AMBISONIC MULTI-DIRECTION DECOMPOSITION METHOD FOR RESOLVING COINCIDENT ACOUSTIC REFLECTIONS

Thomas Röck[1], Franz Zotter[1]

[1]*Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria*

Abstract

Not all reflections in measured Ambisonic room impulse responses (ARIRs) arrive in distinct time intervals. In principle, compact spherical microphone arrays recording higher-order Ambisonics enable decomposing directions of coincident reflections. However, existing approaches for decomposing and upmixing higher-order ARIRs resolve fewer coincident directions than the theoretical $(N+1)^2$ components contained in an order-N ARIR. Moreover, decomposition becomes ill-conditioned whenever directions are not well-separated. This contribution employs the Ambisonic Multi-Direction Decomposition Method (AMDDM) to address these limitations in higher-order ARIR decomposition: $(N+1)^2$ well-separated directional signals are extracted by adaptively rotating a fixed maximum-determinant grid. Two loss functions defining an optimal rotation are compared to find the one most applicable to ARIR analysis. In a numerical study with simulated ARIRs containing multiple reflections coinciding within a single time frame, we compare AMDDM with two other, established state-of-the-art methods. In the tested conditions, the rigid directional grid optimally rotated using a sparsity loss function outperformed the state-of-the-art methods.

Keywords: *Room Impulse Response, Reflection Analysis, Higher-Order Microphone Arrays*

1. Introduction

The directional characteristics of a real acoustic space can be captured with an Ambisonic room impulse response (ARIR). Convolution of that ARIR with an arbitrary source signal is the basis for rendering virtual acoustic scenes in augmented- and virtual-reality applications. The underlying Ambisonics format [1] allows for playback and rotation of these scenes on any loudspeaker or headphone setup.

Tetrahedral microphone arrays are a popular candidate for recording ARIRs but can only record first-order Ambisonics with limited directional resolution. Therefore, a variety of methods aim to enhance directional resolution by estimating spatial parameters and re-encoding the low-order ARIR into higher orders [2], [3], [4], [5], [6]. These methods rely on the detection of a single direction of arrival (DoA) per time-frame.

Higher-order compact spherical microphone arrays (CSMAs) [7] can resolve multiple simultaneous reflections and detect their DoAs. Methods have been proposed to analyze the higher-order signals they capture [8], [9]. Higher-order spatial impulse response rendering (HO-SIRR) [8] first segregates the sound field into individual sectors, then estimates a DoA and diffuseness for every sector signal.

Another method, Reproduction and Parametrization of Array Impulse Responses (REPAIR) [9], first determines the number of active sources via subspace analysis. The active source subspace is decomposed into simultaneous DoAs and signals, and the residual subspace is processed as a diffuse field.

Neither method has addressed that CSMAs do not reach a uniform directional resolution across frequency.

Recently, the Ambisonic Multi-Direction Decomposition Method (AMDDM) [10] has been proposed for upmixing Ambisonic signals of input order N to a higher target order $\tilde{N}$.

AMDDM operates by extracting signals from directions defined by a rotated grid of well-separated directions. It then re-encodes these directional signals at the rotated directions into the higher target order. At the optimal rotation, a cross-covariance based loss function reaches its minimum and promotes uncorrelated directional signals. It is unclear whether this loss function applies well to ARIR analysis.

After discussing the frequency-dependent resolution of CSMAs in Section 2 and briefly revisiting the

Corresponding author: *roeck@iem.at*

AMDDM method in Section 3, this paper introduces an alternative, sparsity-based loss function suited to ARIR analysis. In a simulation study in Section 4, AMDDM is compared with HO-SIRR and REPAIR. For all methods, CSMA signals are processed considering the achievable directional resolution in every frequency band. Section 5 identifies the better-suited loss function and quantifies the spatial accuracy of all three methods on simulated single-frame scenarios and on a full room ARIR, and Section 6 discusses where a fixed, rotated grid gains over a parametric directional/diffuse split and where it reaches its limits.

2. CSMA order is frequency-dependent

CSMAs capture sounds from all directions using M microphones mounted on positions $a\,\boldsymbol{\theta}_l$ on a rigid sphere of radius a. The microphone signals $\boldsymbol{p}$ correspond to sound pressure and relate to the spherical harmonic (SH) coefficients $\boldsymbol{\Psi}$

$$\boldsymbol{p} = \hat{\boldsymbol{Y}}_N \boldsymbol{\Psi} \tag{1}$$

where $\hat{\boldsymbol{Y}}_N$ contains the sampled SHs at the microphone directions $\boldsymbol{\theta}_l$. Minimizing $\|\boldsymbol{p} - \hat{\boldsymbol{Y}}_N \boldsymbol{\Psi}\|^2$ yields the encoding matrix to obtain the coefficients

$$\boldsymbol{\Psi} = \left(\hat{\boldsymbol{Y}}_N^{\mathrm{T}} \hat{\boldsymbol{Y}}_N\right)^{-1} \hat{\boldsymbol{Y}}_N^{\mathrm{T}}\, \boldsymbol{p}. \tag{2}$$

