AMBISONIC DECODER EQUALIZATION IN REVERBERANT ENVIRONMENTS VIA CLOSED-HULL CROSSTALK INVERSION

Alex Tung and Mark Rau

Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, USA
akht@mit.edu

ABSTRACT

In this work, the authors develop a higher order Ambisonic (HOA) decoder that compensates for listening room reverberation by cascading a conventional decode matrix with a crosstalk matrix derived from room impulse responses (RIR) constrained over the listener area, namely by sampling over a spherical boundary according to HOA convention. A decoder is generated for simulated RIRs of mixed specular and diffuse reverberation, and spectral and spatiotemporal energy distributions are shown for directional and diffuse HOA signals rendered through the reverberant room. Results demonstrate that the decoder renders temporally-variant directional sources with higher directivity as compared to a conventional decoder.

1. INTRODUCTION

In order to reconstruct a soundfield, a volumetric distribution of acoustic pressure or velocity must be created or captured using distributions of transducers and rendered such that perceptual cues are preserved. This allows for the recording and transmission of immersive content, transmission of navigable soundfields, and rendering of virtual acoustic environments.

Higher-order Ambisonics (HOA) has arisen as a leading spatial audio capture and rendering paradigm, in which the soundfield is represented by spherical harmonics over an enclosed spherical boundary [1]. When the pressure distribution over the boundary is decomposed at a sufficiently high order of spatial harmonic and sampled with a sufficiently high point density, the enclosed volumetric soundfield can be stored and rendered for playback over point loudspeakers [2].

Several modifications to the HOA workflow account for non-ideal geometries and acoustic-device interactions: This includes compensations for the microphone-encode and loudspeaker-decode arrays [3, 4], as well as intermediate rendering to virtual loudspeakers [5] or subset geometries [6] to mitigate spatial warping from transducer-sparse regions. In addition, extensions for radial extent [7] and parametric soundfield decomposition [8] have been developed for further analysis and transport, all of which have been implemented in plugin suites [9] and standardized workflows for immersive content creation [10]. However, compensations for room reverberation specifically aimed towards HOA decoded signals are limited, with one notable method requiring decomposition of source material into direct and reverberant components for perceptual compensation [11].

Crosstalk matrix inversion is a general soundfield manipulation method that compensates for known transfer functions in the frequency domain; this is accomplished by matrix inversion of cross-path transfer functions between source-loudspeaker and receiver-microphone arrays [12]. This method is popularly used for crosstalk cancellation (CTC), in which acoustic pressure is preserved or minimized at distinct spatial regions. By performing the inversion on head-related transfer functions (HRTF), array transfer functions (ATF), and room control points, CTC has been used for loudspeaker-based reproduction of binaural audio [13, 14], beamformers [15], and personal sound zones (PSZ) [16], respectively.

When performed over room impulse responses (RIR), crosstalk matrix inversion offers compensation of the reverberant characteristic of a listening room and takes form as the multiple-input/output inverse theorem (MINT) [17] [18]. This method has been modified to address high-frequency reconstruction inaccuracies for sparse sensors [19, 20, 21] and extended to equalize microphone array signal for playback in a reverberant room [19]. Such improvements exhibit high robustness to error [22], though the performance evaluations are largely performed in the scope of single-talk speech dereverberation and spectral matching of diffuse fields [23, 24]. There exist few explorations of crosstalk matrix inversion for control points sampled over closed spherical boundaries, and in turn, limited applications of the method for general HOA decoding. Several works have applied crosstalk dereverberation to HOA decoders, but utilize colocated points to restrict the system inversion to reverberant effects only [25, 26].

In this work, the authors introduce an audio renderer consisting of a static HOA decoder cascaded with a crosstalk convolution matrix derived from the RIR of the intended listening room. Compared to prior works, this decoder utilizes a modified matrix of RIRs with control points sampled over a closed hull around the istener area boundary, providing a well-posed system for inversion. Using this decoder, signals are rendered with compensation for the reverberant characteristics of the listening room while preserving the qualities of a soundfield with full periphony. The algorithm is described in Section 2, an evaluation is performed over simulated RIRs in Section 3, and concluding remarks are given in Section 4.

2. ALGORITHMIC DESCRIPTION

2.1. Signal Model

For a listening room consisting of N_s loudspeakers located at points $\mathbf{r}_s = \{\mathrm{r}_{s1}, \cdots, \mathrm{r}_{sN_s}\}$ and N_c control points $\mathbf{r}_c = \{\mathrm{r}_{c1}, \cdots, \mathrm{r}_{cN_c}\}$, the N-point RIRs from loudspeakers to control points can be ar-

ranged as a collection of $N_c \times N_s$ matrices,

$$\mathbf{h} = \{\mathrm{h}_1, \mathrm{h}_2, \cdots, \mathrm{h}_N\}, \tag{1}$$

where for $n \in [1, N]$,

$$\mathrm{h}_n = \begin{bmatrix} h_{1,1}(n) & h_{1,2}(n) & \cdots & h_{1,N_s}(n) \\ h_{2,1}(n) & h_{2,2}(n) & \cdots & h_{2,N_s}(n) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_c,1}(n) & h_{N_c,2}(n) & \cdots & h_{N_c,N_s}(n) \end{bmatrix}, \tag{2}$$

