ICA 2022 24th INTERNATIONAL CONGRESS ON ACOUSTICS October 24(Mon) - 28(Fri), 2022 Gyeongju, Korea

PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

!"

Multi-directional parameterisation and rendering of spatial room impulse responses

Leo MCCORMACK[1]; Nils MEYER-KAHLEN;[1] and Archontis POLITIS[2]

[1] Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

[2] Faculty of Information Technology and Communication Sciences, Tampere University, Finland

ABSTRACT

This paper proposes a framework for parameterising and rendering spatial room impulse responses, such that monophonic recordings may be reproduced over a loudspeaker array and exhibit the same spatial characteristics as the captured space. Due to its general formulation, the rendering framework can either operate directly based on the measured microphone array room impulse responses, or on room impulse responses represented in the spherical harmonic domain. The method employs a sound-field model comprising a variable number of simultaneous reflections, which are combined with an ambient component encapsulating anisotropic diffuse reverberation. The isolated reflections are then reproduced over the target loudspeaker array using amplitude-panning, whereas the diffuse reverberation is reproduced over all loudspeakers and subjected to decorrelation operations. The proposed rendering framework has also been made available as an open-source MATLAB toolbox.

Keywords: spatial room impulse responses, sound-field reproduction, spatial audio

1. INTRODUCTION

An array of microphones may be used to capture spatial room impulse responses (RIRs), which store spatial information regarding the magnitudes and directions of reflections, and the direction-dependent energy distribution of diffuse reverberation. Spatial RIR rendering methods then use this captured information to synthesise RIRs corresponding to a target reproduction setup. Subsequently convolving a monophonic input signal with each channel in the rendered RIR allows the input signal to be reproduced over the playback setup, while also exhibiting the spatial characteristics of the captured space. This category of spatial audio rendering has found application in a number of areas, including: perceptual evaluations of concert halls [1], the acoustical analysis of historical buildings [2], and for artistic purposes.

In many cases, such rendering is based on signal-independent processing of the impulse responses; for example, through the Ambisonics encoding of spatial RIR measured with spherical microphone arrays [3], with subsequent decoding to arbitrary reproduction setups [4]. To overcome the inherent limitations associated with a purely signal-independent reproduction, such as limited spatial resolution, several signal-dependent alternatives have been proposed. These approaches are often referred to as parametric methods [5], due to their inherent nature of operation, which entails the estimation of appropriate spatial parameters that are subsequently used to inform synthesis techniques to produce output signals conforming to an assumed sound-field model. In principle, many existing sound-field reproduction methods operating on running signals may also be applied to spatial RIRs. However, a RIR has a particular structure that can be modelled. For example, it can be assumed to consist of direct sound and prominent reflections represented as distinct peaks in the earlier part in the response, followed by a diffuse exponentially decaying reverberant tail. As the reflection density increases quickly, signal-dependent RIR rendering methods may also benefit from higher temporal resolution. Whereas, methods operating on running signals may favour higher frequency resolution, in order to improve the separation of the different sound sources, and more temporal averaging, in order to mitigate audible time-varying artefacts.

