PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Higher-order processing of spatial impulse responses

Leo MCCORMACK[(1)], Archontis POLITIS[(1)], Oliver SCHEUREGGER[(2)], Ville PULKKI[(1)]

[(1)]Aalto University, Department of Signal Processing and Acoustics, Finland, *leo.mccormack@aalto.fi*
[(2)]Technical University of Denmark, Department of Health Technology, Denmark

ABSTRACT

This article pertains to parametric rendering of microphone array impulse responses, such that the spatial characteristics of a captured space may be imposed onto a monophonic input signal and reproduced over an array of loudspeakers. Parametric methods operate by analysing a set of spatial parameters, dividing the response into components based on an assumed sound-field model, and rendering the components to the loudspeaker array using techniques informed by the analysis. For a direct/diffuse model, the sound is divided into non-diffuse and diffuse components, which are reproduced using directional and surrounding reproduction methods, respectively. In many cases, the input is first divided into frequency bands for the analysis and reproduction. In this article, a method capable of accommodating higher-order spherical harmonic input is proposed, which, based upon initial testing, appears to overcome some of the limitations exhibited by existing methods. The proposed approach operates by partitioning the sound-field into multiple directionally biased sectors, which are then analysed independently. The non-diffuse components are reproduced individually for each sector using amplitude-panning, whereas the diffuse components are encoded back into the spherical harmonic domain, and subsequently reproduced via linear decoding, followed by decorrelation. An open-source implementation of the proposed method is also described.

Keywords: spatial impulse response rendering, spherical harmonic domain, loudspeaker reproduction

1 INTRODUCTION

Over the years, only a small hand-full of methods have been proposed for the task of reproducing an anechoic input signal, such that it may exhibit the spatial characteristics of a captured space. Essentially, these methods operate by synthesising suitable loudspeaker or headphone impulse responses (IRs), using microphone array IR measurements of a space. Ideally, the reproduced transfer function at the listening position, should exactly match that of the transfer function of the original space in the recording position; naturally, the listening room should be as anechoic as possible. Existing methods proposed for this task may be loosely categorised as either non-parametric or parametric in nature. Non-parametric methods operate via a linear combination of the microphone IRs, in order to synthesise the loudspeaker or headphone IRs. Ambisonics is one such example, which has had many decoding strategies proposed over the years [1, 2, 3] and operates on spherical harmonic IRs (or signals) as input; this is an intermediate format, acquired after suitable spatial encoding of the microphone array input [4]. Methods formulated in the spherical harmonic domain (SHD) are of particular interest, as they allow the microphone array specifications to be largely abstracted away from the algorithms that employ them. The main issue with non-parametric methods, however, is that they are inherently limited by the spatial resolution of the input format, which is especially poor when employing first or lower-order spherical harmonic input. This ultimately results in erroneously high signal-coherence between the loudspeaker or binaural channels, which, in turn, leads to the incursion of several perceptual deficiencies, including: localisation ambiguity, signal colouration, comb-filtering and a loss of envelopment [5, 6, 7, 8].

In an attempt to overcome many of the limitations associated with non-parametric rendering approaches, parametric alternatives have been developed [9, 10, 11, 12, 13, 14, 15]. These methods exploit the predicable structure of room IRs, which often exhibit strong peaks in the early response, corresponding to early directional components, followed by a much denser and exponentially decaying tail, corresponding to late reverberation.

Parametric methods, therefore, essentially employ a sound-field model, which lays out a set of assumptions regarding the composition of the sound-field and the response. Many of which are also considered to be perceptually motivated, as they often factor in the known temporal, frequency, and spatial resolution of human hearing [16, 17, 18, 19].

