CHALMERS
UNIVERSITY OF TECHNOLOGY

Spatial Subtraction of Reflections from Room Impulse Responses Measured With a Spherical Microphone Array

SPATIAL SUBTRACTION OF REFLECTIONS FROM ROOM IMPULSE RESPONSES MEASURED WITH A SPHERICAL MICROPHONE ARRAY

Thomas Deppisch[*]*, Jens Ahrens*[*]

Chalmers University of Technology,
412 96 Gothenburg, Sweden
{thomas.deppisch, jens.ahrens}@chalmers.se

Sebastià V. Amengual Garí, Paul Calamia

Facebook Reality Labs Research, Facebook,
1 Hacker Way, Menlo Park, CA 94025
{samengual, pcalamia}@fb.com

ABSTRACT

We propose a method for the decomposition of measured directional room impulse responses (DRIRs) into prominent reflections and a residual. The method comprises obtaining a fingerprint of the time-frequency signal that a given reflection carries, imposing this time-frequency fingerprint on a plane-wave prototype that exhibits the same propagation direction as the reflection, and finally subtracting this plane-wave prototype from the DRIR. Our main contributions are the formulation of the problem as a spatial subtraction as well as the incorporation of order truncation, spatial aliasing and regularization of the radial filters into the definition of the underlying beamforming problem. We demonstrate, based on simulated as well as measured array impulse responses, that our method increases the accuracy of the model of the reflection under test and consequently decreases the energy of the residual that remains in a measured DRIR after the spatial subtraction.

Index Terms— Beamforming, Spherical Harmonics, Spherical Microphone Arrays

1. INTRODUCTION

Parametric artificial reverberation offers flexibility and allows for modification but cannot easily be adapted to resemble the reverberation that would be measured in a real acoustic space. Even state-of-the-art room acoustic simulations fail to render complex acoustic scenarios that are indistinguishable from a reference measurement with the main perceived differences lying in timbre and source position [1]. This contribution tackles the problem of the virtualization of real-world acoustic spaces from a different angle: by facilitating the separation of prominent early reflections from the diffuse reverberation, it functions as a first step toward the parametric modification of measured directional room impulse responses (DRIRs).

Several methods for the parametric processing and rendering of DRIRs have been proposed [2–8]. While many of them explicitly distinguish between a non-diffuse and a diffuse part of the DRIR in their processing, none of the methods achieves a general detection and separation of individual reflections from the DRIR. A separation of directional and diffuse signal components of sound scenes is targeted in Coding and Multidirectional Parametrization of Ambisonic Sound Scenes (COMPASS) [9]. COMPASS decomposes the sound scene into source signals and an ambient signal using dedicated source and ambience beamformers before it pans the source signals and decorrelates the ambience signal for rendering. Our proposed method extends the source and ambience beamformers of COMPASS by taking into account spatial aliasing, order truncation, and scattering off the surface of a spherical microphone array (SMA). Our formulation is a general extension of such beamformers and is therefore also applicable to sound scenes as long as the assumption of a given impinging wave front being planar is met sufficiently well. Prominent reflections in RIRs are assumed to satisfy the plane-wave assumption well and thus are our first choice to employ the proposed method. As is commonly done with signals captured by SMAs, the method operates in the spherical harmonic (SH) domain.

[*]We thank Facebook Reality Labs Research for funding this project.

2. PRESSURE ON THE SURFACE OF A RIGID SPHERE DUE TO AN INCIDENT PLANE WAVE

The sound pressure $p(\omega, \Theta)$ on the surface of a rigid sphere due to a plane wave impinging from direction Θ_0 and carrying the signal $s(\omega)$ is described by [10, Sec. 2.6]

$$p(\omega, \Theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\Theta) b_n(\omega) Y_n^m(\Theta_0) s(\omega)\,, \qquad (1)$$

where $Y_n^m(\Theta)$ are the SHs of order n and degree m that depend on the spherical coordinate $\Theta = (\varphi, \vartheta)$ composed of the azimuth angle φ and the zenith angle ϑ. For notational convenience, we assume a real-valued definition of the SHs. The superposition of the incident plane wave and the scattering off the rigid, spherical microphone body evaluated at the surface of the rigid sphere with radius r, is expressed by [11, Sec. 6.5]