and $h_{i,j}(n)$ is the n^{th} sample of the RIR between the i^{th} control point and the j^{th} loudspeaker. The collection $\mathbf{h}$ can be transformed into the frequency domain,

$$\mathbf{H} = \{\mathrm{H}_1, \mathrm{H}_2, \cdots, \mathrm{H}_K\}, \tag{3}$$

where for frequency k,

$$\mathrm{H} = \begin{bmatrix} H_{1,1}(k) & H_{1,2}(k) & \cdots & H_{1,N_s}(k) \\ H_{2,1}(k) & H_{2,2}(k) & \cdots & H_{2,N_s}(k) \\ \vdots & \vdots & \ddots & \vdots \\ H_{N_c,1}(k) & H_{N_c,2}(k) & \cdots & H_{N_c,N_s}(k) \end{bmatrix}, \tag{4}$$

and $H_{i,j}(k)$ is the transfer function corresponding with $h_{i,j}(n)$ obtained by the N-point discrete Fourier transform (DFT) and evaluated at bin k.

Given a loudspeaker signal $\mathrm{x}_k = [x_1(k), \cdots, x_{N_s}(k)]^{\mathrm{T}}$ and control signal $\mathrm{y}_k = [y_1(k), \cdots, y_{N_c}(k)]^{\mathrm{T}}$ matrix H_k represents a linear map between subspaces $\mathcal{S} \in \mathbb{C}^{N_s}$ and $\mathcal{C} \in \mathbb{C}^{N_c}$,

$$\mathrm{y}_k = \mathrm{H}_k \mathrm{x}_k. \tag{5}$$

For each H_k, a $N_s \times N_c$ matrix C_k that transforms a control signal $\mathrm{y}_k \in \mathcal{C}$ to a loudspeaker signal $\hat{\mathrm{x}}_k \in \mathcal{S}$ may exist,

$$\hat{\mathrm{x}}_k = \mathrm{C}_k \mathrm{y}_k, \tag{6}$$

which can be transformed back to the control space using the original transfer function,

$$\hat{\mathrm{y}}_k = \mathrm{H}_k \hat{\mathrm{x}}_k. \tag{7}$$

2.2. Pre-Processing

The matrix H_k is often ill-conditioned for sparse array geometries and short wavelengths, and an inverse solution of $\mathbf{H}$ commonly results in pre- and post-ringing in the resulting impulse responses arising from the matrix solutions at high frequencies. In order to minimize these artifacts, high-frequency reverberant energy of each RIR is reduced while preserving its main impulse [19, 22]; this is achieved by crossfading the unprocessed RIR with a filtered RIR in the time domain, which is denoted with the operator f,

$$\tilde{h}_{i,j}(n) = f\big(h_{i,j}(n)\big),\ i \in [1, N_c],\ j \in [1, N_s], \tag{8}$$

$$f\big(h(n)\big) = \begin{cases} h(n) & \frac{n}{f_s} \in [0, \tau_0) \\ w_0(n)h(n) + w_1(n)h_{LP}(n) & \frac{n}{f_s} \in [\tau_0, \tau_1) \\ h_{LP}(n) & \frac{n}{f_s} \geq \tau_1 \end{cases}, \tag{9}$$

where w_0 and $w_1 = (1 - w_0)$ are complementary windows over time range $n/f_s \in [\tau_0, \tau_1]$ for gradual transition of the processed RIR, and $h_{LP}(n)$ is the impulse response processed through a low-pass filter (LPF) with passband and stopband criteria $(g_p, f_p), (g_s, f_s)$,

$$10\log_{10}|H_{LP}(j\omega)|^2 \in \pm\, g_p \qquad \omega/2\pi < f_p, \tag{10}$$

$$10\log_{10}|H_{LP}(j\omega)|^2 < -g_s \qquad \omega/2\pi > f_s. \tag{11}$$

Processing each $h_{i,j}$ yields a modified set of system matrices in the frequency domain $\tilde{\mathbf{H}} = \{\tilde{\mathrm{H}}_1, \tilde{\mathrm{H}}_2, \cdots, \tilde{\mathrm{H}}_K\}$ (4).

2.3. System Inversion

In order to account for matrices $\tilde{\mathrm{H}}_k$ that are not full-rank, a solution C_k is obtained via the Tikhonov-regularized left pseudoinverse [27],

$$\mathrm{C}_k = (\tilde{\mathrm{H}}_k^{\mathrm{H}} \tilde{\mathrm{H}}_k + \beta \mathrm{I})^{-1} \tilde{\mathrm{H}}_k^{\mathrm{H}}. \tag{12}$$

The pseudoinverse (12) exists as a closed-form solution that improves inversion robustness by minimizing both the error norm $\|\hat{\mathrm{y}}_k - \mathrm{y}_k\|_2$ and the ℓ_2 solution norm $\|\hat{\mathrm{x}}_k\|_2$. In this work, the maximum ℓ_2 row norm of the solution C_k is used as an additional regularization criterion to construct the pseudoinverse,

$$\varepsilon = \sqrt{\max\left\{\mathrm{C_k C_k^H}\right\}}, \tag{13}$$

in which an optimal regularization constant $\beta_{o,k}$ represents the maximum value that ensures the loudspeakers do not exceed a predetermined gain $\varepsilon < \epsilon_g$. Because the criterion is proportionally related to the error norm, $\beta_{o,k}$ is chosen by binary search for each k, then smoothed by $1/12^{\text{th}}$ octave bands as described in Algorithm 1. The derivation of a pseudoinverse using both the additional regularization criterion and the fractional octave smoothing is consistent with that described in [19].