The model equation relates each component Ψ_{nm} in the vector $\boldsymbol{\Psi}$ to a corresponding Ambisonic signal via radial filters $\rho_n(\omega)$ [1, p. 139]

$$\mathcal{X}_{nm} = \underbrace{\frac{(\Bbbk a)^2 h_n'^{(2)}(\Bbbk a)}{2\pi \mathrm{i}^{n+1}}}_{\rho_n(\omega)} \Psi_{nm} \tag{3}$$

with imaginary unit $\mathrm{i} = \sqrt{-1}$, wave number $\Bbbk = \frac{2\pi f}{c}$ with the frequency f and speed of sound c, and the derivative of the spherical Hankel function of the second kind $h_n'^{(2)}$ [11, Eq. 10.47.6]. According to Eq. (3), the so-called radial filters that equalize high-order coefficients require strong gain boosts at low frequencies. Therefore, radial filters need to be regularized, which can be implemented by cutting on higher orders at certain frequencies [12].

We implement the regularized Eq. (3) as a linear-phase crossover filterbank of $N+1$ bands. Each of the bands $b = 0, ..., N$ carries only the orders $n \leq b$. The crossover frequencies enabling the order n were adapted empirically from the rule in [13, Eq. (13)], such that the maximum gain of any radial filter stays within ± 1.5 dB of a maximum boost $g_{\max}$, for $n > 0$,

$$\Bbbk a_n = (2n-1)/2 \cdot \sqrt[n]{2/g_{\max}}. \tag{4}$$

Figure 1 shows the $0 \leq n \leq 6$, $g_{\max} = 20$, encoding filters for a CSMA of $a = 4.2$ cm, cut on by $(n+1)^{\text{th}}$ order Butterworth high-pass filters at $f_n \approx 130$ Hz, 900 Hz, 2 kHz, 3.2 kHz, 4.4 kHz, 5.5 kHz, for $n > 0$.

The directional resolution of a CSMA recording is hence inherently band-limited: within band b, the ARIR is an Ambisonic signal of order b rather than the maximum-achievable order N. Instead of analyzing the broadband signal, every band b is processed at its native order, upmixed to the target order $\tilde{N}$, before summing the upmixed bands. As the omnidirectional band $b = 0$ permits no direction analysis, it is folded into band 1 by summing the radial filters of both bands. In the evaluation, HO-SIRR and REPAIR receive the same per-band inputs, to permit a meaningful comparison.

3. AMDDM

Instead of estimating a time-varying number of sources and their corresponding directions, AMDDM extracts directional signals $\boldsymbol{s}(t)$ using a fixed grid of well-separated directions (cf. Figure 2)

$$\boldsymbol{s}(t) = \boldsymbol{Y}_N^{-\mathrm{T}} \boldsymbol{R}_N(\varrho_1, \varrho_2, \varrho_3) \boldsymbol{\mathcal{X}}_N(t). \tag{5}$$

The extraction matrix $\boldsymbol{Y}_N^{-\mathrm{T}}$ inverts the real spherical harmonics (SHs) sampled at maximum determinant (MD) grid [14] points, transposed. The number of grid points matches the number of SH signals, thus the MD optimization guarantees a directly invertible, well-conditioned $\boldsymbol{Y}_N$. The Ambisonic signals $\boldsymbol{\mathcal{X}}(t)$ are rotated by an N-th order Wigner-D rotation matrix $\boldsymbol{R}_N$ around the Euler angles $\varrho_1, \varrho_2, \varrho_3$. AMDDM originally evaluated the SHs at MD grid points rotated by a 3×3 Cartesian rotation matrix and approximated the (higher-order) derivatives required for finding the optimal rotation numerically. Defining the rotation as a recombination of the SH signals by the N-th order Wigner-D rotation matrix $\boldsymbol{R}_N$ not only is computationally more efficient, but further allows for analytic differentiation, as one improvement introduced here. The recursion of Ivanic and Ruedenberg [15] assembles $\boldsymbol{R}_N$ order by order from the 3×3 rotation matrix $\boldsymbol{R}$, as detailed in Section 3.1.