1.1 Spatial Impulse Response Rendering (SIRR)

Spatial Impulse Response Rendering (SIRR) [9, 10] was the first proposed parametric reproduction method, which built upon the spatial analysis techniques described in [20, 21, 22, 23]. The method employs a sound-field model which assumes the existence of many narrow-band and time-varying directional components and an isotropic diffuse-field. To realise this model in practice, SIRR operates in the time-frequency domain and estimates a source direction-of-arrival (DoA) and a diffuseness parameter at each time-frequency tile. The temporal and frequency resolution is determined by the short-time Fourier transform (STFT) or filter-bank chosen for this task. As suggested in the original SIRR publication [9], the required parameters may be estimated, although not exclusively, by using the energetic properties of the active-intensity vector [20, 24]. The active-intensity vector may be estimated using first-order spherical harmonic components, which may be derived from 3-D microphone arrays of four or more sensors [4]. Since the active-intensity vector points in the direction of the flow of acoustical energy, an assumption is made that the opposite direction corresponds to the source DoA, while the diffuseness parameter is derived as a single-channel Wiener filter on the diffuse-energy [25]. SIRR then synthesises the loudspeaker IRs by panning the omni-directional component to the analysed DoA using vector-base amplitude-panning (VBAP) [26], and replicating the omni-directional component to all loudspeaker channels, followed by decorrelation. The balance between these direct and diffuse streams is dictated by the temporal and frequency-dependent diffuseness parameter.

While the SIRR method has been shown to be perceptually similar to reference cases in listening tests [10], its sound-field model is limited in two respects. Firstly, only one DoA estimate is extracted per time-frequency tile, which is problematic if multiple reflections land in the same time-frequency tile. Secondly, the original formulation also forces an isotropic diffuse-field, which may not correspond to that of the original space. This latter limitation may, however, be minimised by employing cardioid beampatterns for each loudspeaker direction (akin to ambisonic decoding), rather than replicating the omni-directional component for each loudspeaker channel; this was later suggested in [25]. However, the diffuse stream is still modulated with direction-independent diffuseness values.

1.2 Spatial Decomposition Method (SDM)

Another example of a parametric reproduction method, intended for microphone array IRs, is the Spatial Decomposition Method (SDM) [11], which was developed to overcome SIRR input limitations by employing the microphone array signals directly. While the SHD is often considered a more convenient format, some loss of spatial performance is incurred during the conversion. The SDM approach operates based on the assumption that the sound-field is composed of a single time-varying broad-band source. The temporal resolution of the method is determined by the moving windowing function and its chosen hop-size. In practice, and as suggested by the authors, an open spherical microphone array of four or more omni-directional sensors may be employed. The DoA may then be estimated via the cross-correlation between the microphone array channels, using very short windows of the order of approximately 1 ms. This correlation-based analysis principle is the same as the one proposed earlier in [20]. The authors of SDM also later developed a first-order SHD variant (SDM-B-Format) [27], which operates based on the broadband intensity vector in the same manner as SIRR; albeit, without frequency resolution or dedicated diffuse stream rendering. During the synthesis stage, the omni-directional component (sum of all the sensor signals) is either quantised to the nearest loudspeaker or panned using VBAP or Ambisonics. Since SDM was originally intended for concert hall auralisation, quantisation is generally more preferable as timbral colouration is minimised. Note that while the method employs no dedicated diffuse-stream, it makes the assumption that the DoA will fluctuate randomly in all directions during the late part of the response, thus resulting in a diffuse tail.

One of the main limitations of SDM, however, is that low-frequencies, with wavelengths longer than that of the analysis window length, are effectively divided into shorter temporal components and then panned using broad-band DoA estimates. This processing introduces time-varying distortions and colouration of the loudspeaker output. However, as suggested by the authors, this may be mitigated by applying an adaptive post-filter on the outputs, in order to bring its spectra closer to that of the input. In [11], listening tests were conducted, which compared SDM with their own implementation of the SIRR method. The authors simulated a six sensor array for the SDM test cases, and applied additional spatial encoding to obtain first-order spherical harmonic components for the SIRR test cases. The results indicated that SDM yielded a closer *similarity* to the reference test cases, when compared to SIRR. However, it should be noted that open arrays with omni-directional sensors are not well suited to this conversion [28]. The authors also did not detail how they conducted this conversion.