Algorithm 1 Regularization parameter search.

Input: Preprocessed transfer function $\tilde{\mathbf{H}}$, criterion ϵ_g
Output: Vector of optimal regularization factors β_o

```
for k ∈ [1, K] do
    β_lo ← 1 × 10^−12                          ▷ Initialize
    β_hi ← 20
    repeat
        β ← ½(β_hi + β_lo)
        C_k ← (H̃_k^H H̃_k + βI)^−1 H̃_k^H        ▷ (12)
        ε ← sqrt(max{C_k C_k^H})               ▷ (13)
        if ε > ε_g then
            β_lo ← β                           ▷ Increase β
        else
            β_hi ← β                           ▷ Decrease β
    until β_hi − β_lo < 1 × 10^−14
    β_{o,k} ← β
β_o ← SMOOTH([β_{o,1}, β_{o,2}, ···, β_{o,K}])
return β_o
```

2.4. Geometric Distribution

In a multichannel listening room, loudspeakers are commonly distributed along the walls of the room, which can be interpreted as a discretization of a closed boundary $\mathcal{B}_s$. Control points can be arranged to encapsulate a listening area within the room, similarly

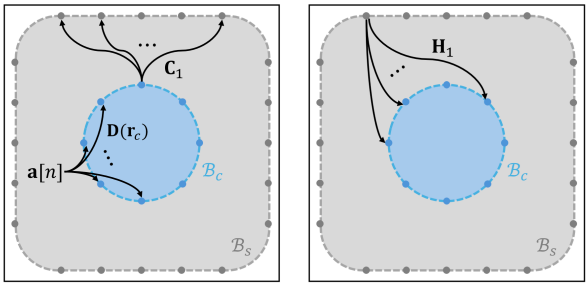

Figure 1: Processing flow from HOA signal $\mathbf{a}(n)$ to intermediate signal $\mathrm{y}_c(n)$ and loudspeaker signal y(n) (a, left), and from intermediate signal to evaluation signal $\mathbf{p}(n)$ (b, right). For matrix TFs $\mathbf{C}, \mathbf{D}$, a subset of paths from one source to several receivers is drawn for readability.

representing a discretization of a closed boundary $\mathcal{B}_c$. From this distribution, it is apparent that the radiant acoustic sources are situated outside of the listening area.

When control points $\mathbf{r}_c \in \mathcal{B}_c$ are arranged to sample the boundary with sufficient density, transforming a multichannel signal through the inverse system $\mathbf{C}$ will yield a soundfield such that the original pressure distribution over the boundary $\mathcal{B}_c$ is preserved without reverberation artifacts present in the forward system $\mathbf{H}$. This differs from a previous work in that the forward system contains a spatial transformation that is radially monotonic from outer loudspeakers to inner control points, rather than a transform between colocated points [25].

When modified in (8), the matrix system $\tilde{\mathbf{H}}$ primarily consists of path differences between loudspeaker locations and control points. Assuming the absence of acoustic baffles in the listening area for the frequencies of interest, enforcing the condition $\mathcal{B}_c \subset \mathcal{B}_s$ is equivalent with the source-exterior constraint of the Kirchhoff-Helmholtz integral equation [28] as illustrated in Figure 1.

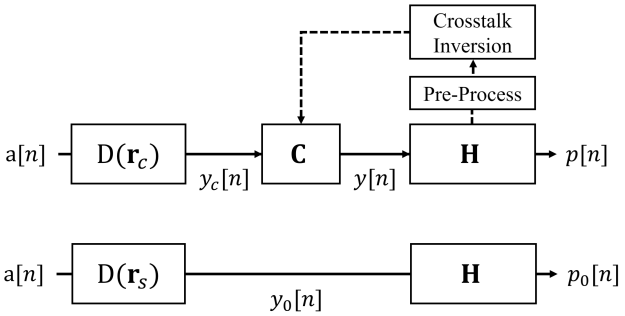

Figure 2: Processing flow for decoding and evaluation signals $\mathrm{y}(n), \mathrm{p}(n)$ using present decoder (top) and similar signals $\mathrm{y}_0(n), \mathrm{p}_0(n)$ using conventional decoder (bottom).