3.1. Optimal Rotation (Off-diagonal loss)

AMDDM considers the directional signal covariance matrix to define optimal rotation by $\varrho_1, \varrho_2, \varrho_3$, attempting to find uncorrelated or independent signals $\boldsymbol{s}$. The expectation is estimated as an average over short time frames, so one optimal rotation is found per frame. Given Eq. (5) and omitting $\varrho_1, \varrho_2, \varrho_3$ for brevity, the directional-signal covariance matrix is

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{s}} = E\{\boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}\} &= \boldsymbol{Y}_N^{-\mathrm{T}} \boldsymbol{R}_N \underbrace{E\{\boldsymbol{\mathcal{X}}_N \boldsymbol{\mathcal{X}}_N^{\mathrm{T}}\}} \boldsymbol{R}_N^{\mathrm{T}} \boldsymbol{Y}_N^{-1} \\ &= \boldsymbol{Y}_N^{-\mathrm{T}} \boldsymbol{R}_N \boldsymbol{C}_{\boldsymbol{\mathcal{X}}} \boldsymbol{R}_N^{\mathrm{T}} \boldsymbol{Y}_N^{-1}. \end{aligned} \tag{6}$$

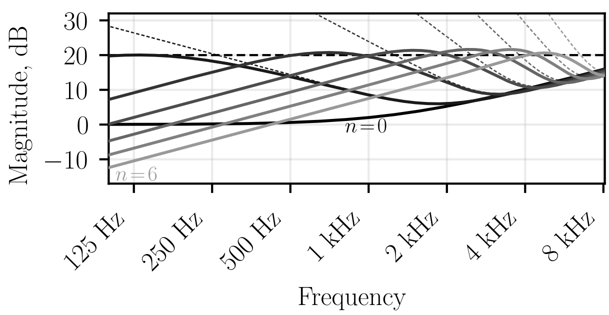

Figure 1: Radial filters of the order-6 CSMA encoding filterbank ($a = 4.2$ cm, $g_{\max} = 20$ dB maximum gain), from $n = 0$ (black) to $n = 6$ (light).

Non-vanishing off-diagonal entries in this matrix indicate correlation between the directional signals extracted. A loss function enforcing independent signals must therefore penalize off-diagonal entries

$$\mathcal{L}_{\mathrm{Cov}}(\boldsymbol{C_s}) = \|\boldsymbol{C_s} - \boldsymbol{I} \odot \boldsymbol{C_s}\|^2_{\mathrm{Fro}} \tag{7}$$

where $\|\cdot\|_{\mathrm{Fro}}$ denotes the Frobenius matrix norm and element-wise (Hadamard) multiplication of $\boldsymbol{C_s}$ with the identity matrix $\boldsymbol{I}$ removes the main diagonal.
Optimal rotation parameters $\boldsymbol{\varrho} = (\varrho_1, \varrho_2, \varrho_3)^{\mathrm{T}}$ minimize the loss

$$\underset{\boldsymbol{\varrho}}{\operatorname{argmin}} \quad \mathcal{L}_{\mathrm{Cov}}(\boldsymbol{C_s}) \tag{8}$$

and can be found iteratively using Newton's method $\boldsymbol{\varrho} \leftarrow \boldsymbol{\varrho} - \mu \boldsymbol{H}^{-1}\boldsymbol{g}$ with the step size μ, the 3×1 gradient $\boldsymbol{g}$ and the 3×3 Hessian $\boldsymbol{H}$ defined as

$$\boldsymbol{g} = \left[\frac{\partial}{\partial \varrho_k}\right]_k \mathcal{L}, \quad \text{and } \boldsymbol{H} = \left[\frac{\partial^2}{\partial \varrho_k \partial \varrho_l}\right]_{kl} \mathcal{L}, \tag{9}$$

with partial derivatives with regard to the k^{th} or l^{th} rotation parameter in $\boldsymbol{\varrho}$, and indexed square brackets denoting an operator shaping single- or double-indexed terms such as our differentials into either a vector $[a_k]_k =: \boldsymbol{a}$ or matrix $[a_{kl}]_{kl} =: \boldsymbol{A}$, respectively. Before inversion, the Hessian is regularized as $\boldsymbol{H} \leftarrow \boldsymbol{H} + \lambda \boldsymbol{I}$ with $\lambda = 10^{-3}$, which keeps the step well-defined wherever $\boldsymbol{H}$ becomes near-singular. Each iteration picks the μ with the smallest loss from the discrete line-search set $\mu \in \{0, \pm 0.01, \pm 0.1, \pm 0.3, \pm 1, \pm 3, \pm 10\}$ instead of a single Newton step. This avoids too short, too long, or reverse steps whenever the implied quadratically approximated stationary point is insufficient or indicates a maximum instead of a minimum.
In $\boldsymbol{C_s}$ Eq. (6), only the SH rotation matrix $\boldsymbol{R}_N$ depends on the rotation parameters. According to [15], it can be evaluated recursively based on the 3×3 rotation matrix for the Euler angles $\varrho_1, \varrho_2, \varrho_3$,

$$\boldsymbol{R}(\varrho_1, \varrho_2, \varrho_3) = \boldsymbol{R}_Z(\varrho_1)\boldsymbol{R}_Y(\varrho_2)\boldsymbol{R}_Z(\varrho_3) \tag{10}$$

constructed from the rotation matrices

$$\boldsymbol{R}_Z = \begin{bmatrix} \cos & 0 & \sin \\ 0 & 1 & 0 \\ -\sin & 0 & \cos \end{bmatrix}, \ \boldsymbol{R}_Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos & -\sin \\ 0 & \sin & \cos \end{bmatrix}. \tag{11}$$

Each entry of $\boldsymbol{R}_N$ is a weighted sum of products of one entry of $\boldsymbol{R}$ and one entry of $\boldsymbol{R}_{N-1}$ [15, Table 2], so the recursion is differentiable in closed form. The gradient and Hessian of $\mathcal{L}_{\mathrm{Cov}}$ are therefore available analytically: the derivatives of $\boldsymbol{R}_N$ propagate through the covariance Eq. (6) and the loss Eq. (7) by the chain and product rules. The individual elementary steps are given in Appendix A.