[1] leo.mccormack@aalto.fi

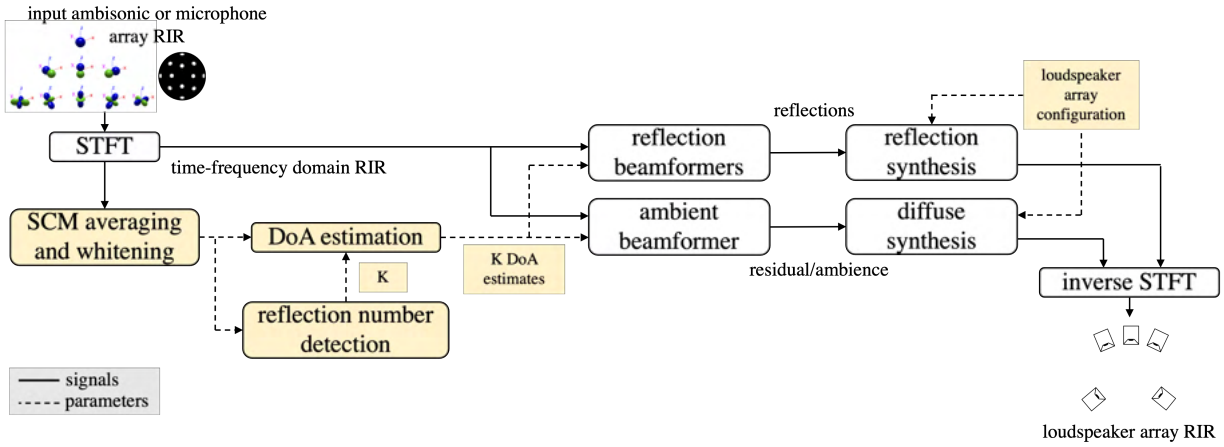


Figure 1: Block diagram of the proposed spatial RIR rendering framework.

The first parametric spatial RIR rendering method was the Spatial Impulse Response Rendering (SIRR) method [6]. The sound-field model employed by the method assumes a single reflection, isotropic diffuse reverberation, or a combination of the two, per time-frequency index. In practice, the method operates based on first-order Ambisonic RIRs, and estimates the direction-of-arrival (DoA) and diffuseness parameter per time-frequency index, through analysis of the pseudo intensity vector. The omni-directional component of the input is reproduced directly over the loudspeaker setup using amplitude panning, and also routed to all loudspeaker channels and decorrelated. The balance between these two streams is then dictated by the diffuseness parameter. The SIRR method was then later extended to higher-order input (HO-SIRR) in [7], by dividing the input into directionally-constrained regions on the sphere, and applying the original SIRR analysis to each. The sector components are panned to the loudspeaker setup, and also: scaled by the diffuseness term, re-encoded into the Ambisonic format, decoded to the target playback setup, and finally decorrelated. Therefore, provided that the reflections land within their own sector, the reproduction of them is more robust with this higher-order formulation. Furthermore, due to the direction-dependent diffuseness term, this method also permits the reproduction of possible anisotropic energy distributions of the diffuse reverberation.

Another example of a parametric spatial RIR rendering approach is the spatial decomposition method (SDM) [8], which employs a sound-field model that assumes a single reflection per time window. In practice, the method employs an open microphone array of four or more omni-directional sensors as input, with the DoA estimated based upon the time-difference of arrivals; although, a spherical harmonic domain variant, employing the pseudo intensity vector, is also available in the SDM toolbox [9].

In this paper, a new multi-directional spatial RIR rendering framework is proposed[1], which builds on the COding and Multi-Directional Parameterisation of Ambisonic Sound Scenes (COMPASS) sound-field model [10], which was recently explored for running signals in [11]. The principles for decomposing RIRs developed recently in [12] are also based on a similar model.

2. PARAMETRIC RENDERING FRAMEWORK

This section provides an overview of the proposed parametric rendering framework, which is also depicted in Fig. 1.

2.1 Sound-field model

It is assumed that the input Q-channel microphone array or Ambisonic RIR has first been transformed into the time-frequency domain $\mathbf{x}(t,f) \in \mathbb{C}^{Q\times 1}$, where t and f denote the down-sampled time and frequency indices, respectively. This is commonly achieved through the application of a short-time Fourier transform (STFT). The spatial covariance matrices (SCM) of the input RIRs are then obtained as $\mathbf{C}_\mathrm{x}(t,f) = \mathbb{E}[\mathbf{x}(t,f)\mathbf{x}^\mathrm{H}(t,f)] \in \mathbb{C}^{Q\times Q}$, where $\mathbb{E}[.]$ denotes the expectation operator. It is then assumed that $K < Q$ simultaneous reflections $\mathbf{r}$ are active at each time-frequency index, which are incident from

[1] A MATLAB toolbox of the proposed framework may be found here: `https://github.com/leomccormack/REPAIR`

directions $\mathbf{\Gamma}_\mathrm{r} = [\gamma_1, ..., \gamma_K]$; where $\gamma_k \in \mathbb{R}$ denotes the direction of the k^{th} reflection.