$$b_n(\omega) = \frac{4\pi \mathrm{i}^{n+1}}{(kr)^2\, h_n'^{(2)}(kr)}\,, \qquad (2)$$

with the wave number $k = \omega/c$ composed of the angular frequency ω and the speed of sound c, the imaginary unit i, and the derivative of the spherical Hankel function of the second kind $h_n'^{(2)}$. Expressions for other SMA configurations are provided in [12]. By limiting the infinite sum of (1) to a finite maximum SH order $\tilde{N}$ and stacking Y_n^m and $b_n(\omega)$ in vectors of length $(\tilde{N}+1)^2$, $\boldsymbol{y}_{\tilde{N}}(\Theta) = [Y_0^0(\Theta),\, Y_1^{-1}(\Theta),\, Y_1^0(\Theta),\, \ldots,\, Y_{\tilde{N}}^{\tilde{N}}(\Theta)]^{\mathrm{T}}$, $\boldsymbol{b}_{\tilde{N}}(\omega) = [b_0(\omega),\, b_1(\omega),\, b_1(\omega),\, \ldots,\, b_{\tilde{N}}(\omega)]^{\mathrm{T}}$, (1) is re-expressed in matrix-vector notation

$$p(\omega, \Theta) = \boldsymbol{y}_{\tilde{N}}^{\mathrm{T}}(\Theta) \boldsymbol{B}_{\tilde{N}}(\omega) \boldsymbol{y}_{\tilde{N}}(\Theta_0) s(\omega)\,, \qquad (3)$$

where $\boldsymbol{B}_{\tilde{N}}(\omega) = \mathrm{diag}(\boldsymbol{b}_{\tilde{N}}(\omega))$ is a diagonal matrix with $\boldsymbol{b}_{\tilde{N}}(\omega)$ on its diagonal. Depending on the highest angular frequency of interest ω and the radius r of the SMA, the order $\tilde{N}$ can be chosen high enough such that the approximation of (3) creates a negligibly

small error. The sound pressure is probed at a set of L locations on the spherical body

$$\boldsymbol{p}(\omega) = \boldsymbol{Y}_{\tilde{N},\mathrm{M}}^{\mathrm{T}} \boldsymbol{B}_{\tilde{N}}(\omega) \boldsymbol{y}_{\tilde{N}}(\Theta_0) s(\omega)\,, \tag{4}$$

which is expressed by the multiplication with the matrix $\boldsymbol{Y}_{\tilde{N},\mathrm{M}} = [\boldsymbol{y}_{\tilde{N}}(\Theta_1), \ldots, \boldsymbol{y}_{\tilde{N}}(\Theta_L)]$ that contains the SHs up to order $\tilde{N}$, evaluated at all microphone locations. From the pressure $\boldsymbol{p}(\omega)$ at the microphone locations, an order-N SH representation

$$\boldsymbol{p}_N(\omega) = \boldsymbol{E}_N \boldsymbol{Y}_{\tilde{N},\mathrm{M}}^{\mathrm{T}} \boldsymbol{B}_{\tilde{N}}(\omega) \boldsymbol{y}_{\tilde{N}}(\Theta_0) s(\omega) \tag{5}$$

is obtained via the discrete spherical Fourier transform (DSFT) matrix $\boldsymbol{E}_N$ [10, Sec. 3.6] that in the general case employs the pseudoinverse $\boldsymbol{E}_N = (\boldsymbol{Y}_{N,\mathrm{M}}^{\mathrm{T}})^{\dagger}$, but can for some sampling configurations also utilize quadrature weights $\boldsymbol{\alpha}$, $\boldsymbol{E}_N = \boldsymbol{Y}_{N,\mathrm{M}} \mathrm{diag}(\boldsymbol{\alpha})$. Note that the maximum SH order N in the DSFT is limited by the number L of microphones in the array and their spatial distribution, and is in most practical cases smaller than the order $\tilde{N}$ for approximation of ideal plane waves. This is referred to as order truncation in the SMA literature and affects the output primarily at high frequencies. The spatial information that is extracted by the SMA is additionally corrupted at high frequencies by spatial aliasing, which is a consequence of the employment of a discrete microphone grid [13].