1.3 Parametric methods for continuous microphone array input

While parametric methods intended for microphone array IRs are few in number, there are many for the task of reproducing continuous microphone array input. For example, the sound-field model employed in SIRR was later used in the first formulation of Directional Audio Coding (DirAC) [25], which operated on first-order spherical harmonic *signals* as input. Later formulations of DirAC also accommodate higher-order input signals [29, 30], and operate by segregating the sound-field into several spatially-selective sectors. This partitioning of the sound-field allows DirAC to subsequently determine multiple DoA and diffuseness parameters at each time-frequency tile; which has been shown to improve the perceived spatial accuracy for challenging sound scenes [29, 30]. High-Angular Plane-wave Expansion (HARPEX) [31] is another example of a parametric method, which is also intended for continuous spherical harmonic signals as input. It employs a sound-field model that assumes two plane-wave components in each time-frequency tile and no diffuse rendering. The Sparse-Recovery method, proposed in [32], employs a sound-field model consisting of a number of source components, up to half the number of input channels, using arbitrary order input. More recently, the Coding and Multi-Parameterisation of Ambisonic Sound Scenes (COMPASS) method was proposed in [33], which also employs a sound-field model consisting of multiple source components, up to half the number of input channels, but also includes an additional residual component. This residual component is obtained by subtracting the source components from the input sound-field, which therefore encapsulates all of the remaining ambient and diffuse components. An ambisonic decoder is then employed to reproduce the ambient stream.

1.4 Motivation for this work

In recent years, the realm of parametric rendering has seen far greater advances with regard to the reproduction of continuous microphone array signals, compared with IRs. Indeed, existing parametric methods for IRs, such as SIRR and SDM, do not take advantage of many of the advancements that their continuous signal counterparts employ today. Therefore, this paper proposes a higher-order formulation of the SIRR method (HO-SIRR). Much like recent formulations of the DirAC method [29, 30], HO-SIRR extracts multiple DoA estimates per time-frequency tile by segregating the sound-field into individual sectors; provided higher-order input signals are available. The spatially biased components are then panned to their respective analysed directions using VBAP. Furthermore, unlike the diffuse stream rendering conducted in legacy SIRR, the proposed formulation first scales the sector signals with the corresponding diffuseness values and encodes them back into spherical harmonics. This SHD diffuse stream is then decoded to the loudspeaker set-up using an ambisonic decoder, followed by decorrelation of the loudspeaker channels; an approach reminiscent of the ambient rendering conducted by the COMPASS method [33]. The proposed formulation is intended to be more robust in cases where multiple directional components arrive in the same time-frequency tile, and to also reproduce a non-isotropic diffuse-field more faithfully. Visual depictions of energy spectrograms using different parametric methods, alongside a reference case, are also presented, which appear to validate these objectives. Furthermore, an open-source implementation of the proposed method is detailed.

2 LEGACY SPATIAL IMPULSE RESPONSE RENDERING (SIRR)

The original SIRR formulation [9, 10] operated on first-order, $N = 1$, spherical harmonic signals as input $\mathbf{a_1}(t,f) = [\mathrm{a}_{00}(t,f), \mathrm{a}_{1(-1)}(t,f), \mathrm{a}_{10}(t,f), \mathrm{a}_{11}(t,f)]^{\mathrm{T}}$, in the time-frequency domain; where t and f refer to the time and frequency indices, respectively. A parameter vector, $\mathbf{p}_1 = \mathscr{A}[\mathbf{a_1}] = [\theta, \phi, \psi]$, is then estimated for each time-frequency tile, which consists of the azimuth-elevation angles, θ, ϕ, and the diffuseness, ψ. As suggested in the original publication [9], these parameters may be extracted via the energetic properties of the active-intensity vector [20, 24], which can be derived from the zeroth and first-order spherical harmonic components (note that the time and frequency indices are henceforth omitted for the brevity of notation)

$$\mathbf{i}_{\mathrm{a}} = -\Re[p\mathbf{u}^*], \quad \text{with}^1 \quad \mathbf{u} \simeq -\frac{1}{\rho_0 c\sqrt{3}} \begin{bmatrix} \mathrm{a}_{11} \\ \mathrm{a}_{1(-1)} \\ \mathrm{a}_{10}, \end{bmatrix}, \quad \text{and} \quad p \simeq \mathrm{a}_{00} \tag{1}$$