2.5. Decoder Construction

In order to reproduce an HOA signal $\mathbf{a}(n)$ with compensation for listening room reverberation described in Sections 2.2, 2.3 and 2.4, an intermediate signal $\mathrm{y}_c(n)$ is first rendered to the control points by a conventional decoder $\mathrm{D}(\mathbf{r}_c)$,

$$\mathrm{y}_c(n) = \mathrm{D}(\mathbf{r}_c)\mathrm{a}(n). \tag{14}$$

The intermediate signal is then transformed to the final loudspeaker signal $\mathrm{y}(n)$ via matrices C_k (12) collected into a matrix transfer function $\mathbf{C}$ in the same manner as (3),

$$\mathrm{y}(n) = \mathbf{C}\bar{\circledast}\mathrm{y}_c(n), \tag{15}$$

where $\bar{\circledast}$ denotes matrix convolution between the rows of $\mathbf{C}$ across frequency and the vector of signals $\mathrm{y}_c(n)$,

$$y_i(n) = \sum_{j=1}^{N_c} \mathrm{C}_{i,j}(n) \circledast y_{c,j}(n), \tag{16}$$

and $\circledast$ denotes convolution between the $(i,j)^{\text{th}}$ filter of $\mathbf{C}$ and the j^{th} signal of $\mathrm{y}_c(n)$. An illustration of this algorithm flow is shown in Figure 1a.

3. EVALUATION

The reconstruction accuracy of the present renderer can be examined in the expected listening room by its full system matrix $\mathbf{M} \in \mathcal{C}^{N_s \times N_s}$, obtained from matrix-convolving the original RIRs $\mathbf{H}$ with its solution $\mathbf{C}$,

$$\mathbf{M} = \mathbf{H}\bar{\circledast}\mathbf{C}. \tag{17}$$

Additionally, in order to evaluate the robustness of the present renderer against overfitting to the control points, the loudspeaker output (15) can be processed through an evaluation RIR matrix $\mathbf{H}_E$ containing RIRs measured on the contol boundary $\mathcal{B}_c$ at different points different than the control points $\mathbf{r}_c$.

$$\mathbf{p}(n) = \mathbf{H}_E\bar{\circledast}\mathrm{y}(n) = \mathbf{H}_E\bar{\circledast}\left(\mathbf{C}\bar{\circledast}\left(\mathrm{D}(\mathbf{r}_c)\mathrm{a}(n)\right)\right). \tag{18}$$

This output is compared to a control condition, obtained by decoding the HOA soundfield directly to the loudspeaker locations $\mathbf{D}(\mathbf{r}_s)$ to yield $\mathrm{y}_0(n)$, which is matrix convolved through the same RIRs,

$$\mathrm{p}_0(n) = \mathbf{H}_E\bar{\circledast}\mathrm{y}_0(n) = \mathbf{H}_E\bar{\circledast}\big(\mathrm{D}(\mathbf{r}_s)\mathrm{a}(n)\big). \tag{19}$$

The processing flow of the present decoder evaluation is illustrated in Figure 1b, and the processing flow of both decoders under evaluation are depicted in Figure 2.

3.1. Experimental Setup

The algorithm is evaluated in a simulated reverberant environment with $N_s = 26$ loudspeakers distributed over an upper partial ellipsoid and $N_c = 25$ control points regularly arranged over a spherical surface of unit radius ($r = 1$ m) [29]. The arrangement of loudspeakers and control points is shown in Figure 3. HOA signals consist of an impulse placed at the standard front-center location, or $(\theta, \phi) = (0°, 0°)$, and an impulse placed at the cardinal left direction $(\theta, \phi) = (90°, 0°)$.

RIRs are simulated for a rectangular shoebox room of dimensions $3.9 \times 3.6 \times 3.0$ m using a mixed model of specular and stochastic reflections from image-source and ray-tracing methods,

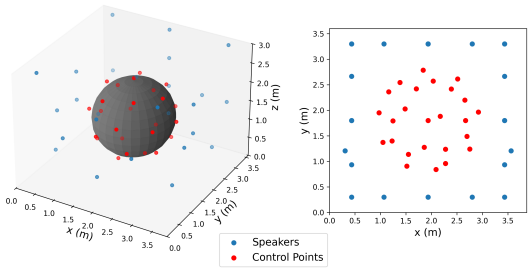

Figure 3: Loudspeaker and control point arrays, orthographic view (left) and top-down view (right).

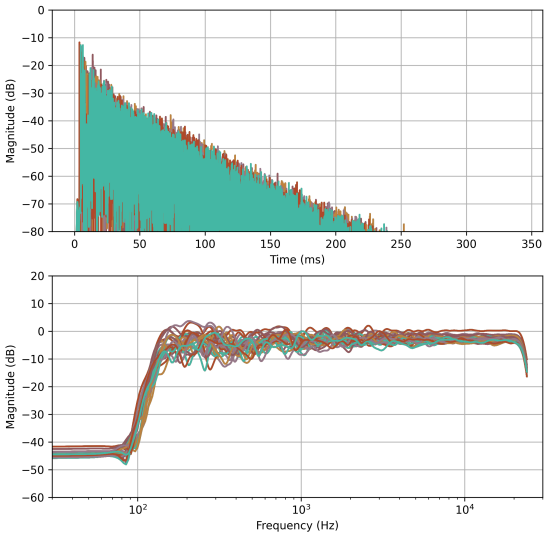

Figure 4: Simulated RIRs, power response (top) and magnitude frequency response (bottom).

respectively [30]. All faces of the room exhibit a broadband absorption coefficient of $\alpha = 0.35$ representing the reverberant characteristic of a plausible listening room. A summary of the RIRs collected is shown in Figure 4.