3. EXTRACTION OF PLANE-WAVE SIGNALS AND SPATIAL SUBTRACTION

The objective of the SH beamformer is to retrieve the plane-wave signal $s(\omega)$ (cf. (1)) from the SH domain pressure $\boldsymbol{p}_N(\omega)$ using beamformer weights $\boldsymbol{d}_N(\omega)$, such that the extracted signal

$$\gamma(\omega) = \boldsymbol{d}_N^{\mathrm{T}}(\omega)\, \boldsymbol{p}_N(\omega) \tag{6}$$

equals $s(\omega)$. This is achieved by removing the impact of the scattering via the inverse scattering matrix $\boldsymbol{B}_N^{-1}$ and normalization with reference to the amplitude caused by a plane wave

$$\boldsymbol{d}_N^{\mathrm{T}}(\omega) = \frac{\boldsymbol{y}_N^{\mathrm{T}}(\Theta_0) \tilde{\boldsymbol{B}}_N^{-1}(\omega)}{\boldsymbol{y}_N^{\mathrm{T}}(\Theta_0) \boldsymbol{y}_{N,\mathrm{pro}}(\Theta_0)}\,. \tag{7}$$

We study the influence of different assumptions on the plane wave that is also used for spatial subtraction and hence refer to it as plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro}}(\Theta_0)$. In practice, the scattering matrix $\boldsymbol{B}_N(\omega)$ has to be regularized before inversion (cf. Sec. 3.2); the regularized version is denoted as $\tilde{\boldsymbol{B}}_N(\omega)$ and often referred to as radial filter. The extracted signal $\gamma(\omega)$ characterizes a specific sound event such as a room reflection and hence is in the following referred to as time-frequency fingerprint of that event.

Note that, as in [9], this beamformer can be extended to consider multiple simultaneous directions by replacing the vectors $\boldsymbol{y}_N(\Theta_0)$ and $\boldsymbol{y}_{N,\mathrm{pro}}(\Theta_0)$ by matrices whose columns contain such vectors for multiple directions.

3.1. Ideal Case

If the SMA configuration supports an error-free DSFT up to order $\tilde{N}$, i.e. $\boldsymbol{E}_N \boldsymbol{Y}_{\tilde{N},\mathrm{M}}^{\mathrm{T}} = \boldsymbol{I}$, it does not introduce order truncation or spatial aliasing artifacts. This can be achieved in practice by distributing the microphones according to a spherical t-design [14] of degree $t \geq 2\tilde{N}$ [10, Sec. 3.4]. If additive noise is disregarded, $\boldsymbol{B}_{\tilde{N}}(\omega)$ can be inverted without regularization (except for $\omega = 0$) such that $\tilde{\boldsymbol{B}}_N^{-1}(\omega) \boldsymbol{B}_{\tilde{N}}(\omega) = \boldsymbol{I}$, and the suitable plane-wave prototype to extract the fingerprint $\gamma(\omega)$ is expressed as

$$\boldsymbol{y}_{N,\mathrm{pro},1} = \boldsymbol{y}_N(\Theta_0)\,. \tag{8}$$

A SH beamformer defined by the weights $\boldsymbol{d}_N$ from (7) using this plane-wave prototype is commonly referred to as plane-wave decomposition beamformer or maximum directivity beamformer [10, Sec. 6.1]. Note that in the beamforming literature, $\boldsymbol{p}_N(\omega)$ and $\boldsymbol{d}_N(\omega)$ are commonly defined without taking into account the regularization of $\boldsymbol{B}_N(\omega)$, the order truncation, or the spatial aliasing, and $\boldsymbol{p}_N(\omega)$ is termed the steering vector. Beamforming and spatial subtraction based on the conventional definition of the steering vector was proposed in COMPASS [9]. Our main contributions are the extended definition of the steering vector in (5) and the corresponding adaptation of the beamformer weights as well as of the spatial subtraction method in Sec. 3.2–3.4. However, for the spatial subtraction it is beneficial to define plane-wave prototypes that carry the time-frequency fingerprint $\gamma(\omega)$ instead of steering vectors.