where $\Re$ denotes the real operator, p is the sound pressure, ρ_0 is the mean density of the medium, c is the speed of sound, and $\mathbf{u}$ is the particle velocity. The particle velocity may be estimated in this manner, with the assumption that the sound sources are received as plane-waves [34]. The parameters may then be estimated as

$$\theta, \phi = \angle - \frac{\mathbf{i}_{\mathrm{a}}}{||\mathbf{i}_{\mathrm{a}}||}, \quad \text{and} \quad \psi = 1 - \frac{2||\mathbf{i}_{\mathrm{a}}||}{|p|^2 + \mathbf{u}^{\mathrm{H}}\mathbf{u}}. \tag{2}$$

[1] Assuming ortho-normalised (N3D) real SHs with ACN indexing. Omit the $1/\sqrt{3}$ term if using the semi-normalised (SN3D) convention.

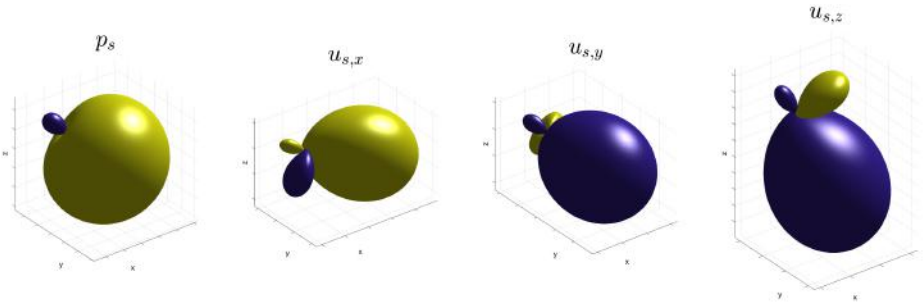


Figure 1. Example directivity patterns of sector components, employing second-order spherical harmonics and a hyper-cardioid beamformer with maxRE weighting [2].

The direct, $\mathbf{y}_{\text{dir}} \in \mathbb{R}^{L\times 1}$, and diffuse, $\mathbf{y}_{\text{diff}} \in \mathbb{R}^{L\times 1}$, loudspeaker IRs are then independently synthesised, for each time-frequency tile, by employing the analysed parameters as

$$\mathbf{y}_{\text{dir}} = \sqrt{1-\psi}\ \mathbf{g}(\theta,\phi)\mathrm{a}_{00}, \quad \text{and} \quad \mathbf{y}_{\text{diff}} = \sqrt{\frac{\psi}{L}}\ \mathscr{D}\left[\mathbf{1}_L \mathrm{a}_{00}\right], \tag{3}$$

where L is the number of loudspeakers; $\mathbf{g}(\theta,\phi) \in \mathbb{R}^{L\times 1}$ are the VBAP gains corresponding to the estimated DoA; $\mathbf{1}_L \in \mathbb{R}^{L\times 1}$ is a vector of ones to replicate the omni-directional component, a_{00}, to each loudspeaker channel; and $\mathscr{D}[.]$ denotes a de-correlation operation on the enclosed signals. The final time-domain loudspeaker IRs may be acquired by summing the two streams, followed by an appropriate inverse time-frequency transform.

3 PROPOSED HIGHER-ORDER FORMULATION (HO-SIRR)

The original SIRR formulation has been shown to perform well in listening tests, under certain conditions [10]. However, the method is limited in two key respects: firstly, the method can only estimate one DoA at a time, which is problematic if multiple reflections land in the same time-frequency tile; and secondly, the method forces the diffuse stream to be isotropic, which may not correspond to the true input diffuse properties. Therefore, the proposed HO-SIRR formulation incorporates support for higher-order input, to allow for multiple simultaneous DoA estimates, and conducts the diffuse stream rendering in a manner that can more faithfully reproduce a non-isotropic diffuse-field.