Given the listening room under evaluation, a LPF with a passband of 2000 Hz is constructed for pre-processing of the RIRs into $\tilde{\mathbf{H}}$(9). In addition, the resulting inverse system $\mathbf{C}$ was passed through a high-pass filter with a corner frequency at 200 Hz to suppress further artifacts at long wavelengths. Parameters related to the reverberant characteristic of the simulated room, as well as parameters for pre-processing, are listed in Table 1.

For evaluation, the point distribution of evaluation RIRs $\mathbf{H}_E$ is chosen to be a 45×30 equiangular grid with azimuth and elevation increments $\Delta_\phi = 8°, \Delta_\theta = 6°$. This sampling allows for a spatially dense analysis of soundfields rendered by the present decoder, with the ability to make direct comparison to soundfields rendered by a conventional HOA decoder.

Table 1: Parameters used in evaluation.

	Value	Unit	Description
α	0.35	–	Absorptivity
RT_{60}	245	ms	60 dB decay
τ_0	1	ms	LPF fade-in time
τ_1	4	ms	LPF fade-out time
f_{pass}	2000	Hz	Passband frequency
f_{stop}	4000	Hz	Stopband frequency
g_{pass}	3	dB	Passband gain
g_{stop}	60	dB	Stopband gain
ϵ_g	0	dB	Inversion criterion

Table 2: Nearest control points for evaluation in Figure 6.

Nominal Location	Angle (°)	Nearest Angle on $\mathbf{r}_c$ (°)
Front Center	$(0, 0)$	$(9.5, 3.1)$
Front Left	$(30, 0)$	$(54.6, -0.5)$
Front Right	$(-30, 0)$	$(-34.4, 7.8)$

3.2. Evaluation Results

Figure 5 shows the system matrix (17) condensed by column. For each column $i \in [1, N_s]$, the transfer function of the diagonal element $\mathbf{M}_{i,i}$ is highlighted. Energy over this element corresponds with energy that had originated from the original loudspeaker. From this figure, it is apparent that at reconstruction, the i^{th} loudspeaker receives 10 dB more energy from the same loudspeaker channel than from other channels at frequencies from 500 Hz to 4 kHz. This is interpreted as a preservation of loudspeaker directivity when processing decoded HOA signals through $\mathbf{C}$. For frequencies outside of this range, the loudspeaker receives more relative energy from other loudspeakers (i.e. elements along each row of $\mathbf{M}$). For interpretability, the highlighted magnitude responses correspond with the nearest locations on $\mathbf{r}_c$ as listed in Table 2.

Figure 6 displays the output through the present decoder, corresponding with processing through $\mathbf{M}$ (17), as well as that through the conventional decoder. Using the present decoder, it is apparent that the control points closest to the originally encoded sources exhibit higher energy at frequencies from 2 kHz to 8 kHz: This difference is approximately 10 dB for the single source and 6 dB for the pair of coherent sources. Although the difference in prominence may be attributed to differences between nominal source and control point locations, it is apparent that there is a local concentration of energy close to the intended location of the HOA soundfield. Using the conventional decoder, the loudspeaker prominences are not present due to spectral mixing of reverberant energy in the room.

Figures 7, 8, and 9 depict the spatiotemporal distribution of the solution through the evaluation RIRs (18), (19). Energy is integrated over windows of $[0, 2000]$ ms, $[0, 10]$ ms, $[10, 50]$ ms, $[50, 80]$ ms, and $[80, 2000]$ ms relative to the impulse and represent the entire decay response as well as common ranges for direct and reverberant portions of an RIR. In turn, the resulting integrated energy levels correspond with quantities for calculating direct-to-reverberant ratio (DRR) and clarity indices ($\mathrm{C}_{50}, \mathrm{C}_{80}$). For all such figures, color ranges are normalized for readability and kept consistent across both decoders to allow for comparisons to be made in accumulated energy.