3.2. Non-Ideal Radial Filtering

The direct inversion of $\boldsymbol{B}_N$ to remove the impact of the scattering body is problematic if the measured pressure signals contain noise, as the weights in (2) tend to zero for low frequencies. Common solutions [11, Sec. 6.8] [15–17] apply regularization to limit the strong boost of low frequencies and refer to the inverted filters as radial filters. We again assume a SMA that is able to achieve an error-free DSFT up to order $\tilde{N}$ but now include the regularized radial filter $\tilde{\boldsymbol{B}}_N^{-1}(\omega)$ into the definition of the plane-wave prototype

$$\boldsymbol{y}_{N,\mathrm{pro},2}(\omega) = \tilde{\boldsymbol{B}}_N^{-1}(\omega) \boldsymbol{B}_N(\omega)\, \boldsymbol{y}_N(\Theta_0)\,. \tag{9}$$

3.3. Order Limitation and Spatial Aliasing

In addition to the regularized radial filtering, we include the discretization of the sound pressure by the array microphones via $\boldsymbol{Y}_{\tilde{N},\mathrm{M}}^{\mathrm{T}}$ and the truncation to order N carried out by the DSFT matrix $\boldsymbol{E}_N$ into the definition of the plane-wave prototype

$$\boldsymbol{y}_{N,\mathrm{pro},3}(\omega) = \tilde{\boldsymbol{B}}_N^{-1}(\omega) \boldsymbol{E}_N \boldsymbol{Y}_{\tilde{N},\mathrm{M}}^{\mathrm{T}} \boldsymbol{B}_{\tilde{N}}(\omega)\, \boldsymbol{y}_{\tilde{N}}(\Theta_0)\,, \tag{10}$$

and by that model the prototype as being order truncated and exhibiting spatial aliasing.

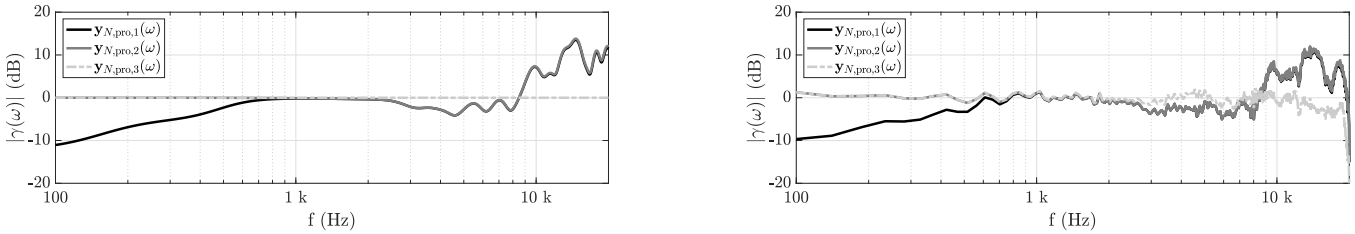
3.4. Spatial Subtraction

The spatial subtraction separates the signal components that are directly related to a single plane wave from the rest of the signal, in the following denoted as residual. To calculate the residual, the time-frequency fingerprint $\gamma(\omega)$ of the wave front under consideration is imposed on a plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro}}(\omega)$ and subtracted from the measured, spherical Fourier transformed, and radial filtered pressure signal $\mathring{\boldsymbol{p}}_N(\omega) = \tilde{\boldsymbol{B}}_N^{-1}(\omega) \boldsymbol{E}_N \mathring{\boldsymbol{p}}(\omega)$,

$$\boldsymbol{s}_{N,\mathrm{res}}(\omega) = \mathring{\boldsymbol{p}}_N(\omega) - \boldsymbol{y}_{N,\mathrm{pro}}(\omega) \gamma(\omega)\,. \tag{11}$$

Both, $\mathring{\boldsymbol{p}}(\omega)$ and $\mathring{\boldsymbol{p}}_N(\omega)$, denote measured quantities that are theoretically described by (4) and (5), respectively.

In the following, a subscript index denotes which specific plane-wave prototype is employed in the definition of the residual whenever a distinction is necessary, e.g., $\boldsymbol{s}_{N,\mathrm{res},1}(\omega)$ is defined using the plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro},1}$ from (8).



(a) Simulation.

(b) Measurement.