The input RIR may therefore be described as

$$\mathbf{x}(t,f) = \mathbf{A}_\mathrm{r}(f)\mathbf{r}(t,f) + \mathbf{d}(t,f) + \mathbf{n}(t,f), \tag{1}$$

where $\mathbf{A}_\mathrm{r} = [\mathbf{a}(\gamma_1), ..., \mathbf{a}(\gamma_K)] \in \mathbb{C}^{Q\times K}$ is a matrix containing the array transfer functions corresponding to each reflection direction; $\mathbf{d} \in \mathbb{C}^{Q\times 1}$ is a vector encapsulating diffuse reverberation; and $\mathbf{n} \in \mathbb{C}^{Q\times 1}$ denotes sensor noise. Note that for spherical microphone arrays, descriptions of array radius and sensor positions may be used to obtain analytical array transfer functions [13]. In the general case, the array transfer functions may be obtained from free-field array impulse response measurements, or through numerical simulations, of the array in question. Alternatively, for a spherical harmonic domain receiver, the array transfer functions may be replaced by frequency-independent spherical harmonic weights [3]. It is henceforth assumed that these array transfer functions $\mathbf{A} \in \mathbb{C}^{Q\times V}$ are available for a dense uniformly distributed grid, $\mathbf{\Gamma} = [\gamma_1, ..., \gamma_V]$, of V directions.

2.2 SCM frequency-averaging

Contrary to parametric methods operating on running signals, RIRs typically involve a single source/receiver combination, and thus the reflection signals are likely to be more coherent compared to the direct path signals of the (typically) multiple source/receiver scenarios that are captured and rendered by parametric methods operating based upon running signals. Therefore, to alleviate problems with subspace localisation methods, which assume full-rank/incoherent directional sounds, it may be beneficial to average the SCMs over frequency. This has the benefit of increasing the effective rank of the SCMs used for the spatial analysis, given the same temporal resolution; with the penalty of reduced frequency-resolution. Although, it is noted that high frequency-resolution may be less useful for single source/receiver combinations, as also shown in the perceptual studies conducted in [7]. Averaging the SCMs more across frequency, and less across time, may also make more intuitive sense, when considering how the reflection density in RIRs increases greatly over time.

In the proposed framework, the WINGS coherent-focusing method [14] is employed for the task of averaging the uniformly-spaced STFT input SCMs to form octave band averaged SCMs. The SCM corresponding to each octave-band centre frequency f_0 is obtained as [15]

$$\mathbf{C}_\mathrm{x}^{(\mathrm{OCT})}(f_0) = \sum_{f_i=f_l}^{f_u} \mathbf{T}_\mathrm{coh}(f_i, f_0)\mathbf{C}_\mathrm{x}(f_i)\mathbf{T}_\mathrm{coh}^\mathrm{H}(f_i, f_0), \tag{2}$$

where f_l and f_u denote the lower and upper frequency indices that define the octave-band grouping, and $\mathbf{T}_\mathrm{coh} \in \mathbb{C}^{Q\times Q}$ is the coherent focusing matrix, which is computed as [14]

$$\mathbf{T}_\mathrm{coh}(f, f_0) = [\mathbf{A}(f_0)\mathbf{Y}^\mathrm{T}][\mathbf{A}(f)\mathbf{Y}^\mathrm{T}]^\dagger, \tag{3}$$

where † denotes the Moore-Penrose pseudo inverse, and $\mathbf{Y} \in \mathbb{C}^{Q\times(N+1)^2}$ are spherical harmonic weights up to order N for the same directions as used to measure/simulate the array transfer functions. Note that in the case of a spherical harmonic domain receiver (using broad-band spherical harmonic weights as steering vectors), this coherent focusing operation is intrinsically bypassed since $\mathbf{T}_\mathrm{coh}$ becomes an identity matrix.

2.3 SCM whitening

The parametric spatial analysis techniques described in the following sections are based on the subspace principles of array signal processing. When attempting to detect the number of reflections, for example, many available algorithms rely on the eigenvalues of the array SCM all being equal when the SCM is describing only sensor noise. However, in the present case, it may be assumed that the diffuse reverberation would likely have more energy than the noise floor, i.e. $\mathbf{tr}[\mathbf{dd}^\mathrm{H}] > \mathbf{tr}[\mathbf{nn}^\mathrm{H}]$. Therefore, it may be beneficial to apply a spatial whitening operation on the array SCMs, such that they become more diagonal when the array is capturing only diffuse reverberation, which may subsequently lead to a more accurate detection of the number of reflections.