3.2. Kurtosis Loss

$\mathcal{L}_{\mathrm{Cov}}$ penalizes correlation between the extracted signals. In an ARIR, however, coincident reflections are delayed and attenuated copies of the direct sound and might be correlated whenever their delay difference is small compared to the analysis window. We instead propose promoting the joint spatio-temporal sparsity of the extracted frame. Collecting the T frame samples in $\boldsymbol{S} = [\boldsymbol{s}(t_1), ..., \boldsymbol{s}(t_T)]$, cf. $\boldsymbol{C_s} \propto \boldsymbol{S}\boldsymbol{S}^{\mathrm{T}}$, the negative-kurtosis loss

$$\mathcal{L}_{\mathrm{Kurt}}(\boldsymbol{S}) = -\frac{\sum_{i,t} S_{it}^4}{\left(\sum_{i,t} S_{it}^2\right)^2} = -\frac{\|\boldsymbol{S}\|_4^4}{\|\boldsymbol{S}\|_{\mathrm{Fro}}^4} \tag{12}$$

is minimized when the frame energy concentrates in few direction-time cells, i.e. when only a few grid directions carry pronounced signals with pronounced temporal peaks. The loss is smooth, scale-invariant, and indifferent to correlation between the extracted signals. As a smooth rational function of the extracted samples, it inherits its dependence on $\boldsymbol{\varrho}$ solely through $\boldsymbol{R}_N$, so the very same differentiated recursion serves both criteria and the rotation search proceeds unchanged. Its gradient and Hessian are derived in Appendix B.
All analytic derivatives were verified against automatic differentiation. Further sparsity criteria based on per-direction crest factors were investigated as well, but performed consistently worse and are not considered further.

3.3. Upmix Synthesis

Given the optimal rotation of a single frame, the extracted signals are re-encoded into the target order as plane waves arriving from the rotated grid directions,

$$\tilde{\boldsymbol{\mathcal{X}}}_{\tilde{N}}(t) = \boldsymbol{R}_{\tilde{N}}^{\mathrm{T}} \boldsymbol{Y}_{\tilde{N}}^{\mathrm{T}} \boldsymbol{s}(t), \tag{13}$$

where $\boldsymbol{Y}_{\tilde{N}}$ contains the order-$\tilde{N}$ SHs at the unrotated MD grid and $\boldsymbol{R}_{\tilde{N}}$ is the order-$\tilde{N}$ rotation matrix of the optimal $\boldsymbol{\varrho}$, whose top-left block is $\boldsymbol{R}_N$. This formulation avoids inverting a rotated SH matrix and stays well-conditioned for any rotation. The overlap-added frames of all bands sum to the upmixed ARIR.

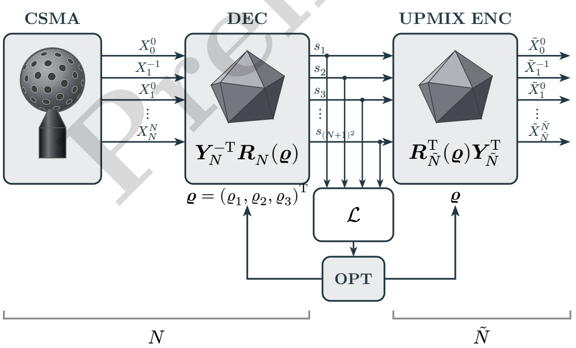

Figure 2: The Ambisonic Multi-Direction Decomposition Method (AMDDM) decomposes the input onto an optimally rotated maximum-determinant grid. A loss function $\mathcal{L}$ evaluated on the extracted signals steers the rotation and the signals are re-encoded into the target order $\tilde{N}$ at the rotated grid directions.

4. Simulation Study

All experiments simulate an Eigenmike em64 [16] with $M = 64$ capsules on a rigid sphere of radius $a = 4.2$ cm. The capsule signals are encoded into an order $N = 6$ ARIR using the filterbank of Section 2 with a maximum gain $g_{\max} = 20$ dB and linear-phase filters of 1024 taps, and every method upmixes this common input to the target order $\tilde{N} = 7$, i.e. $(\tilde{N} + 1)^2 = 64$ channels. Every output is compared to the ideal order-7 plane wave encoding of the same sound field, free of scattering and aliasing. All signals are sampled at 48 kHz, and all comparisons are restricted to below the aliasing frequency $f_{\text{alias}} \approx 8$ kHz.