Figure 1: Magnitude responses $|\gamma(\omega)|$ extracted from a plane wave with a flat magnitude response by a beamformer (6) utilizing the different plane-wave prototypes $\boldsymbol{y}_{N,\mathrm{pro}}(\omega)$ from (8)–(10).

4. EVALUATION USING SMA IMPULSE RESPONSES

Before the method is applied to a DRIR, we investigate the effectiveness using measured free-field SMA impulse responses. In particular, we compare the magnitude responses $|\gamma(\omega)|$ extracted by the beamformer (6), using the three different plane-wave prototypes $\boldsymbol{y}_{N,\mathrm{pro}}(\omega)$ of (8)–(10) in the definition of its weights $\boldsymbol{d}_N(\omega)$. Then, we perform the spatial subtraction and investigate the SH vector norm of the residuals to compare the effect of the three different plane-wave prototypes. All simulation and measurement data in this contribution use $N = 32$, a DSFT matrix employing the pseudoinverse $\boldsymbol{E}_N = (\boldsymbol{Y}_{N,\mathrm{M}}^{\mathrm{T}})^{\dagger}$, and Tikhonov-regularized radial filters [15]

$$\tilde{b}_n^{-1}(\omega) = \frac{b_n^*(\omega)}{|b_n(\omega)|^2 + 0.01}\,, \tag{12}$$

that are collected on the diagonal of $\tilde{\boldsymbol{B}}_N^{-1}(\omega)$ whereby $(\cdot)^*$ denotes the complex conjugate.

4.1. Evaluation of the Beamformer

We evaluate the beamformer (6) that extracts the time-frequency fingerprint $\gamma(\omega)$ of an impinging wave front. A rigid-sphere array with $L = 26$ microphones distributed according to a Lebedev grid with radius $r = 8.5$ cm was simulated, enabling a DSFT up to order $N = 3$. Fig. 1a depicts $|\gamma(\omega)|$ assuming an ideal plane-wave signal with flat magnitude response as described by (5) for the three different plane-wave prototypes. Due to the regularized radial filtering that is not included in the definition of the plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro},1}$, the corresponding magnitude response deviates from the flat response below 700 Hz. For high frequencies above 2.5 kHz, the magnitude responses that are extracted using $\boldsymbol{y}_{N,\mathrm{pro},1}$ and $\boldsymbol{y}_{N,\mathrm{pro},2}(\omega)$ deviate from the flat response as they do not take into account the order truncation and the spatial aliasing. Only the beamformer using the plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro},3}(\omega)$ is able to extract the flat magnitude response over the full frequency range.

To evaluate the accuracy of the underlying model assumptions, impulse responses of several SMAs were measured in the anechoic chamber of the Division of Applied Acoustics at Chalmers University of Technology and are made available online [18]. In the following, we again consider the same SMA with $L = 26$ microphones distributed according to a Lebedev grid. The SMA was emulated using the VariSphear [19] sequential measurement system, which comprises a single DPA 4060 pressure microphone and a wooden rigid sphere of radius $r = 8.5$ cm as scattering body. The distance between the emulated SMA and the employed Genelec 8030A loudspeaker was 2.15 m. Additionally, the impulse response between the loudspeaker and the microphone without the scattering body was measured and compensated for by deconvolution such that $|\gamma(\omega)|$ may be considered constant. Fig. 1b shows $|\gamma(\omega)|$ based on the measured impulse responses. The simulation of Fig. 1a and the measurement of Fig. 1b are nearly identical: in contrast to $\boldsymbol{y}_{N,\mathrm{pro},1}$ and $\boldsymbol{y}_{N,\mathrm{pro},2}(\omega)$, the beamformer based on the plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro},3}(\omega)$ is able to extract a nearly flat magnitude response. Only for very high frequencies close to 20 kHz a significant roll-off is visible that is caused by a band limitation in the measurement processing chain.