This spatial whitening is conducted by first obtaining diffuse-coherence matrices averaged over octave

band groupings, similarly as in Eq. 2, and decomposing them as [11]

$$\mathbf{D}^{(\mathrm{OCT})}(f_0) = \sum_{f_i=f_l}^{f_u} \mathbf{T}_{\mathrm{coh}}(f_i, f_0)\mathbf{D}(f_i)\mathbf{T}_{\mathrm{coh}}^{\mathrm{H}}(f_i, f_0), \tag{4}$$

$$= \mathbf{R}\mathbf{\Lambda}\mathbf{R}^{\mathrm{H}}, \tag{5}$$

where $\mathbf{D} = \mathbf{A}\mathbf{A}^{\mathrm{H}} \in \mathbb{C}^{Q\times Q}$ is the diffuse coherence matrix of the array per STFT frequency bin. Note that time and frequency indices are henceforth omitted, unless required for clarity.

This decomposition then permits the acquisition of a spatial whitening matrix $\mathbf{T}_{\mathrm{w}} = \mathbf{\Lambda}^{-1/2}\mathbf{R}^{\mathrm{H}} \in \mathbb{C}^{Q\times Q}$, in order to obtained spatially-whitened SCMs as

$$\hat{\mathbf{C}}_{\mathrm{x}}^{(\mathrm{OCT})} = \mathbf{T}_{\mathrm{w}}\mathbf{C}_{\mathrm{x}}^{(\mathrm{OCT})}\mathbf{T}_{\mathrm{w}}^{\mathrm{H}}. \tag{6}$$

Note that in the case of a spherical harmonic domain receiver (using broad-band spherical harmonic weights as steering vectors), this spatial whitening operation is also intrinsically bypassed.

2.4 Reflection number detection

For detecting the number of reflections over time and per octave band, the frequency-averaged and spatially whitened SCMs are first decomposed as

$$\hat{\mathbf{C}}_{\mathrm{x}}^{(\mathrm{OCT})} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}} = \sum_{k=1}^{K} \lambda_k \mathbf{v}_k \mathbf{v}_k^{\mathrm{H}} + \sum_{k=K+1}^{Q} \lambda_k \mathbf{v}_k \mathbf{v}_k^{\mathrm{H}}, \tag{7}$$

where $\lambda_1 > ... > \lambda_Q$ are the eigenvalues in descending order, and $\mathbf{v}_k$ are their respective eigenvectors.

The SORTE detection algorithm [16] is then employed, which is based on first determining the differences between eigenvalues as

$$\nabla\lambda_i = \lambda_i - \lambda_{i+1}, \quad \text{for } i = 1, ..., Q-1, \tag{8}$$

with an estimate of the number of reflections obtained with the following:

$$K_{\mathrm{SORTE}} = \underset{k}{\operatorname{argmin}}\, f(k) \quad \text{for } k = 1, ..., Q-3, \tag{9}$$

$$f(k) = \begin{cases} \frac{\sigma_{k+1}^2}{\sigma_k^2}, & \sigma_k^2 > 0 \\ +\infty, & \sigma_k^2 = 0 \end{cases}, \quad \text{for } k = 1, ..., Q-2, \tag{10}$$

$$\sigma_k^2 = \frac{1}{Q-k}\sum_{i=k}^{Q-1}\left(\nabla\lambda_i - \frac{1}{Q-k}\sum_{i=k}^{Q-1}\nabla\lambda_i\right)^2. \tag{11}$$

However, it is noted that the SORTE algorithm can have the tendency to overestimate the true number of reflections in practice, often during periods of low direct-to-diffuse energy ratios. Therefore, it can be beneficial to constrain the estimate based on a diffuseness measure, in order to obtain a more conservative reflection number estimate as

$$K_{\mathrm{SORTED}} = \min\big(K_{\mathrm{SORTE}}, \lfloor (Q-1)\psi + 1 \rfloor\big) \tag{12}$$

where $\lfloor . \rfloor$ denotes the floor operator, and $\psi \in [0, 1]$ is a diffuseness measure. In this framework, the diffuseness parameter is estimated using the COMEDIE algorithm [17] as

$$\psi = 1 - \frac{\beta}{\beta_0}, \tag{13}$$

where $\beta_0 = 2(Q-1)$, $\beta = \frac{1}{\langle\lambda\rangle}\sum_{q=1}^{Q}|\lambda_q - \langle\lambda\rangle|$, and $\langle\lambda\rangle = \frac{1}{Q}\sum_{q=1}^{Q}\lambda_q$.

