OPEN ACCESS Freely available online
C. Hold, T. Mckenzie, G. Götz, S. J. Schlecht, and V. Pulkki, “Resynthesis of Spatial Room Impulse Response Tails With Anisotropic Multi-Slope Decays”
J. Audio Eng. Soc., vol. 70, no. 6, pp. 526–538, (2022 June).
DOI: https://doi.org/10.17743/jaes.2022.0017

Resynthesis of Spatial Room Impulse Response Tails With Anisotropic Multi-Slope Decays

CHRISTOPH HOLD,[1] *AES Student Member*, **THOMAS MCKENZIE,**[1] **GEORG GÖTZ,**[1]
(christoph.hold@aalto.fi) (thomas.mckenzie@aalto.fi) (georg.gotz@aalto.fi)
SEBASTIAN J. SCHLECHT,[1 2] *AES Associate Member* **AND VILLE PULKKI,**[1] *AES Fellow*
(sebastian.schlecht@aalto.fi) (ville.pulkki@aalto.fi)

[1]*Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland*
[2]*Media Lab, Department of Art and Media, Aalto University, Espoo, Finland*

Spatial room impulse responses (SRIRs) capture room acoustics with directional information. SRIRs measured in coupled rooms and spaces with non-uniform absorption distribution may exhibit anisotropic reverberation decays and multiple decay slopes. However, noisy measurements with low signal-to-noise ratios pose issues in analysis and reproduction in practice. This paper presents a method for resynthesis of the late decay of anisotropic SRIRs, effectively removing noise from SRIR measurements. The method accounts for both multi-slope decays and directional reverberation. A spherical filter bank extracts directionally constrained signals from Ambisonic input, which are then analyzed and parameterized in terms of multiple exponential decays and a noise floor. The noisy late reverberation is then resynthesized from the estimated parameters using modal synthesis, and the restored SRIR is reconstructed as Ambisonic signals. The method is evaluated both numerically and perceptually, which shows that SRIRs can be denoised with minimal error as long as parts of the decay slope are above the noise level, with signal-to-noise ratios as low as 40 dB in the presented experiment. The method can be used to increase the perceived spatial audio quality of noise-impaired SRIRs.

0 INTRODUCTION

Room impulse responses (RIRs) capture the reverberation characteristics of a space. Rooms with simple geometries and uniform absorption tend to feature isotropic sound energy decays, for which reverberation time is constant in all directions and the sound energy decays with a single slope [1]. However, non-uniform spaces, such as coupled rooms, feature multiple-slope decays, which typically comprise a mix of the single slopes for each space [1–3]. These are perceivable [4, 5], direction-dependent [6], and vary with inter-room position and coupling aperture size [7, 8].

Spatial RIRs (SRIRs), such as those measured using a spherical microphone array (SMA), contain the directional characteristics of a room’s reverberation. They are therefore also often called directional RIRs, which is used interchangeably in this document. Spatial room characteristics play a crucial role in localization and other room-related properties, such as sound source width, envelopment, perceived loudness, and clarity [9, 10].

Using an SMA to measure SRIRs allows encoding the microphone signals to the spherical harmonic domain (SHD) in higher-order Ambisonic (HOA) format [11, 12], which is independent of the used capture or playback devices. SRIRs allow for greater flexibility after measurement than monophonic or stereophonic alternatives because they can be analyzed using beamforming directional approaches and reproduced over both loudspeaker arrays and headphones.

However, commercial SMAs tend to be noisier than single-capsule microphones. This noise is particularly prominent because of the number of capsules and encoding process of microphone signals to the SHD. Potential noise sources include the microphone capsules and circuitry noise, as well as electromagnetic interference and quantization noise. Additionally, measurements taken in non-laboratory environments can suffer from environment background noise, such as air conditioning, people, and traffic, which intensifies with large distances between the source and receiver [13]. All of the above contribute to the problem of a noisy measurement in practice.