Figures 7 and 8 depict the decay from impulses located at

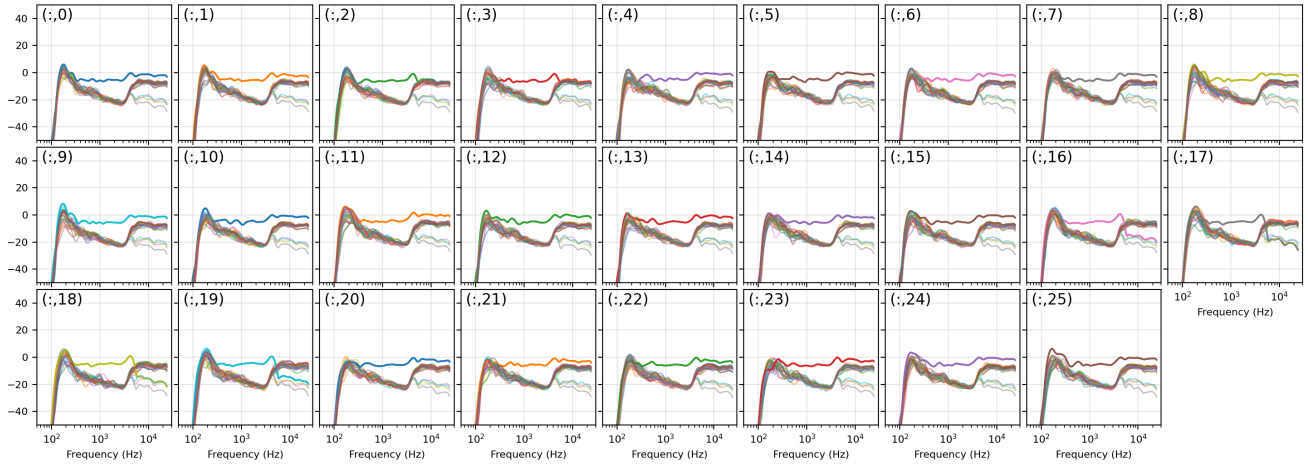

Figure 5: System matrix TF $\mathbf{M}$ (17), condensed by column. Bold plots denote diagonal elements of the matrix TF.

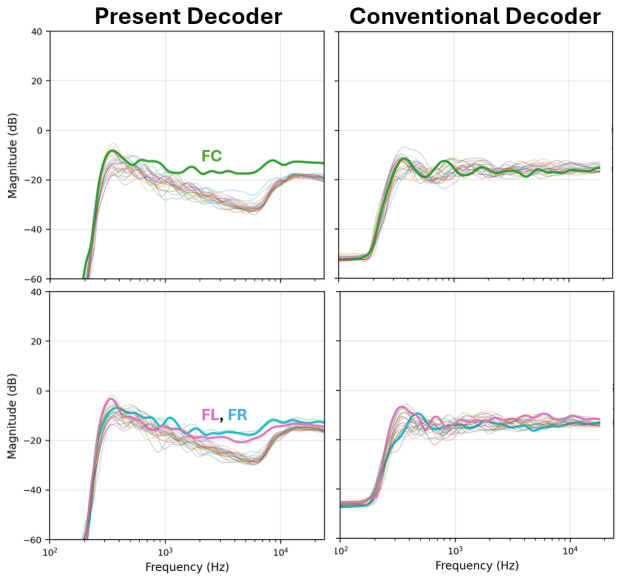

Figure 6: Present decoder and original decoder evaluated around control boundary $\mathcal{B}_c$ for a single impulse at $(0°, 0°)$ (top), coherent impulses at $(\pm 30°, 0°)$ (bottom).

$(0°, 0°)$ and $(90°, 0°)$ respectively. For both sources, it is apparent that energy is concentrated accurately at the HOA source locations, and that the rendered soundfield exhibits higher source directivity using the present decoder as compared to the conventional decoder. The apparent directivities of these sources corroborate the broadband prominence of the magnitude in Figure 6a, suggesting that processing through the compensation matrix $\mathbf{C}$ does not result in overfitting at the control points. For the source at $(90°, 0°)$, the late reverberant field of the signal rendered through the present decoder exhibits higher isotropy, suggesting that prominent reflections are mitigated.

Figure 9 depicts the decay from the interrupted noise stimulus

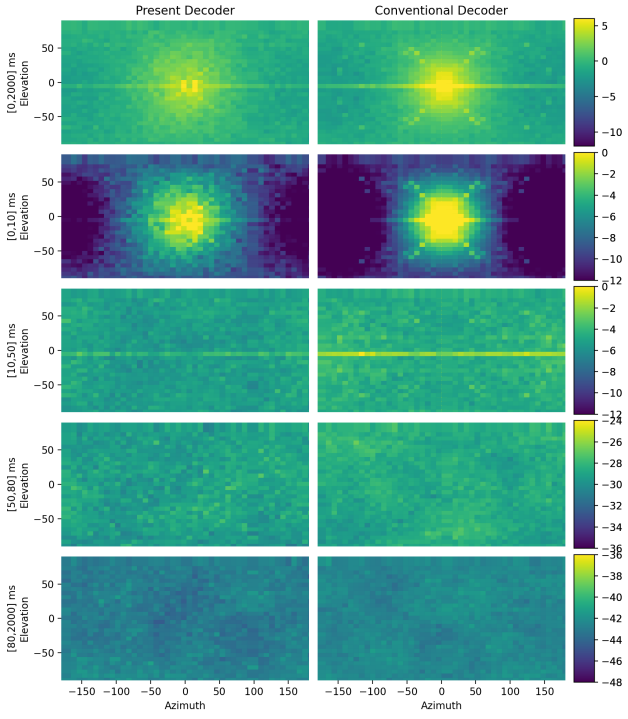

Figure 7: Spatiotemporal distribution of an impulse at $\Omega = (0°, 0°)$ rendered by the present decoder (left) and conventional decoder (right), through listening room RIRs. (dB)