In order to conduct the higher-order analysis, $N > 1$, the sound-field is first weighted multiple times, with each weighted variant favouring energetic contributions arriving in one specific direction. This operation is similar to the higher-order analysis employed by DirAC for reproduction purposes [29], and to the directional re-assignment of beamformer energy approach for visualisation purposes [35]. This spatial weighting allows for the estimation of multiple spatially-localised active-intensity vectors

$$\mathbf{i}_{\mathrm{a},s} = \Re[p_s \mathbf{u}_s^*], \quad \text{with} \quad \begin{bmatrix} p_s \\ \mathbf{u}_s \end{bmatrix} = \mathbf{W}_s\ \mathbf{a}_N. \tag{4}$$

where $\mathbf{W}_s \in \mathbb{R}^{4\times(N+1)^2}$ is a matrix of beamforming weights, which provide the spatially weighted pressure and particle velocity components for sector s. Note that the sector directions should, ideally, uniformly sample the sphere, and higher input orders allow the sound-field to be more finely segregated. Example sector patterns are depicted in Fig. 1. For information regarding how to compute these beamforming weights, the reader is referred to [36].

A higher-order parameter vector is then obtained by extracting the SIRR parameters for all S sectors

$$\mathbf{p}_N = \mathscr{A}\left[\mathbf{a}_N\right] = [\theta_1, \phi_1, \psi_1, ..., \theta_S, \phi_S, \psi_S]. \tag{5}$$

If higher-order components are not available, then the analysis reverts back to first-order SIRR analysis.

The sector signals are then panned using VBAP to the loudspeaker channels and accumulated, which yields the direct stream as

$$\mathbf{y}_{\text{dir}} = \sum_{s=1}^{S} \sqrt{\frac{1-\psi_s}{S}}\ \mathbf{g}(\theta_s, \phi_s)\ p_s. \tag{6}$$

For the diffuse stream, when employing higher-order input, the sector signals are first scaled with the corresponding diffuseness values and re-encoded back into the SHD. Whereas, when using first-order input, the signals are simply scaled with the diffuseness value

$$\mathbf{a}_{\text{diff}} = \left\{ \begin{array}{ll} \sqrt{\frac{\psi_1}{(N+1)^2}}\ \mathbf{a}_1, & \text{for } N = 1 \\ \sum_{s=1}^{S} \sqrt{\frac{\psi_s}{S}}\ \mathbf{y}_N^{(s)} p_s, & \text{for } N > 1 \end{array} \right\}, \tag{7}$$

where $\mathbf{y}_N^{(s)}$ are the spherical harmonic weights for the corresponding sector direction, and $\mathbf{a}_{\text{diff}}$ is the diffuse stream expressed in the SHD. The loudspeaker diffuse stream, $\mathbf{y}_{\text{diff}}$, is then obtained via ambisonic decoding, followed by a decorrelation operation on the loudspeaker channels

$$\mathbf{y}_{\text{diff}} = \mathscr{D}\left[\mathbf{D}_L\ \mathbf{a}_{\text{diff}}\right], \tag{8}$$

where $\mathbf{D}_L \in \mathbb{R}^{L\times(N+1)^2}$ is an ambisonic decoding matrix of real-valued gains, which can be calculated using one of a number of methods [1, 2, 3]. Note that when $N > 1$, the method renders the diffuse components in a non-isotropic manner. Furthermore, contrary to the legacy SIRR approach, the proposed formulation still employs an ambisonic decoder when $N = 1$, rather than replicating the omni to the loudspeaker channels. This is generally preferable, as it allows for some degree of spatial separation across the loudspeaker channels, which has been found to sound more *natural* in [25].