2.5 Reflection DoA estimation

Now that the number of reflections K at each time and in each octave band has been detected, the MUltiple-Signal Classification (MUSIC) approach [18] is be employed to estimate their directions

$$P_{\text{MUSIC}}(\boldsymbol{\gamma}) = \frac{1}{||\mathbf{V}_{\text{n}}^{\text{H}} \mathbf{T}_{\text{w}} \mathbf{a}(\boldsymbol{\gamma}, f_0)||^2}, \quad \text{for } \boldsymbol{\gamma} \in \boldsymbol{\Gamma}, \tag{14}$$

where $\mathbf{V}_{\text{n}} \in \mathbb{C}^{Q \times (Q-K)}$ is the noise subspace, consisting of the eigenvectors corresponding to the lowest $Q-K$ eigenvalues. A peak finding algorithm is then used to obtain the DoA estimates.

2.6 Rendering reflections

The signals of the direct sound and reflections are then estimated based upon the application of a beamforming matrix as

$$\mathbf{r} = \mathbf{W}_{\text{r}} \mathbf{x}, \tag{15}$$

with one suitable beamforming design for this task being [10]

$$\mathbf{W}_{\text{r}} = (\mathbf{A}_{\text{r}}^{\text{H}} \mathbf{A}_{\text{r}} + \beta^2 \mathbf{I}_K)^{-1} \mathbf{A}_{\text{r}}^{\text{H}}, \tag{16}$$

where $\beta > 0$ is a regularisation parameter, and $\mathbf{I}$ is an identity matrix. Note that for a single direction, this beamformer reverts to a matched-filter beamformer (or hyper-cardioid/maximum directivity beamformer in the spherical harmonic domain). Whereas, for multiple directions, each row of beamforming weights correspond to beamforming weights derived with a unity-gain constraint towards the respective reflection direction, and null-constraints towards the other reflections.

The estimated reflection signals may then be spatialised directly over an L-channel loudspeaker setup, with directions $\boldsymbol{\Gamma}_L$, as

$$\mathbf{y}_{\text{r}} = \mathbf{G}_{\text{r}} \mathbf{r} = \mathbf{G}_{\text{r}} \mathbf{W}_{\text{r}} \mathbf{x}, \tag{17}$$

where $\mathbf{G}_{\text{r}} = [\mathbf{g}(\boldsymbol{\gamma}_1), ..., \mathbf{g}(\boldsymbol{\gamma}_K)] \in \mathbb{R}^{L \times K}$ is a matrix of vector-base amplitude panning gains $\mathbf{g} = [g_1, ..., g_L]^{\text{T}}$ [19], which correspond to the same DoAs as used to steer the beamformers.

2.7 Rendering diffuse reverberation

The estimation of the diffuse reverberant signals $\mathbf{d} \in \mathbb{C}^{Q \times 1}$ is conducted by spatially subtracting the reflection signals from the input, in order to obtain a residual component

$$\mathbf{d} = \mathbf{W}_{\text{d}} \mathbf{x}, \tag{18}$$

which encapsulates an anisotropic representation of diffuse components in the response. This ambient extraction matrix is calculated as [10]

$$\mathbf{W}_{\text{d}} = \mathbf{I}_Q - \mathbf{A}_{\text{r}} \mathbf{W}_{\text{r}}. \tag{19}$$

Once the ambient array signals have been determined, they are subsequently reproduced over the same loudspeaker setup using a beamforming matrix, $\mathbf{G}_{\text{d}} \in \mathbb{C}^{L \times Q}$, as

$$\mathbf{y}_{\text{d}} = d_{\text{EQ}} \mathbf{G}_{\text{d}} \mathbf{d} = d_{\text{EQ}} \mathbf{G}_{\text{d}} \mathbf{W}_{\text{d}} \mathbf{x}, \tag{20}$$