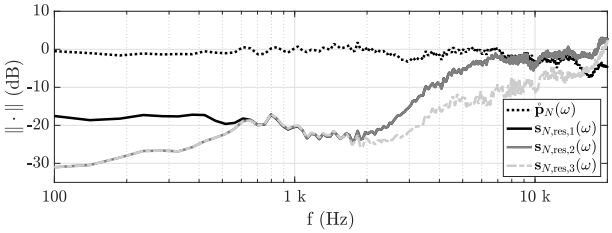


Figure 2: The vector norm of the measured SH sound pressure $\mathring{\boldsymbol{p}}_N(\omega)$ of a free-field SMA impulse response is reduced over a wide frequency range after spatial subtraction. The magnitude of the reduction depends on the employed plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro}}(\omega)$ from (8)–(10).

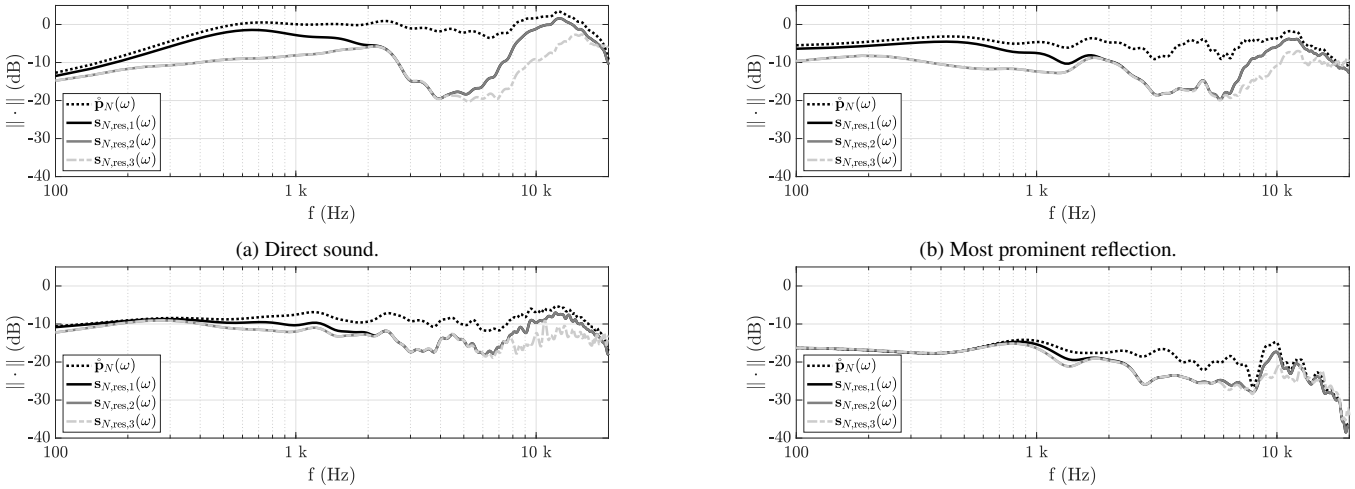
4.2. Evaluation of the Spatial Subtraction

The effectiveness of the spatial subtraction is analyzed via the norm of the residual $\|\boldsymbol{s}_{N,\mathrm{res}}(\omega)\|$ obtained after subtraction of the plane-wave prototype from the measured, spherical Fourier transformed, and radial filtered DRIR $\mathring{\boldsymbol{p}}_N(\omega)$. According to Parseval's theorem for the spherical Fourier transform [13]

$$\|\boldsymbol{s}_{N,\mathrm{res}}(\omega)\|^2 = \int_{\Theta\in\mathcal{S}^2} |s_{\mathrm{res}}(\omega,\Theta)|^2 d\Theta\,, \tag{13}$$

the squared SH vector norm equals the total spatial energy over the whole unit sphere $\mathcal{S}^2$ of the residual in space domain $s_{\mathrm{res}}(\omega,\Theta)$, which makes the SH vector norm a suitable measure. The subtraction of a plane-wave prototype that is an exact copy of the impinging wave front leads to a vector norm of $-\infty$ dB.

Fig. 2 depicts the results for the three plane-wave prototypes. The measured free-field SMA impulse responses are the same as in Sec. 4.1. All three plane-wave prototypes achieve a substantial subtraction of energy for frequencies below 7 kHz. Between 600 Hz and 2 kHz all three variants show the same subtraction performance as the impact of the regularization of the employed radial filters as well as of the order truncation and the spatial aliasing are negligible at these frequencies. The incorporation of the regularized



(a) Direct sound.