3.1 Open-source Implementation

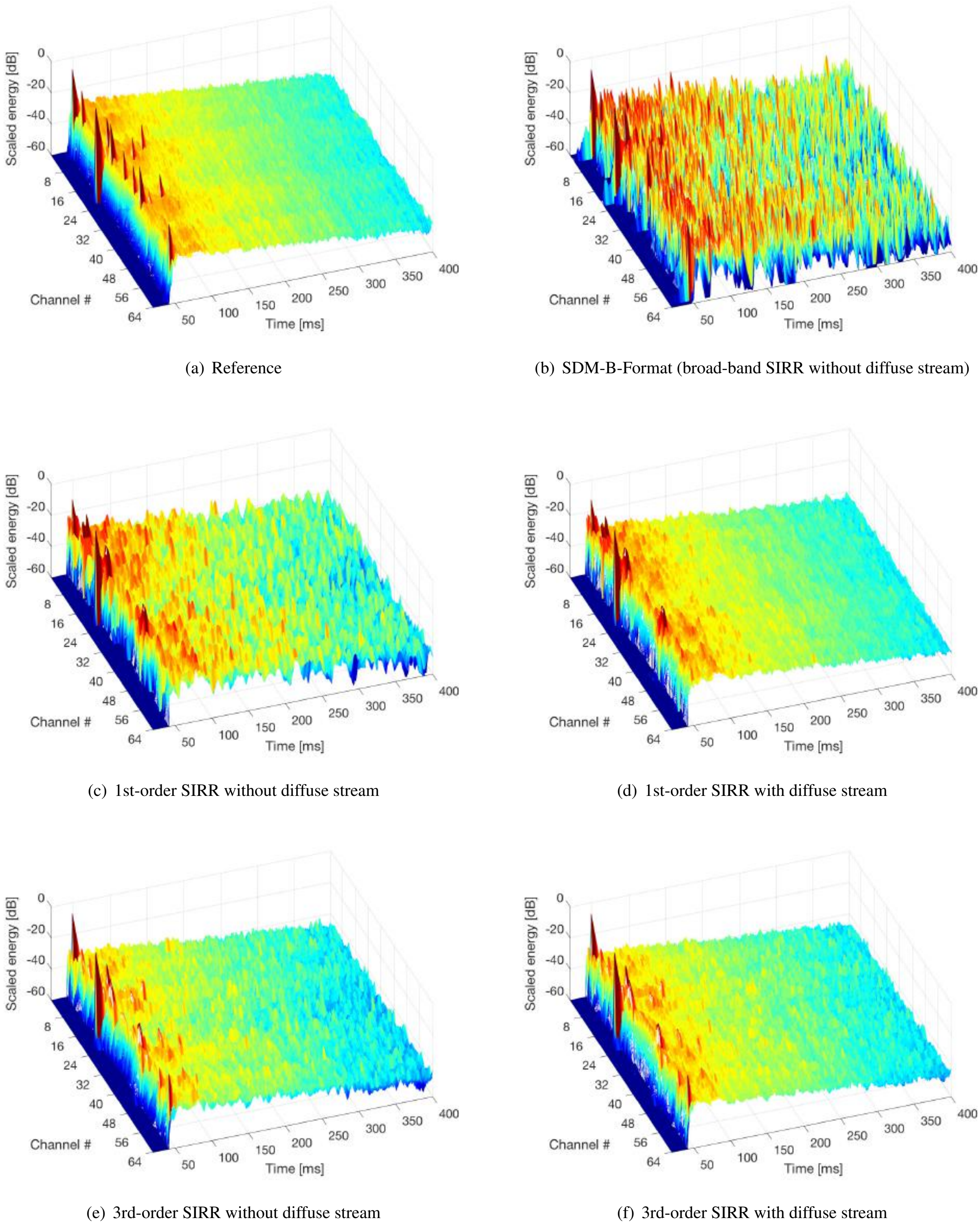
An open-source MATLAB implementation[2] of the proposed method has also been developed. By default, the method employs a frequency-resolution of 128 uniformly-spaced bins and 50% window overlap. However, the implementation also supports multi-resolution STFT processing, allowing the user to dictate different temporal and frequency resolution trade-offs for different frequency ranges. Furthermore, an option to isolate the direct peak and pan it using a broadband DoA estimate is included, which may reduce timbral colouration in some cases.

4 EXAMPLE ENERGY SPECTROGRAMS

To gain some insight into the performance of the different parametric methods, acoustical simulation tools were used to create reference impulse responses for a 64-channel spherical loudspeaker array. Spherical harmonic IRs were then synthesised by applying the appropriate direction-dependent weights for all rays that arrived at the listening position in the simulation. These spherical harmonic IRs were then passed through different rendering methods, in order to generate loudspeaker IRs for comparison with the reference. The SDM-B-Format implementation by the original authors was employed [27], alongside the open-source implementation of the proposed HO-SIRR method; the latter was configured to utilise a 128 uniform frequency resolution and a 50% temporal overlap.