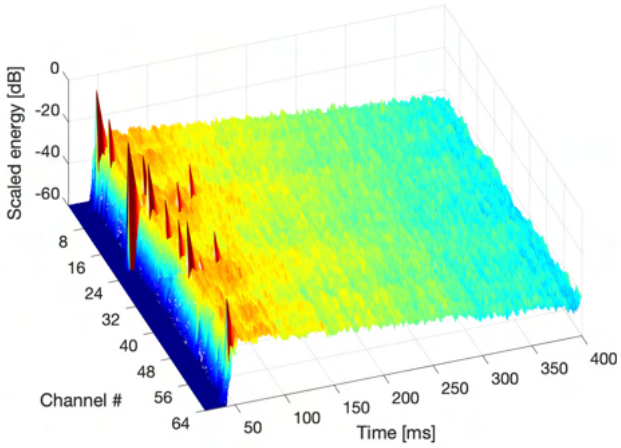
where $d_{\text{EQ}} = \mathbf{tr}[\mathbf{A}\mathbf{A}^{\text{H}}]^{-1/2}$ is a diffuse-field equalisation term.

In the simplest case, this beamforming matrix may be based on a set of matched filters as

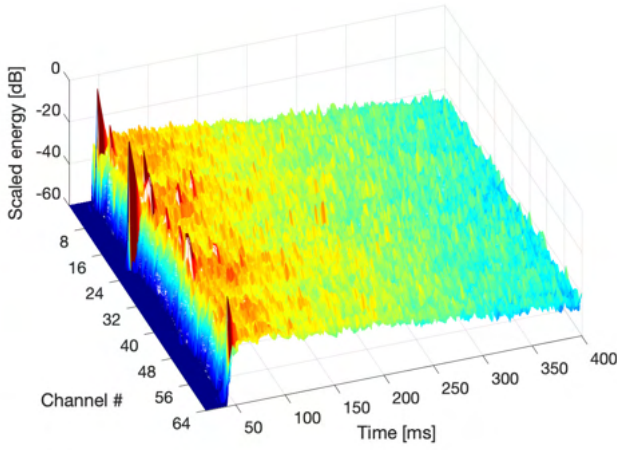
$$\mathbf{G}_{\text{d}} = \mathbf{A}_{\text{d}}^{\text{H}} \, \text{Diag}[\mathbf{A}_{\text{d}}^{\text{H}} \mathbf{A}_{\text{d}}]^{-1}, \tag{21}$$

where $\mathbf{A}_{\text{d}} = [\mathbf{a}(\boldsymbol{\gamma}_1), ..., \mathbf{a}(\boldsymbol{\gamma}_L)] \in \mathbb{C}^{Q \times L}$ are the array steering vectors for each loudspeaker direction, and Diag[.] denotes the construction of a diagonal matrix based on the diagonal entries of the enclosed matrix. Note, that in the spherical harmonic domain, and in Ambisonics terminology, this beamforming matrix reverts to a sampling Ambisonics decoding matrix.

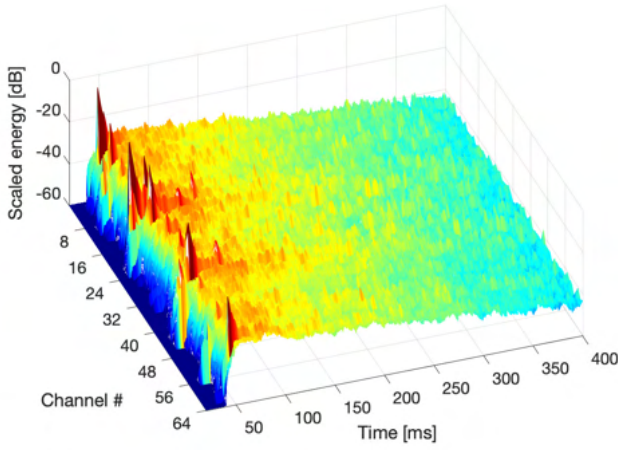
Optionally, an energy-preserving alternative may be derived by forcing the beamforming matrix to be