4.1. Simulated Array Signals

Sound fields are synthesized as plane waves in the SH domain at order $N_{\text{sim}} = 30$. The coefficients pass through the rigid-sphere transfer function, i.e. the inverse of Eq. (3), and are sampled at the capsule directions.

A full room ARIR is simulated with a hybrid model [17] of a shoebox room ($8 \times 6 \times 5$ m, $T_{60} = 0.5$ s, source-array distance 4.5 m): the image source method [18] up to order 10 provides the direct sound and early reflections as discrete plane waves, and the late part is an isotropic diffuse field synthesized with one plane wave of random direction per sample whose energy envelope follows a stochastic ray-tracing simulation of the same room. Both parts are joined by an equal-power crossfade at 100 ms.

Additionally, three single-frame scenarios with known DoAs are simulated to isolate directional questions:

1) Single source
2) Two sources 40° apart
3) 12 arrivals from random directions with levels within 9 dB and individual delays below 1 ms and at least 12° apart from each other.

4.2. Compared methods

HO-SIRR and REPAIR are run in their reference implementations[1] for CSMAs at their native analysis resolutions, per band as described in Section 2. Both render to a loudspeaker layout, here a spherical t-design of degree 14 (114 directions) that is dense enough to carry an order-7 field. REPAIR estimates its DoAs with MUSIC on a t−60 design directional grid. AMDDM processes Hann-windowed frames of 50 samples with 50% overlap per band.

4.3. Metrics

The main figure of merit is the spatial correlation coefficient

$$c = \frac{\sum_{f \in \mathcal{F}} \boldsymbol{\mathcal{X}}_{\text{ref}}^{\text{H}}(f)\, \boldsymbol{\mathcal{X}}(f)}{\sqrt{\left(\sum_{f \in \mathcal{F}} \|\boldsymbol{\mathcal{X}}_{\text{ref}}(f)\|^2\right)\left(\sum_{f \in \mathcal{F}} \|\boldsymbol{\mathcal{X}}(f)\|^2\right)}}, \quad (14)$$

which correlates the vectors $\boldsymbol{\mathcal{X}}(f)$ of SH-channel spectra of method and reference over all DFT frequencies $\mathcal{F}$ below f_{alias}. Its magnitude reaches $|c| = 1$ if and only if the two ARIRs agree up to a single global complex factor, so $|c|$ is insensitive to an overall gain but penalizes any deviation in the directional composition.

[1] HO-SIRR: https://github.com/leomccormack/HO-SIRR
REPAIR: https://github.com/leomccormack/REPAIR

Table 1: Spatial correlation coefficient $|c|$ of the AMDDM output per loss criterion, for the simulated room ARIR and the three single-frame scenarios of Section 4 (best per row in bold).

Case	$\mathcal{L}_{\text{Cov}}$	$\mathcal{L}_{\text{Kurt}}$
Simulated ARIR	0.813	**0.833**
Single source	0.994	**0.995**
Two sources	0.902	**0.931**
Twelve random sources	0.663	**0.687**

5. Results

5.1. Loss Criteria

Table 1 compares the off-diagonal covariance loss with the proposed kurtosis loss, evaluated in terms of the magnitude of the spatial correlation coefficient Eq. (14) for the AMDDM output. For the multi-source scenarios, the two simultaneous sources and the twelve source scene, the kurtosis loss outperforms the off-diagonal loss. This supports the reasoning presented in Section 3.2, and the kurtosis loss is therefore adopted for all subsequent results.

5.2. Single-frame Scenarios

Figure 3 depicts broadband directional energy maps of the three scenarios, rendered by a max-r_{E}-weighted plane-wave decomposition of each order-7 output for visualization with reduced sidelobes.

The single source is localized correctly by all methods. AMDDM reproduces the ideal field almost exactly ($|c| = 0.99$), while HO-SIRR and REPAIR render a visibly broader spot ($|c| = 0.93$ and 0.86 respectively). The two sources 40° apart are resolved by all methods, with AMDDM staying closest to the ideal field ($|c| = 0.95$ vs. 0.90 and 0.91). The 12-source scene degrades every method: HO-SIRR and REPAIR blur most arrivals into a broad energy distribution ($|c| = 0.49$ and 0.60), whereas AMDDM still reproduces the strongest arrivals distinctly ($|c| = 0.72$).

5.3. Full ARIR

For the simulated room ARIR, Figure 4 resolves the spatial correlation coefficient over frequency. Above 300 Hz, AMDDM clearly exceeds both reference methods and reaches $|c(f)| \approx 0.95$ around 4 kHz. Towards lower frequencies, all methods drop to the modest coherence the low-order bands afford. Below 700 Hz, REPAIR performs somewhat better than HO-SIRR, while above that both reference methods achieve an almost identical spatial correlation value. It should be stressed, however, that $|c|$ measures wavefront reconstruction and not perceptual quality. The parametric diffuse streams are decorrelated by design and thus score low in it.