Excessive noise in the impulse response can substantially degrade the resulting audio quality and needs to be suppressed whenever possible. A noisy impulse response leads to degradation in rendering because it introduces unwanted temporal smearing of the input signal [14–16]. For these reasons, there is a need for a robust method to denoise measured SRIRs.

Recent studies have proposed tailored denoising algorithms for SRIRs [17–19]. They are centered around the concept of segregating late reverberation components in either spherically isotropic [19] or anisotropic components [18, 20]. The articles [18, 20] are in that sense related to this current work. They presented a conceptual framework for tackling the problem of directional SRIR denoising, suggesting to first carry out a plane-wave decomposition in many directions, followed by a mixing time estimation based on spatial coherence. They have then approximated the reverberation tail with exponential slopes, potentially accounting for multi-slope situations. These methods then proposed to resynthesize the reverberation tail in order to achieve signal-to-noise–ratio (SNR) improvements. From the published literature, however, details about the resynthesis have been left unaddressed. The present authors have identified key issues related to the spatial analysis and resynthesis to the SHD, which they improved upon by the recently formalized spherical filter bank (SFB) in the SHD [21]. Furthermore, they could not find a detailed performance analysis of the presented framework, e.g., in terms of input-to-output error, which is addressed more explicitly in the present article.

This paper describes a method for denoising SRIRs, incorporating both anisotropic directional analysis and multi-slope sound energy decay fitting. The approach aims to be robust to measurements with low SNRs and non-Gaussian noise. A brief description of the approach is as follows. Firstly, an SFB is used to extract directionally constrained signals of the SHD SRIR. These beamformer outputs form a set of spatially filtered signals that subdivides the spatial impulse input signal. Next, each directional component is analyzed and described as multiple exponential decay slopes, using a recently proposed neural network parameter fitting method [22], which also provides a noise-level estimate. The proposed approach then replaces the presumed noisy components below this noise level, using modal synthesis to resynthesize the late decay tails to match the timbre and decay rate of the estimated slopes. Finally, the SFB design allows re-encoding of the processed signals back to the SHD while preserving their energy.

This paper is structured as follows. SEC. 1 presents an analysis with insights about SRIRs, further motivating the multi-slope decay fit. SEC. 2 details the methodology and algorithm design for the multi-slope directional denoising approach. The approach is evaluated in SEC. 3 both technically, which includes directional SNR comparisons, using both simulated SRIRs with added Gaussian noise and measured SRIRs with captured noise, and perceptually, through a listening test on denoised SRIR measurements with varying SNRs. The results of the evaluation are discussed in SEC. 4, and the paper is concluded, along with further work proposed, in SEC. 5.

1 ANISOTROPIC MULTI-SLOPE DECAY SRIR

To better understand anisotropic reverberation, Fig. 1 presents the input amplitude density of a measured SRIR ($\mathrm{SRIR}_{\mathrm{meas}}$), which is used throughout this paper. The depicted directional dependency of the SRIR supports the need for directional processing in the following.

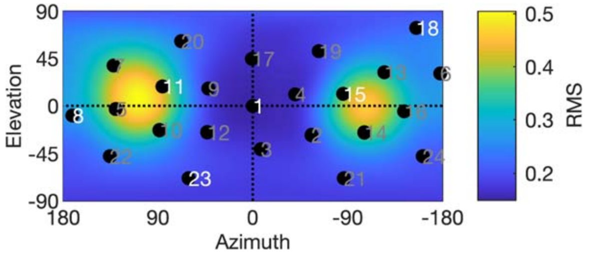

Fig. 1. The scaled input RMS of the measured spatial room impulse response $\mathrm{SRIR}_{\mathrm{meas}}$ ($N_{\mathrm{sph}} = 3$) under test, evaluated on a dense grid. Numbers indicate beamformer steering directions utilized within this article. The white numbers indicate the filter steering directions close to the Cartesian axes used for further performance analysis visualizations.