To illustrate the differences between SDM-B-Format, SIRR and HO-SIRR (third-order), energy spectrograms of the rendered and reference 64-channel IRs are depicted in Fig. 2. The energy spectrograms were low-pass filtered using a first-order IIR filter with a cut-off frequency of 100 Hz, in order to improve graphical clarity. The reference case is shown in Fig. 2(a), where the direct path and early reflections are visible as distinct peaks, and the late reverberation exhibits a relatively smooth downward trending surface. The SDM-B-Format case is shown in Fig. 2(b), where it can be observed that the early peaks are similar to that of the original response. However, the late reverberation consists of a multitude of sharp peaks which are not present in the reference case. The first-order SIRR energy spectrogram with the diffuse-stream disabled, exhibits a similar early response to the SDM-B-Format case, as shown in Fig. 2(c). However, the introduction of bin-wise processing and/or larger temporal windows, appears to yield a smoother late response, but is still quite sporadic when compared to the reference. The inclusion of the diffuse stream, however, clearly improves the surface of the late energy decay response, as shown in Fig. 2(d). The HO-SIRR test case *without* a diffuse-stream, Fig. 2(e), appears to be more stable in the early response than that of its first-order counterpart, which may indicate that the sector-based directional analysis is more robust. Finally, HO-SIRR *with* the dedicated diffuse stream, as shown in Fig. 2(f), appears to yield an energy spectrogram that most closely resembles the reference. However, it is important to note that these figures are provided only to give a rough idea of the differences between the methods. The perceptual consequences of these observable differences have to be investigated with formal listening tests, which is a topic of future work.

[2] The MATLAB source code (GPLv3 license) of the proposed method is available here: https://github.com/leomccormack/HO-SIRR



(a) Reference

(b) SDM-B-Format (broad-band SIRR without diffuse stream)

(c) 1st-order SIRR without diffuse stream

(d) 1st-order SIRR with diffuse stream

(e) 3rd-order SIRR without diffuse stream

(f) 3rd-order SIRR with diffuse stream

Figure 2. Energy spectrograms of impulse responses rendered for a 64-channel loudspeaker array. The default SDM-B-Format implementation by the original authors [27] and the open-source implementation of HO-SIRR were employed for the renderings.

5 CONCLUSION

This paper has proposed a new formulation of the Spatial Impulse Response Rendering (SIRR) method, which can accommodate higher-order input, thus allowing it to address many of the shortcomings of the original formulation. The proposed method is able to analyse multiple simultaneous directional components, by segregating the sound-field into individual sectors and extracting parameters for each. The higher-order processing also allows for a directional diffuse stream, which is contrary to the majority of parametric methods proposed to date. Informal visual inspection of energy spectrograms indicate that the higher-order analysis and a dedicated diffuse-stream, appear to yield rendering outputs which more closely resemble the reference; when compared to the original SIRR formulation and the Spatial Decomposition Method (SDM). A more comprehensive evaluation of the method is a topic of future work.

ACKNOWLEDGEMENTS

This research has received funding from the Aalto ELEC Doctoral School and the Academy of Finland project no 317341.

References

[1] M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[2] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *Journal of the audio engineering society*, vol. 60, no. 10, pp. 807–820, 2012.

[3] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 37–47, 2012.

[4] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015, vol. 8.

[5] O. Santala, H. Vertanen, J. Pekonen, J. Oksanen, and V. Pulkki, "Effect of listening room on audio quality in Ambisonics reproduction," in *126th Convention of the AES*, Munich, Germany, 2009.

[6] S. Braun and M. Frank, "Localization of 3D ambisonic recordings and ambisonic virtual sources," in *1st Int. Conf. on Spatial Audio (ICSA)*, Detmold, Germany, 2011.

[7] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order Ambisonics reproduced sound sources," *Acta Acustica united with Acustica*, vol. 99, no. 4, pp. 642–657, 2013.