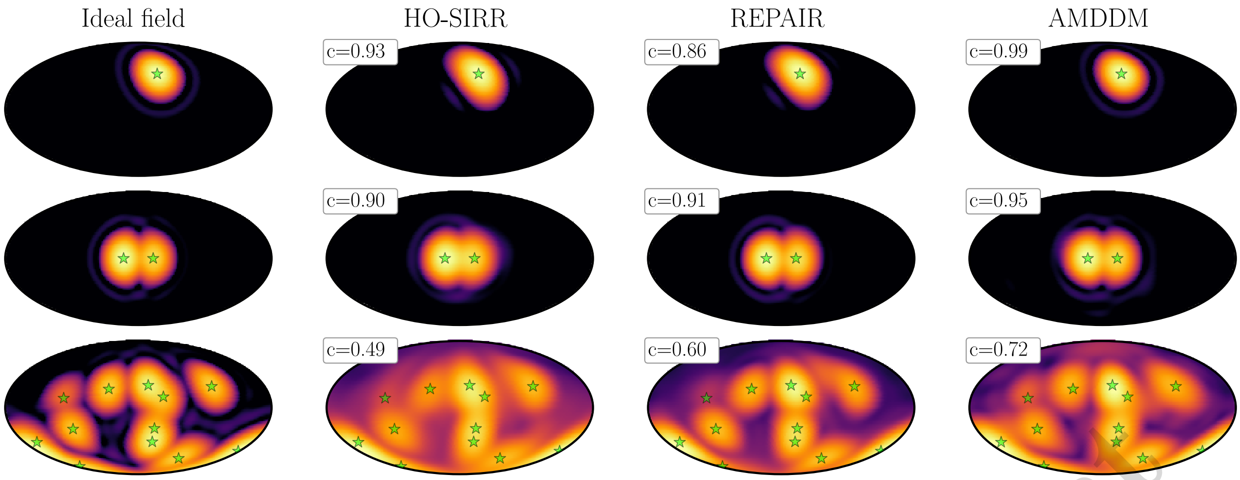

Figure 3: Comparison of HO-SIRR, REPAIR, and AMDDM to the ideal order-7 reference for different numbers of sources. First row: single source. Second row: two sources 40° apart. Third row: 12 random sources.

6. Conclusion

This paper adapted the Ambisonic Multi-Direction Decomposition Method to the analysis and upmixing of ARIRs recorded with compact spherical microphone arrays. Three ingredients were introduced: an SH-domain Wigner-D formulation of the rotation search with analytic gradients and Hessians, a kurtosis loss that favors joint spatio-temporal sparsity over decorrelation of coincident, mutually correlated reflections, and band-wise processing at the frequency-dependent native order of the array, applied uniformly to all compared methods. In a simulation study with an Eigenmike em64, AMDDM matched or outperformed per-band HO-SIRR and REPAIR. Its clearest advantage lies in the fidelity of the reconstructed wavefronts and in resolving coincident reflections, up to the point where the number of arrivals exceeds what a single rotated grid can represent. The consistent advantage of AMDDM stems from its structure: It is a linear, signal-preserving operation without a parametric split into a directional and a diffuse stream.
Perceptual validation and measured ARIRs of real rooms remain future work.

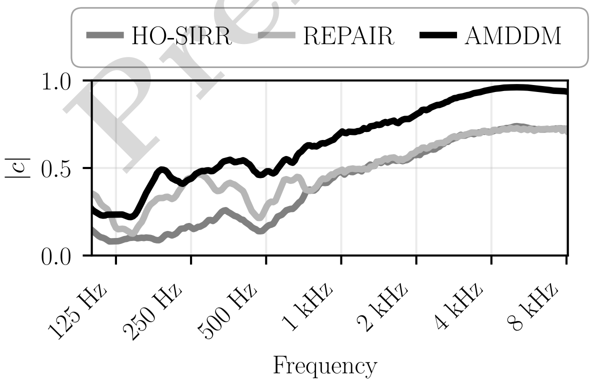

Figure 4: Spatial correlation coefficient $|c|$ over frequency for the three methods HO-SIRR, REPAIR, and AMDDM with third-octave band smoothing.

References

[1] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality.* Springer Nature, 2019.

[2] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering I: Analysis and Synthesis," *Journal of the audio engineering Society*, vol. 53, pp. 1115–1127, 2005.

[3] S. Tervo, J. P. Atynen, and A. Kuusinen, "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc*, vol. 61, pp. 17–28, 2013.

[4] M. Zaunschirm, M. Frank, and F. Zotter, "BRIR synthesis using first-order microphone arrays," 2018.