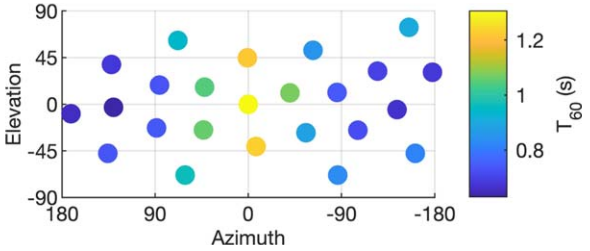

Fig. 2. Broadband T_{60}, estimated from spatial Butterworth filters, on the measured spatial room impulse response $\mathrm{SRIR}_{\mathrm{meas}}$ ($N_{\mathrm{sph}} = 3$) input shown in Fig. 1. The markers' colors show the estimated broadband T_{60} over each beamformer, indicating an anisotropic reverberation tail.

However, not only the reverberation level but also the reverberation time can change with the angle of incidence. Therefore, it is not sufficient to treat only reverberation level as a function of direction, but also the decay rate. This becomes apparent in Fig. 2, which shows a frequency band average of the T_{60} reverberation length for each of the beamformer directions shown in Fig. 1. Here, the T_{60} is the time taken for the energy of the impulse response to decay to −60 dB lower than the initial impulse [23]. In insufficient SNR situations, it is suggested to infer T_{60} by doubling the decay time to −30 dB [23]. This observed directional dependency calls for adequate directional processing, which is addressed with an SFB in this article.

Furthermore, the energy decay curve (EDC) may be composed of multiple decaying exponential slopes at nonidentical decay rates. Multi-slope decays are used for quantifying complex acoustic phenomena, such as room coupling or non-uniform absorptive material distributions [8, 24]. Fig. 4 demonstrates that the reverberation decay of $\mathrm{SRIR}_{\mathrm{meas}}$ is composed of multiple slopes. It can therefore be concluded that in the present case, a typical single-slope assumption is violated and hence requires further methods, which are detailed in the following.

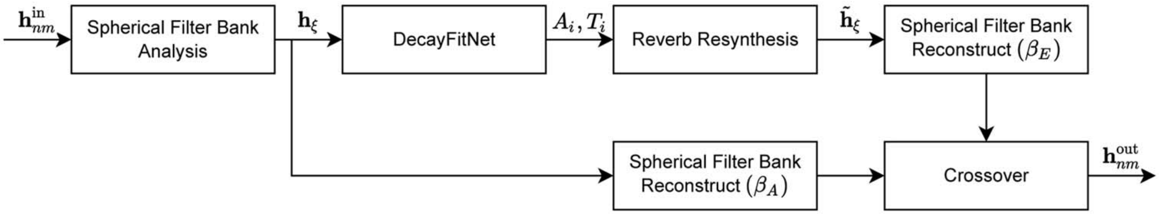
$\mathbf{h}_{nm}^{\text{in}}$
Spherical Filter Bank Analysis
$\mathbf{h}_\xi$
DecayFitNet
A_i, T_i
Reverb Resynthesis
$\tilde{\mathbf{h}}_\xi$
Spherical Filter Bank Reconstruct (β_E)
Spherical Filter Bank Reconstruct (β_A)
Crossover
$\mathbf{h}_{nm}^{\text{out}}$

Fig. 3. Block diagram of the denoising approach presented in this paper. The subscript *nm* marks spherical harmonic domain signals. Each step is detailed in SEC. 2.