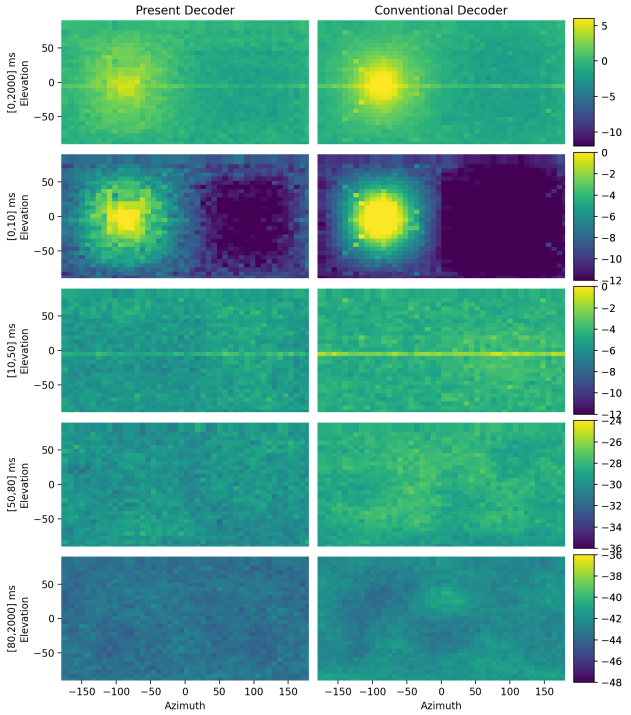

Figure 8: Spatiotemporal distribution of an impulse at $\Omega = (90^\circ, 0^\circ)$ rendered by the present decoder (left) and conventional decoder (right), through listening room RIRs. (dB)

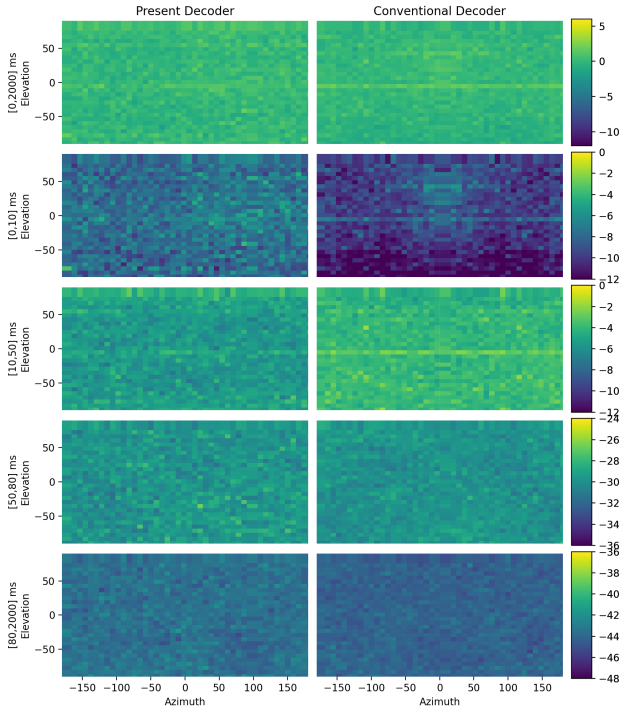

Figure 9: Spatiotemporal distribution of an interrupted diffuse field rendered by the present decoder (left) and conventional decoder (right), through listening room RIRs. (dB)

of a diffuse field; in this case, the integration windows are set to 128 samples prior to the interruption of the noise signal. For the initial $[0, 10]$ ms window, the soundfield from the conventional decoder contains less accumulated energy than that from the present decoder. For remaining time windows, the soundfield decay from both renderers is similar in accumulated energy and spatial distribution. It is worth noting that both signals contain a similar amount of energy when accumulated over the entire decay window, $[0, 2000]$ ms. Combined with the difference at the initial window, it may be interpreted that perceived early reflections are mitigated while maintaining the overall distribution of energy over the room decay.

4. CONCLUSIONS

In this work, the authors introduce a soundfield renderer that compensates for reverberant energy using a crosstalk inversion matrix constructed with geometric conventions compatible with HOA and its associated physical principles. Results suggest that, when rendering HOA signals using this decoder and evaluating its response in the expected listening room, the resulting soundfield exhibits higher spatiotemporal directivity for directional sources, a mild reduction of perceived early reflections, and minimal artifacts in the late reverberant field.

5. REFERENCES

[1] M. J. Gerzon, "Periphony: With-Height Sound Reproduction," *Journal of The Audio Engineering Society*, vol. 21, pp. 2–10, 1973. [Online]. Available: https://api.semanticscholar.org/CorpusID:110210326

[2] J. Daniel, "Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia," Jan. 2000.

[3] J. Meyer and G. Elko, "A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. II–1781–II–1784.

[4] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric Ambisonic Encoding of Arbitrary Microphone Arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2062–2075, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9795678/

[5] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *Journal of the Audio Engineering Society*, vol. 60, no. 10, 2012.