[5] E. Hoffbauer and M. Frank, "Four-directional ambisonic spatial decomposition method with reduced temporal artifacts," *Journal of the Audio Engineering Society*, vol. 70, no. 12, pp. 1002–1014, 2022.

[6] L. Gölles and M. Frank, "Ambisonic spatial decomposition method with salient / diffuse separation," in *158th AES Convention*, 2025.

[7] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE International Conference on Acoustics Speech and Signal Processing*, Orlando, FL, USA: IEEE, May 2002, p. II–1781–II–1784. doi: 10.1109/ICASSP.2002.5744968.

[8] L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall, "Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution," *Journal of the Audio Engineering Society*, vol. 68, no. 5, pp. 338–354, Jun. 2020, doi: 10.17743/jaes.2020.0026.

[9] L. McCormack, N. Meyer-Kahlen, and A. Politis, "Spatial Reconstruction-Based Rendering of

Microphone Array Room Impulse Responses," *Journal of the Audio Engineering Society*, vol. 71, no. 5, pp. 267–280, May 2023, doi: 10.17743/jaes.2022.0072.

[10] Patrick Fahrngruber, Thomas Röck, and Franz Zotter, "Ambisonic Multi-Direction Decomposition Method." Accessed: Jun. 16, 2026. [Online]. Available: https://pub.dega-akustik.de/DAGA_2026/konferenz-2225.html

[11] F. W. J. Olver *et al.*, "NIST digital library of mathematical functions." Accessed: Jul. 28, 2026. [Online]. Available: https://dlmf.nist.gov/

[12] F. Zotter, *A Linear-Phase Filter-Bank Approach to Process Rigid Spherical Microphone Array Recordings*. 2018.

[13] F. Zotter, "Filterbank rule for spherical microphone array encoding and EQs for above aliasing, transducers, and binaural playback," Jun. 2024.

[14] R. S. Womersley and I. H. Sloan, "How good can polynomial interpolation on the sphere be?," *Advances in Computational Mathematics*, vol. 14, no. 3, pp. 195–226, Apr. 2001, doi: 10.1023/A:1016630227163.

[15] J. Ivanic and K. Ruedenberg, "Rotation Matrices for Real Spherical Harmonics. Direct Determination by Recursion," *The Journal of Physical Chemistry*, vol. 100, no. 15, pp. 6342–6347, Jan. 1996, doi: 10.1021/jp953350u.

[16] mh acoustics LLC, "em64 Eigenmike® Spatial Microphone." Accessed: Jul. 26, 2026. [Online]. Available: https://eigenmike.com/eigenmike-64

[17] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 351–355. doi: 10.1109/ICASSP.2018.8461310.

[18] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979, doi: 10.1121/1.382599.

A. Derivatives of the Off-Diagonal Covariance Loss

Applying the product rule to rotation recurrence relations of [15] yields the partial derivatives $\partial \boldsymbol{R}_N / \partial \varrho_k$ and $\partial^2 \boldsymbol{R}_N / (\partial \varrho_k \partial \varrho_l)$ depending on the corresponding derivatives of $\boldsymbol{R}_{N-1}$, terminating in the closed-form derivatives of the 3×3 rotation matrix $\boldsymbol{R}$.

Written entry-wise in the elements C_{ij} of $\boldsymbol{C}_{\boldsymbol{s}}$, the loss Eq. (7) and its entry derivatives are, with the Kronecker delta δ_{ij},

$$\begin{aligned} \mathcal{L}_{\mathrm{Cov}} &= \sum_{ij} (1 - \delta_{ij})\, C_{ij}^2, \\ \frac{\partial \mathcal{L}_{\mathrm{Cov}}}{\partial C_{ij}} &= 2(1 - \delta_{ij})\, C_{ij}, \end{aligned} \tag{15}$$

so that the chain rule accumulates the contributions of all entries into the scalar derivatives

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathrm{Cov}}}{\partial \varrho_k} &= \sum_{ij} \frac{\partial \mathcal{L}_{\mathrm{Cov}}}{\partial C_{ij}} \frac{\partial C_{ij}}{\partial \varrho_k} = 2 \sum_{ij} (1 - \delta_{ij})\, C_{ij} \frac{\partial C_{ij}}{\partial \varrho_k}, \\ \frac{\partial^2 \mathcal{L}_{\mathrm{Cov}}}{\partial \varrho_k \partial \varrho_l} &= 2 \sum_{ij} (1 - \delta_{ij}) \left(\frac{\partial C_{ij}}{\partial \varrho_k} \frac{\partial C_{ij}}{\partial \varrho_l} + C_{ij} \frac{\partial^2 C_{ij}}{\partial \varrho_k \partial \varrho_l} \right). \end{aligned} \tag{16}$$