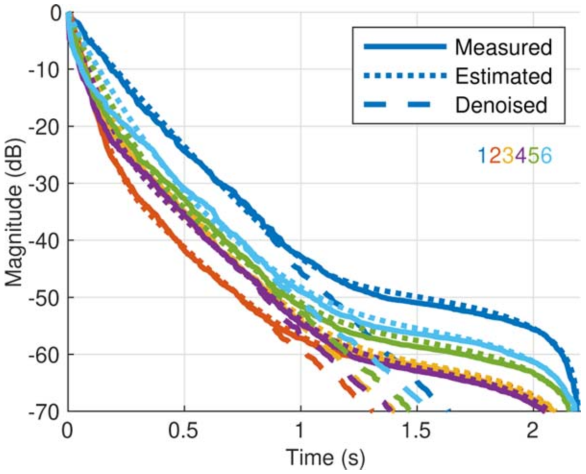

Fig. 4. Energy decay curve (EDC) fit (multi-slopes) in the 1-kHz frequency band on measured SRIR, as shown in Fig. 1, beamformers steering close to [+*x*, −*x*, +*y*, −*y*, +*z*, −*z*], indexed by colors. The estimated EDC (dotted) stemming from the model parameterization matches closely the measured input EDC (solid), and the denoised EDC (dashed) obtained from the output shows the SNR improvement.

2 METHODS

This section explains the methodology of the directional denoising approach of impulse response tails with anisotropic multi-slope decays, beginning with the spatial deconstruction of the SRIR and the spatially filtered decay slope estimation, to the resynthesis of the late decay and reconstruction of the SHD SRIR. Fig. 3 illustrates the proposed denoising approach, consisting of the following steps:

1. Input noisy HOA SRIR $\boldsymbol{h}_{nm}^{\text{in}}(t)$.
2. Extract spatially filtered signals $\boldsymbol{h}_\xi(t)$ in directions Ω_ξ, see Eq. (4).
3. Use DecayFitNet [22] to estimate multi-slope decay model parameters $A_i, T_i, A_{\text{noise}}$ on $h_\xi(t)$, see SEC. 2.2.
4. Resynthesize late tail from estimated parameters using modal synthesis while zeroing the noise component ($A_{\text{noise}} = 0$).
5. Ensure spatial band-limitation of synthesized $\tilde{\boldsymbol{h}}_\xi(t)$, see Eq. (17).
6. Calculate crossover point between input $\boldsymbol{h}_\xi(t)$ and synthesis $\tilde{\boldsymbol{h}}_\xi(t)$ from estimated A_{noise}.
7. Re-encode to SHD by perfect reconstruction of $\boldsymbol{h}_\xi(t)$ and energy preserving reconstruction of $\tilde{\boldsymbol{h}}_\xi(t)$, see Eqs. (14) and (15), respectively.
8. Output denoised HOA SRIR $\hat{\boldsymbol{h}}_{nm}^{\text{out}}(t)$.

Spherical harmonics in this paper are orthonormalized as defined in [25, Eq. (6.20)], on the unit sphere $\Omega = [\phi, \theta] \in S^2$, with the azimuth angle ϕ and zenith/colatitude angle θ. A signal model may be formulated with additive noise s^{noise} as

$$s(t, \Omega) = \hat{s}(t, \Omega) + s^{\text{noise}}(t, \Omega), \tag{1}$$

in the SHD as

$$\sigma_{nm}(t) = \hat{\sigma}_{nm}(t) + \sigma_{nm}^{\text{noise}}(t). \tag{2}$$

It will be further assumed that all signals are spatially band-limited to spherical harmonic (SH) order N_{sph}. The discrete spherical harmonic transform (SHT), up to order N_{sph}, and the inverse SHT (iSHT) are given as [26, Eqs. (3.34) and (3.35)]. In order to avoid confusion, the SHD spectrum is explicitly marked with the subscript $(\,\cdot\,)_{nm}$. Because SRIRs typically decay exponentially temporally, whereas the noise terms are temporally static, the model allows estimating the clean SRIR $\hat{\boldsymbol{h}}_{nm}^{\text{out}}(t)$ from a noisy input SRIR $\boldsymbol{h}_{nm}^{\text{in}}(t)$ in the early decay, if the SNR is sufficient, and hence the noise term is negligible.