[6] J. Trevino, T. Okamoto, Y. Iwaya, and Y. Suzuki, "High Order Ambisonic Decoding Method for Irregular Loudspeaker Arrays," in *Proceedings of 20th International Congress on Acoustics*, 2010.

[7] J. Daniel, "Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format," in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, May 2003.

[8] V. Pulkki, S. Delikaris-Manias, and A. Politis, Eds., *Parametric Time-Frequency Domain Spatial Audio*, 1st ed. Wiley, Nov. 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119252634

[9] A. Politis, S. Tervo, and V. Pulkki, "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, AB: IEEE, Apr. 2018, pp. 6802–6806. [Online]. Available: https://ieeexplore.ieee.org/document/8462608/

[10] International Telecommunication Union, "ITU-R BS.2127-1 Audio Definition Model Renderer for Advanced Sound Systems," Nov. 2023.

[11] A. Fallah, "A New Approach for Rendering Ambisonics Recordings on Loudspeakers in a Reverberant Environment," *Forum Acusticum*, 2020.

[12] O. Kirkeby, P. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast Deconvolution of Multichannel Systems using Regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, Mar. 1998. [Online]. Available: http://ieeexplore.ieee.org/document/661479/

[13] P. Damaske, "Head-Related Two-Channel Stereophony with Loudspeaker Reproduction," *The Journal of the Acoustical Society of America*, vol. 50, no. 4B, pp. 1109–1115, Oct. 1971. [Online]. Available: https://pubs.aip.org/jasa/article/50/4B/1109/719652/Head-Related-Two-Channel-Stereophony-with

[14] T. Takeuchi, "Systems for Virtual Acoustic Imaging Using the Binaural Principle," Ph.D. dissertation, University of Southampton, 2001.

[15] W.-K. Ma, P.-C. Ching, and B.-N. Vo, "Crosstalk Resilient Interference Cancellation in Microphone Arrays Using Capon Beamforming," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 468–477, Sep. 2004. [Online]. Available: http://ieeexplore.ieee.org/document/1323083/

[16] F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, and P. A. Nelson, "Generation of Private Sound With a Circular Loudspeaker Array and the Weighted Pressure Matching Method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1579–1591, Aug. 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7918631/

[17] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local Sound Field Reproduction Using Digital Signal Processing," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1584–1593, Sep. 1996. [Online]. Available: https://pubs.aip.org/jasa/article/100/3/1584/558300/Local-sound-field-reproduction-using-digital

[18] M. Miyoshi and Y. Kaneda, "Inverse Filtering of Room Acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, Feb. 1988. [Online]. Available: https://ieeexplore.ieee.org/document/1509/

[19] European Telecommunications Standards Institute, "ETSI TS 103 224 A Sound Field Reproduction Method for Terminal Testing Including a Background Noise Database," Nov. 2025.

[20] B. Radlovic, R. Williamson, and R. Kennedy, "Equalization in an Acoustic Reverberant Environment: Robustness Results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000. [Online]. Available: https://ieeexplore.ieee.org/document/841213/

[21] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the Use of Channel Shortening in Multichannel Acoustic System Equalization," *International Workshop on Acoustic Echo and Noise Control*, 2010.

[22] I. Kodrasi, S. Goetze, and S. Doclo, "Increasing the Robustness of Acoustic Multichannel Equalization by Means of Regularization," *International Workshop on Acoustic Echo and Noise Control*, 2012.

[23] S. Gannot and M. Moonen, "Speech Dereverberation Via Sub-Band Implementation of Subspace Methods," *International Workshop on Acoustic Echo and Noise Control*, 2003.

[24] K. Furuya and A. Kataoka, "Robust Speech Dereverberation Using Multichannel Blind Deconvolution With Spectral Subtraction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1579–1591, Jul. 2007. [Online]. Available: https://ieeexplore.ieee.org/document/4244518/

[25] P. Lecomte, P.-A. Gauthier, C. Langrenne, A. Berry, and A. Garcia, "Cancellation of Room Reflections Over an Extended Area Using Ambisonics," *The Journal of the Acoustical Society of America*, vol. 143, no. 2, pp. 811–828, Feb. 2018. [Online]. Available: https://pubs.aip.org/jasa/article/143/2/811/606340/Cancellation-of-room-reflections-over-an-extended

[26] T.-J. Bäumer, S. V. D. Par, and A. Fallah, "Comparative Evaluation of Room Compensation Approaches in Ambisonics Reproduction," 2025.

[27] A. N. Tichonov, V. J. Arsenin, and A. N. Tichonov, *Solutions of Ill-Posed Problems*, ser. Scripta series in mathematics. New York, NY: Wiley, 1977.

[28] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. San Diego, Calif: Academic Press, 1999.

[29] J. Fliege and U. Maier, "A Two-Stage Approach for Computing Cubature Formulae for the Sphere," 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:16541569

[30] R. Scheibler, E. Bezzam, and I. Dokmanic, "Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Apr. 2018, pp. 351–355.