(b) Most prominent reflection.

(c) Second-most prominent reflection.

(d) Least prominent of the extracted reflections.

Figure 3: The vector norm of the measured SH sound pressure $\mathring{\boldsymbol{p}}_N(\omega)$ of the direct sound and prominent reflections in a DRIR is reduced in the residual $\boldsymbol{s}_{N,\mathrm{res}}(\omega)$ obtained from spatial subtraction. The magnitude of the reduction depends on the employed plane-wave prototype $\boldsymbol{y}_{N,\mathrm{pro}}(\omega)$ from (8)–(10) and the prominence of the reflection.

radial filter definition into the prototype improves the subtraction in $\boldsymbol{s}_{N,\mathrm{res},2}(\omega)$ and $\boldsymbol{s}_{N,\mathrm{res},3}(\omega)$ below 600 Hz. Between 2–16 kHz, the norm of the residual $\boldsymbol{s}_{N,\mathrm{res},3}(\omega)$ is up to 13 dB smaller than the norm of the other residuals. Very close to 20 kHz, the subtraction is not successful which is again caused by the band limitation in the processing chain of the measurement.

5. CASE STUDY: SPATIAL SUBTRACTION OF REFLECTIONS IN A DRIR

To evaluate the method in a more realistic scenario where the directions of arrival (DOAs) Θ_0 of the impinging wave fronts are not known precisely and possibly several wave fronts overlap in time, the spatial subtraction method is applied to a DRIR measured in a 10.3 x 5.8 x 3.1 m conference room [20]. The DRIR was measured with an Eigenmike em32 microphone array ($N = 4$) and a Genelec 1030A loudspeaker at a distance of 2.5 m. In particular, this study uses measurement position 2 of the data set, where a table is located between the source and the receiver. To detect the most prominent directional peaks in the DRIR, the peak finding algorithm from [21, 22] was implemented. The algorithm analyzes the averaged magnitude of the pseudo-intensity vector $\boldsymbol{i}(t) = w(t)\big[x(t),\, y(t),\, z(t)\big]^{\mathrm{T}}$, that is obtained from the zeroth and first-order SH components of the DRIR, i.e., the first four channels $w(t)$, $y(t)$, $z(t)$, $x(t)$, if ordered according to the Ambisonic channel number (ACN) sequence [23]. It then finds prominent peaks of the averaged pseudo-intensity vector magnitude above a prominence threshold that is defined using the lower quartile of the magnitude and exponential slopes decaying from the individual peaks. We use the onsets of the exponential slopes to define the borders of rectangular windows for the isolation of the individual peaks. Using this algorithm, the 9 most prominent peaks, including the direct sound and 8 reflections that arrive within 31 ms after the direct sound, were isolated. The corresponding rectangular windows have lengths between 1.6–3.1 ms. Subsequently, SH-MUSIC [24] was used to estimate the DOAs of the 9 most prominent peaks. As in Sec. 4.2, the proposed spatial subtraction method is evaluated by analysis of the SH vector norm of the DRIR before and after the subtraction.

Fig. 3 shows the SH vector norm of the residual obtained for the direct sound, the two most prominent and the least prominent of the detected reflections. These four given examples also resemble the results obtained for the other five detected reflections: for frequencies below 2 kHz, the spatial subtraction using the two proposed extended definitions of the plane-wave prototypes $\boldsymbol{y}_{N,\mathrm{proto},2}(\omega)$ and $\boldsymbol{y}_{N,\mathrm{proto},3}(\omega)$ outperform the conventional method using $\boldsymbol{y}_{N,\mathrm{proto},1}$ in most cases by achieving a more significant reduction of the SH vector norm of the corresponding residual. For high frequencies above 4 kHz, the spatial subtraction based on $\boldsymbol{y}_{N,\mathrm{proto},3}(\omega)$ outperforms the other two versions due to consideration of the order truncation and the spatial aliasing. The magnitude of the improvement depends on the prominence of the reflection. For less prominent reflections, the differences in performance are smaller. The explanation for this is twofold: Firstly, less prominent reflections in this scenario also occur later in time, at which point the DRIR contains a more significant amount of diffuse energy, such that the subtraction of a single plane wave removes a smaller part of the total energy. Secondly, due to the increasingly diffuse conditions and the possible overlap of several reflections in time, the accuracy of the DOA estimate is reduced which impairs the congruence of the plane wave prototype with the real wave front. To facilitate a more intuitive insight into the results obtained with the proposed algorithm, we provide online examples for listening [25].