[8] P. Stitt, S. Bertet, and M. van Walstijn, "Off-centre localisation performance of Ambisonics and HOA for large and small loudspeaker array radii," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 937–944, 2014.

[9] J. Merimaa and V. Pulkki, "Spatial impulse response rendering i: Analysis and synthesis," *Journal of the Audio Engineering Society*, vol. 53, no. 12, pp. 1115–1127, 2005.

[10] V. Pulkki and J. Merimaa, "Spatial impulse response rendering ii: Reproduction of diffuse sound and listening tests," *Journal of the Audio Engineering Society*, vol. 54, no. 1/2, pp. 3–20, 2006.

[11] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, 2013.

[12] M. Frank and F. Zotter, "Spatial impression and directional resolution in the reproduction of reverberation," *Fortschritte der Akustik, DAGA*, 2016.

[13] P. Coleman, A. Franck, P. Jackson, R. Hughes, L. Remaggi, and F. Melchior, "Object-based reverberation for spatial audio," *Journal of the Audio Engineering Society*, vol. 65, no. 1/2, pp. 66–77, 2017.

[14] M. Zaunschirm, M. Frank, and F. Zotter, "Brir synthesis using first-order microphone arrays," in *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018.

[15] J. Ahrens, "Auralization of omnidirectional room impulse responses based on the spatial decomposition method and synthetic spatial data," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 146–150.

[16] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[17] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.

[18] M. A. Akeroyd and L. R. Bernstein, "The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses," *The Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2516–2526, 2001.

[19] B. C. Moore, *An introduction to the psychology of hearing*. Brill, 2012.

[20] Y. Yamasaki and T. Itow, "Measurement of spatial information in sound fields by closely located four point microphone method," *Journal of the Acoustical Society of Japan (E)*, vol. 10, no. 2, pp. 101–110, 1989.

[21] G. Schiffrer and D. Stanzial, "Energetic properties of acoustic fields," *The Journal of the Acoustical Society of America*, vol. 96, no. 6, pp. 3645–3653, 1994.

[22] D. Stanzial, N. Prodi, and G. Schiffrer, "Reactive acoustic intensity for general fields and energy polarization," *The Journal of the Acoustical Society of America*, vol. 99, no. 4, pp. 1868–1876, 1996.

[23] M. Karjalainen, J. Merimaa, T. Peltonen, and T. Lokki, "Measurement, analysis, and visualization of directional room responses," in *Audio Engineering Society Convention 111*. Audio Engineering Society, 2001.

[24] F. J. Fahy and V. Salmon, "Sound intensity," *The Journal of the Acoustical Society of America*, vol. 88, no. 4, pp. 2044–2045, 1990.

[25] V. Pulkki, "Spatial sound reproduction with Directional Audio Coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[26] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the audio engineering society*, vol. 45, no. 6, pp. 456–466, 1997.

[27] "SDM matlab toolbox," https://se.mathworks.com/matlabcentral/fileexchange/56663-sdm-toolbox, accessed: 2019-05-09.

[28] S. Moreau, J. Daniel, and S. Bertet, "3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, 2006, pp. 20–23.

[29] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852–866, 2015.

[30] A. Politis, L. McCormack, and V. Pulkki, "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,*, 2017.

[31] S. Berge and N. Barrett, "A new method for B-format to binaural transcoding," in *40th Int. Conf. of AES*, Tokyo, Japan, 2010.

[32] A. Wabnitz, N. Epain, A. McEwan, and C. Jin, "Upscaling ambisonic sound scenes using compressed sensing techniques," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2011, pp. 1–4.

[33] 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)*. IEEE, 2018, pp. 6802–6806.

[34] H. Hacihabiboglu, "Theoretical analysis of open spherical microphone arrays for acoustic intensity measurements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 465–476, 2014.

[35] L. McCormack, A. Politis, and V. Pulkki, "Sharpening of angular spectra based on a directional reassignment approach for ambisonic sound-field visualisation," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 576–580.

[36] A. Politis, "Diffuse-field coherence of sensors with arbitrary directional responses," *arXiv preprint arXiv:1608.07713*, 2016.