dietzen, R. Ali, and T. van Waterschoot, "State-space estimation of spatially dynamic room impulse responses using a room acoustic model-based prior," *Frontiers in Signal Processing*, vol. 4, pp. 1–18, 2024.
- [10] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "Sound-field measurement with moving microphones," *The Journal of the Acoustical Society of America*, vol. 141, no. 5, pp. 3220–3235, 2017.
- [11] F. Katzberg, M. Maass, and A. Mertins, "Spherical Harmonic Representation For Dynamic Sound-Field Measurements," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 426–430.
- [12] J. Brunnström, M. B. Møller, and M. Moonen, "Bayesian sound field estimation using moving microphones," *IEEE Open Journal of Signal Processing*, p. 1–10, 2025.
- [13] T. Deppisch, J. Ahrens, S. V. Amengual Garí, and P. Calamia, "Spatial Room Impulse Response Identification from Rotating Equatorial Microphone Arrays," in *32nd European Signal Processing Conference*, 2024, pp. 116–120.
- [14] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 1, pp. 2–16, 2010.
- [15] F. Schultz and S. Spors, "Data-based binaural synthesis including rotational and translatory head-movements," in *Proc. 52nd AES International Conference*, 2013, pp. 1–11.
- [16] P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 3, pp. 647–658, 2014.
- [17] N. Hahn and S. Spors, "Modal Bandwidth Reduction in Data-based Binaural Synthesis including Translatory Head-movements," in *German Annual Conference on Acoustics (DAGA)*, 2015.
- [18] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 135–139, 2018.
- [19] 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.
- [20] J. Ahrens, H. Helmholtz, D. L. Alon, and S. V. Amengual Garí, "A Head-Mounted Microphone Array for Binaural Rendering," in *Int. Conf. on Immersive and 3D Audio*, 2021, pp. 1–7.
- [21] H. Helmholtz, T. Deppisch, and J. Ahrens, "End-to-End Magnitude Least Squares Binaural Rendering for Equatorial Microphone Arrays," in *Proc. of the German Annual Conference on Acoustics (DAGA)*, 2023, pp. 1679–1682.
- [22] H. Helmholtz, J. Crukley, S. V. Amengual Garí, Z. Ben-Hur, and J. Ahrens, "Quality of Binaural Rendering From Baffled Microphone Arrays Evaluated Without an Explicit Reference," *Journal of the Audio Engineering Society*, pp. 691–704, 2024.
- [23] M. Poletti, "A Unified Theory of Horizontal Holographic Sound Systems," *J. Audio Eng. Soc.*, vol. 48, no. 12, p. 1155–1182, 2000.
- [24] J. Ahrens, H. Helmholtz, 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.
- [25] 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>
- [26] S. Iyanaga and Y. Kawada, *Encyclopedic Dictionary of Mathematics*. Cambridge, MA: MIT Press, 1980.
- [27] T. Deppisch and J. Ahrens, "Symmetries in Complex-Valued Spherical Harmonic Processing of Real-Valued Signals," in *Proc. of the Annual German Conference on Acoustics (DAGA)*, 2025.
- [28] 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.
- [29] 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.
- [30] 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.