6. CONCLUSION

We proposed a spatial subtraction method for the decomposition of DRIRs into prominent reflections and a residual. The method extracts the time-frequency fingerprint of a reflection and effectively subtracts the corresponding components from the DRIR. It outperforms an existing method at low frequencies by taking into account the regularization of the employed radial filters, as well as at high frequencies by considering the impact of order truncation and spatial aliasing.

7. REFERENCES

[1] F. Brinkmann, L. Aspöck, D. Ackermann, S. Lepa, M. Vorländer, and S. Weinzierl, "A round robin on room acoustical simulation and auralization," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2746–2760, 2019.

[2] 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.

[3] F. Menzer, C. Faller, and H. Lissek, "Obtaining binaural room impulse responses from B-format impulse responses," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 2, pp. 395–405, 2011.

[4] 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.

[5] D. Romblom, P. Depalle, C. Guastavino, and R. King, "Diffuse field modeling using physically-inspired decorrelation filters and B-format microphones: Part I algorithm," *Journal of the Audio Engineering Society*, vol. 64, no. 4, pp. 177–193, 2016.

[6] M. Zaunschirm, M. Frank, and F. Zotter, "BRIR synthesis using first-order microphone arrays," *144th Convention of the Audio Eng. Soc.*, 2018.

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

[8] L. Gölles and F. Zotter, "Directional Enhancement of First-Order Ambisonic Room Impulse Responses By The 2 + 2 Directional Signal Estimator," in *Proceedings of Audio Mostly (AM'20)*. Graz: ACM, 2020.

[9] A. Politis, S. Tervo, and V. Pulkki, "COMPASS: Coding and multidirectional parameterization of ambisonic sound scenes," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6802–6806.

[10] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer, 2015.

[11] F. Zotter and M. Frank, *Ambisonics, A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer, 2019.

[12] B. Rafaely, "Spatial sampling and beamforming for spherical microphone arrays," in *Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2008, pp. 5–8.

[13] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1003–1010, 2007.

[14] R. H. Hardin and N. J. Sloane, "McLaren's improved snub cube and other new spherical designs in three dimensions," *Discrete and Computational Geometry*, vol. 15, no. 4, pp. 429–441, 1996.

[15] 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 Audio Eng. Soc.*, 2006.

[16] C. T. Jin, N. Epain, and A. Parthy, "Design, optimization and evaluation of a dual-radius spherical microphone array," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 1, pp. 193–204, 2014.

[17] B. Bernschütz, "Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording," Ph.D. dissertation, Technical University of Berlin, 2016.

[18] T. Deppisch, "Dataset: Free-field impulse responses of several rigid spherical microphone arrays," 2021. [Online]. Available: https://zenodo.org/record/4715801

[19] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio," in *Proc. DAGA*, Berlin, 2010, pp. 717–718.

[20] A. Neidhardt, C. Schneiderwind, F. Klein, and S. Fichna, "Dataset: Eigenmike-DRIRs, KEMAR 45BA-BRIRs, RIRs and 360° pictures captured at five positions of a small conference room," 2019. [Online]. Available: https://zenodo.org/record/2593714

[21] K. Müller and F. Zotter, "Auralization based on multiperspective ambisonic room impulse responses," *Acta Acustica*, vol. 6, no. 25, 2020.

[22] K. Müller, "Variable-Perspective Rendering of Virtual Acoustic Environments based on Distributed First-Order Room Impulse Responses," Master's thesis, University of Music and Performing Arts, Graz, 2020.

[23] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "Ambix - A Suggested Ambisonics Format," in *International Symposium on Ambisonics and Spherical Acoustics*, vol. 3, 2011, pp. 1–11.

[24] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2009, pp. 221–224.

[25] "Audio Examples WASPAA 2021b." [Online]. Available: http://www.ta.chalmers.se/research/audio-technology-group/audio-examples/waspaa-2021b/