2.1 Directional Analysis and Resynthesis

Reverberation models often make the assumption of isotropic decay. However, as highlighted in SEC. 1, this simplification may not apply in practice. Assuming potentially anisotropic reverberation, the analysis requires directional processing. Hence, in this study, the analysis, resynthesis, and therefore denoising are carried out on directionally constrained parts of the sound field. This directional constrain is imposed by beamformers, i.e., spatial filters in the SHD. Partitioning the SHD input signal with a complete set of beamformers, uniformly covering the domain (i.e., the sphere), may therefore be interpreted as an SFB. The SFB thereby enables intuitive techniques of directional processing in the SHD. The key idea here is that the SFB analysis and re-synthesis form a pair that allows transforming between domains under preservation and reconstruction criteria detailed in the following.

The SFB extracts signals from the spatially continuous SHD on a spatially discrete grid, allowing the transform

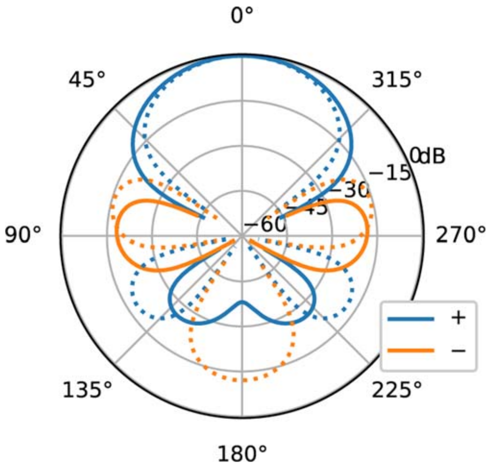

Fig. 5. Comparison of spatial Butterworth filter (solid) for $k = 5$, $l_c = 2.5$, and normalized plane-wave decomposition (dotted) for third-order ($N_{\text{sph}} = 3$) spherical harmonic signals.

back to the SHD. The key difference of SFBs compared to an unmodified SHT and iSHT is the possibility to choose the underlying (axisymmetric) filter pattern. The evaluated spatial pattern of the iSHT is proportional to the maximum directivity (max DI) beamformer or plane-wave decomposition, i.e., a higher-order hyper-cardioid. Despite showing maximum directivity, the back-lobe and side-lobe suppression can be improved in the context of SRIR analysis. Previously, patterns with greater front-to-back ratio generally increased the performance, because differences in opposite directions are typically easily audible and thus need to be resolved. The utilized pattern is a design choice and should be tailored to the application and requirements on hand. For the purpose of demonstration, this study uses a spatial Butterworth filter, which may be formulated as $c_n^{\text{Butterworth}} = \frac{1}{\sqrt{1+(n/n_c)^{2k}}}$, where $n \in [0, N_{\text{sph}}]$ is the SH order, N_{sph} is the maximum SH order, n_c the cut-on SH order, and $k \in \mathbb{N}$ the filter order [27, Table 3.1]. The filter is then normalized to have unit amplitude in steering direction. Its polar pattern is compared to the max DI beamformer in Fig. 5, which demonstrates superior back-lobe suppression while only minimally widening the main lobe. Another suitable alternative would be the max r_{E} pattern, which is conceptually similar to a higher-order super-cardioid. In comparison, the max DI beamformer is given by $c_n^{\text{max DI}} = \frac{4\pi}{(N_{\text{sph}}+1)^2}$. Any axisymmetric analysis pattern is defined in terms of $N_{\text{sph}} + 1$ modal weighting coefficients c_n, where the operator diag_N formalizes repeating m times the weight of order n. The SH coefficients of these beamformers are, for axisymmetric patterns, given directly as

$$\boldsymbol{w}_{nm} = \text{diag}_N(\boldsymbol{c}_n)\, \boldsymbol{y}_n^m(\Omega')^{\text{H}}, \tag{3}$$

with the steering vector $\boldsymbol{y}_n^m(\Omega')$ evaluating the spherical harmonics in direction Ω', and the Hermetian transpose $(\cdot)^{\text{H}}$.