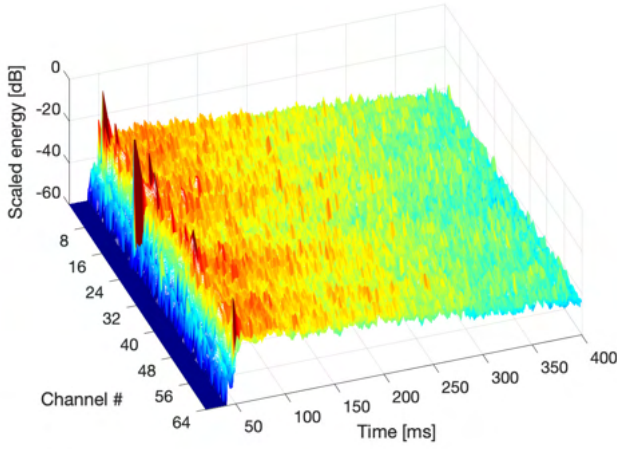
(a) Reference loudspeaker array RIR



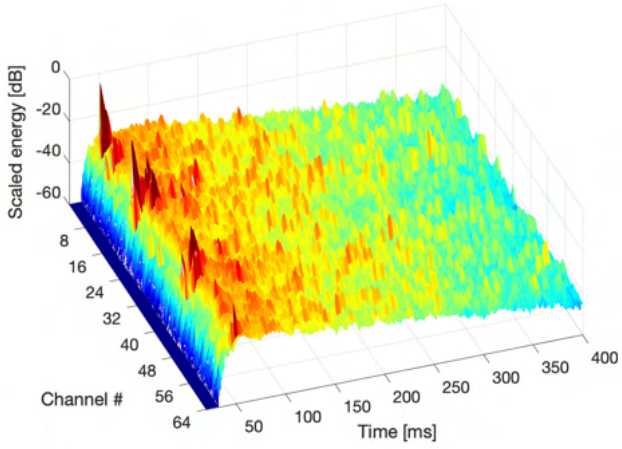
(b) HO-SIRR rendering of fourth-order Ambisonic RIR



(c) REPAIR rendering of fourth-order Ambisonic RIR



(d) HOSIRR rendering of fourth-order encoded Eigenmike32 RIR



(e) REPAIR rendering of Eigenmike32 RIR

Figure 2: Energy plotted over time and loudspeaker channel for the rendered spatial RIR.

unitary, as described in [11]

$$\mathbf{A}_{\mathrm{d}}^{\mathrm{H}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{H}}, \tag{22}$$

$$\hat{\mathbf{G}}_{\mathrm{d}} = \frac{1}{\sqrt{V}}\mathbf{U}^{(\mathrm{trunc})}\mathbf{V}^{\mathrm{H}}, \tag{23}$$

where $\mathbf{U}^{(\mathrm{trunc})} \in \mathbb{C}^{Q\times Q}$ denotes a truncation, in order to retain only the first Q rows. Note that in Ambisonics terminology, and if broad-band spherical harmonic weights are used as the array steering vectors, this reverts to the energy-preserving ambisonic decoder (EPAD) design proposed in [20].

2.8 Overall rendering

The final loudspeaker array RIR is then obtained as

$$\mathbf{y}(t,f) = \mathbf{y}_{\mathrm{r}}(t,f) + \mathscr{D}[\mathbf{y}_{\mathrm{d}}(t,f)], \tag{24}$$

where $\mathscr{D}[.]$ denotes an optional decorrelation operation on the enclosed loudspeaker signals, which may be used to enforce diffuse properties of the rendering.

3. Example multi-channel energy plots

To provide some informal insights into the performance of the proposed spatial RIR rendering framework, multi-channel energy plots were produced using the same reference 64-channel loudspeaker RIRs described in [7]. This reference RIR is plotted in Fig. 2(a). The proposed rendering method was applied to an ideal fourth-order Ambisonic receiver and a simulated Eigenmike32 (a 32-channel spherical microphone array with a radius of 42 mm), which captured the same acoustic scenario. The rendered responses using the proposed framework, and targeting the same loudspeaker setup, are depicted in Figs. 2(c,e). Serving as an additional visual reference, renders for the same loudspeaker array using the HO-SIRR method [7] are also included in Figs. 2(b,d). It can be observed that for the ideal Ambisonic receiver, the rendered responses using the two methods appear to be very similar. Whereas for the Eigenmike32, the early part of the response appears to be slightly closer to the reference when using the proposed rendering method, when compared to the HO-SIRR render, while the later part of the response appears more turbulent with the proposed framework. This is most likely due to the fact that SORTE tends to detect multiple individual reflections also in the late part of the response, which the SORTED modification does not mitigate fully. Evaluating the perceived performance of the proposed framework and investigating more accurate source number estimators, which are better behaved in the late part of the response, is a topic of future work.