Only $\boldsymbol{R}_N$ in Eq. (6) depends on $\boldsymbol{\varrho}$, hence the entry derivatives required above are collected by the matrices, using the symmetry of $\boldsymbol{C}_{\boldsymbol{\mathcal{X}}}$,

$$\begin{aligned} \left[\frac{\partial C_{ij}}{\partial \varrho_k} \right]_{ij} &= \boldsymbol{M}_k + \boldsymbol{M}_k^{\mathrm{T}}, \quad \text{with} \\ \boldsymbol{M}_k &= \boldsymbol{Y}_N^{-\mathrm{T}} \frac{\partial \boldsymbol{R}_N}{\partial \varrho_k} \boldsymbol{C}_{\boldsymbol{\mathcal{X}}} \boldsymbol{R}_N^{\mathrm{T}} \boldsymbol{Y}_N^{-1}, \end{aligned} \tag{17}$$

and, by the product rule applied once more,

$$\begin{aligned} \left[\frac{\partial^2 C_{ij}}{\partial \varrho_k \partial \varrho_l} \right]_{ij} &= \boldsymbol{N}_{kl} + \boldsymbol{N}_{kl}^{\mathrm{T}}, \quad \text{with} \\ \boldsymbol{N}_{kl} = \boldsymbol{Y}_N^{-\mathrm{T}} &\left(\frac{\partial^2 \boldsymbol{R}_N}{\partial \varrho_k \partial \varrho_l} \boldsymbol{C}_{\boldsymbol{\mathcal{X}}} \boldsymbol{R}_N^{\mathrm{T}} + \frac{\partial \boldsymbol{R}_N}{\partial \varrho_k} \boldsymbol{C}_{\boldsymbol{\mathcal{X}}} \frac{\partial \boldsymbol{R}_N^{\mathrm{T}}}{\partial \varrho_l} \right) \boldsymbol{Y}_N^{-1}. \end{aligned} \tag{18}$$

B. Derivatives of the Kurtosis Loss

The kurtosis loss Eq. (12) is differentiated in the same manner. Again only $\boldsymbol{R}_N$ depends on $\boldsymbol{\varrho}$, so the extracted frame differentiates as $\partial \boldsymbol{S} / \partial \varrho_k = \boldsymbol{Y}_N^{-\mathrm{T}} (\partial \boldsymbol{R}_N / \partial \varrho_k)\, \boldsymbol{\mathcal{X}}$, where $\boldsymbol{\mathcal{X}}$ collects the frame of Ambisonic signals, and analogously for the second derivative.

With the elementwise sums $A = \sum_{i,t} S_{it}^4$ and $B = \sum_{i,t} S_{it}^2$, the loss is $\mathcal{L}_{\mathrm{Kurt}} = -A/B^2$. The building blocks are

$$\begin{aligned} \frac{\partial A}{\partial \varrho_k} &= 4 \sum_{i,t} S_{it}^3 \frac{\partial S_{it}}{\partial \varrho_k}, \quad \frac{\partial B}{\partial \varrho_k} = 2 \sum_{i,t} S_{it} \frac{\partial S_{it}}{\partial \varrho_k}, \\ \frac{\partial^2 A}{\partial \varrho_k \partial \varrho_l} &= 12 \sum_{i,t} S_{it}^2 \frac{\partial S_{it}}{\partial \varrho_k} \frac{\partial S_{it}}{\partial \varrho_l} + 4 \sum_{i,t} S_{it}^3 \frac{\partial^2 S_{it}}{\partial \varrho_k \partial \varrho_l}, \\ \frac{\partial^2 B}{\partial \varrho_k \partial \varrho_l} &= 2 \sum_{i,t} \frac{\partial S_{it}}{\partial \varrho_k} \frac{\partial S_{it}}{\partial \varrho_l} + 2 \sum_{i,t} S_{it} \frac{\partial^2 S_{it}}{\partial \varrho_k \partial \varrho_l}, \end{aligned} \tag{19}$$

and the quotient rule on $-A/B^2$ gives

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathrm{Kurt}}}{\partial \varrho_k} &= \frac{2A \frac{\partial B}{\partial \varrho_k}}{B^3} - \frac{\frac{\partial A}{\partial \varrho_k}}{B^2}, \\ \frac{\partial^2 \mathcal{L}_{\mathrm{Kurt}}}{\partial \varrho_k \partial \varrho_l} &= -\frac{\frac{\partial^2 A}{\partial \varrho_k \partial \varrho_l}}{B^2} + \frac{2 \left(\frac{\partial A}{\partial \varrho_k} \frac{\partial B}{\partial \varrho_l} + \frac{\partial A}{\partial \varrho_l} \frac{\partial B}{\partial \varrho_k} \right)}{B^3} \\ &\quad + \frac{2A \frac{\partial^2 B}{\partial \varrho_k \partial \varrho_l}}{B^3} - \frac{6A \frac{\partial B}{\partial \varrho_k} \frac{\partial B}{\partial \varrho_l}}{B^4}. \end{aligned} \tag{20}$$