4. SUMMARY

This paper has proposed a multi-directional parametric framework for rendering a microphone array room impulse response (RIR), in order to synthesise a RIR corresponding instead to an arbitrary loudspeaker setup. The framework can operate either directly on the microphone array RIRs themselves, or on a spherical harmonic domain representation of the array RIRs (i.e. an Ambisonic RIR). Through informal observations of the energy of the multi-channel responses, it is shown that the proposed rendering framework can produce similar renderings to an existing state-of-the-art rendering method, when compared alongside a reference response, and using Ambisonic RIRs as input. Whereas, when using RIRs corresponding to a 32-channel spherical microphone array as input, the renders appear to be slightly different, and thus formally evaluating the perceptual performance of the framework is a topic of future work. An open-source MATLAB toolbox of the rendering framework described in this paper is also made publicly available.

REFERENCES

1. Lokki T, Pätynen J, Kuusinen A, Tervo S. Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences. The Journal of the Acoustical Society of America. 2016 Jul;140(1):551-62.
2. Katz BF, Weber A. An Acoustic Survey of the Cathédrale Notre-Dame de Paris before and after the

Fire of 2019. In: Acoustics. vol. 2. Multidisciplinary Digital Publishing Institute; 2020. p. 791-802.
3. Rafaely B. Fundamentals of spherical array processing. vol. 8. Springer; 2015.
4. Zotter F, Frank M. Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality. Springer Nature; 2019.
5. Pulkki V, Delikaris-Manias S, Politis A. Parametric time-frequency domain spatial audio. John Wiley & Sons; 2017.
6. Merimaa J. Spatial Impulse Response Rendering I: Analysis and Synthesis. J Audio Eng Soc. 2005 Dec;53(12):1115-27.
7. McCormack L, Pulkki V, Politis A, Scheuregger O, Marschall M. Higher-order Spatial Impulse Response Rendering: Investigating the perceived effects of spherical order, dedicated diffuse rendering, and frequency resolution. J Audio Eng Soc. 2020 May;68(5):368-54.
8. Tervo S, Tynen JP, Kuusinen A, Lokki T. Spatial Decomposition Method for Room Impulse Responses. J Audio Eng Soc. 2013 Mar;61(1):17-28.
9. Tervo S. SDM Toolbox;. Available from: `https://se.mathworks.com/matlabcentral/fileexchange/56663-sdm-toolbox`.
10. Politis A, Tervo S, Pulkki V. COMPASS: Coding and multidirectional parameterization of ambisonic sound scenes. In: IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP); 2018. .
11. McCormack L, Politis A, Gonzalez R, Lokki T, Pulkki V. Parametric Ambisonic Encoding of Arbitrary Microphone Arrays. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2022.
12. Deppisch T, Ahrens J, Garí SVA, Calamia P. Spatial Subtraction of Reflections from Room Impulse Responses Measured with a Spherical Microphone Array. In: 2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). IEEE; 2021. p. 346-50.
13. Teutsch H. Modal array signal processing: principles and applications of acoustic wavefield decomposition. vol. 348. Springer; 2007.
14. Beit-On H, Rafaely B. Focusing and frequency smoothing for arbitrary arrays with application to speaker localization. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2020;28:2184-93.
15. McCormack L, Gonzalez R, Fernandez J, Hold C, Politis A. Parametric Ambisonic Encoding using a Microphone Array with a One-plus-Three Configuration. In: Audio Engineering Society Conference: 2022 AES International Conference on Audio for Virtual and Augmented Reality. Audio Engineering Society; 2022. .
16. Han K, Nehorai A. Improved source number detection and direction estimation with nested arrays and ULAs using jackknifing. IEEE Trans Signal Processing. 2013;61(23):6118-28.
17. Epain N, Jin CT. Spherical harmonic signal covariance and sound field diffuseness. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2016;24(10):1796-807.
18. Schmidt R. Multiple emitter location and signal parameter estimation. IEEE transactions on antennas and propagation. 1986;34(3):276-80.
19. Pulkki V. Virtual sound source positioning using vector base amplitude panning. Journal of the audio engineering society. 1997;45(6):456-66.
20. Zotter F, Pomberger H, Noisternig M. Energy-preserving ambisonic decoding. Acta Acustica united with Acustica. 2012;98(1):37-47.