The filter bank framework allows establishing the interpretation of analysis and synthesis. Extracting beamformer $\xi \in [1, \ldots, J]$ signals $\boldsymbol{s}_\xi$ as

$$\boldsymbol{s}_\xi = \boldsymbol{A}\boldsymbol{\sigma}_{nm}^{\text{in}}, \tag{4}$$

where the analysis matrix $\boldsymbol{A}$ of size $J \times (N_{\text{sph}} + 1)^2$ is comprised of stacked beamformers $\boldsymbol{w}_{nm}$.

The latter is formulated accordingly with stacked steering vectors $\boldsymbol{y}_n^m(\Omega_\xi)$ to a matrix $\boldsymbol{Y}$

$$\boldsymbol{A} = \left[\text{diag}_N(\boldsymbol{c}_n^{\text{an}})\boldsymbol{Y}^{\text{H}}\right]^{\text{H}} = \boldsymbol{Y}\,\text{diag}_N(\boldsymbol{c}_n^{\text{an}}). \tag{5}$$

The re-encoding of the beamformer output signals back to the SHD is formulated accordingly as

$$\boldsymbol{\sigma}_{nm}^{\text{out}} = \boldsymbol{B}^{\text{H}}\boldsymbol{s}_\xi, \tag{6}$$

where the matrix $\boldsymbol{B}$ of size $J \times (N_{\text{sph}} + 1)^2$ is

$$\boldsymbol{B} = \boldsymbol{Y}\,\text{diag}_N(\boldsymbol{c}_n^{\text{syn}}). \tag{7}$$

Covering the SHD uniformly, the steering vectors $\boldsymbol{y}_n^m(\Omega_\xi)$ are evaluated on a uniform grid. The re-encoding to the SHD requires a grid supporting the spatial integration of a polynomial of order $2N_{\text{sph}}$, which is discretized by, e.g., a spherical t-design [28]. The grid may be rotated without any loss of generality, e.g., in order to align the first steering direction to $[1, 0, 0]^{\text{T}}$. Although spherical t-designs are generally over-determined and therefore not optimal in terms of minimal spatial interaction, their property of constant quadrature weights is essential in the current formulation.

Preservation factors for a uniform grid are derived in [21] as [21]:

$$\beta_A = \frac{\sqrt{4\pi}}{w_{00}^{\text{an}} J}, \tag{8}$$

$$\beta_E = \frac{4\pi}{[\boldsymbol{w}_{nm}^{\text{an}}]^{\text{H}}\boldsymbol{w}_{nm}^{\text{an}} J}. \tag{9}$$

These factors ensure that the sum over all J analysis filter weightings w_ξ either preserves amplitude as

$$\sum_{\xi}^{J} \beta_A w_\xi(\Omega) = 1, \quad \forall\Omega \in S^2, \tag{10}$$

or preserves energy as

$$\sum_{\xi}^{J} \beta_E w_\xi^2(\Omega) = \sum_{\xi}^{J} \left[\sqrt{\beta_E}\, w_\xi(\Omega)\right]^2 = 1, \quad \forall\Omega \in S^2. \tag{11}$$

The reconstruction of amplitude, defined in this case more strictly as perfect reconstruction, is achieved if the signal restored from the filter bank $\boldsymbol{\sigma}_{nm}^{\text{out}}(t)$ exactly matches the input signal $\boldsymbol{\sigma}_{nm}^{\text{in}}(t)$

$$\boldsymbol{\sigma}_{nm}^{\text{out}}(t) \overset{!}{=} \boldsymbol{\sigma}_{nm}^{\text{in}}(t), \tag{12}$$

for any time t. If perfect reconstruction is not possible, the reconstruction of energy over time may be demanded (projection of signal onto itself) such that

$$\left\langle \boldsymbol{\sigma}_{nm}^{\text{out}}(t), \boldsymbol{\sigma}_{nm}^{\text{out}}(t)\right\rangle \overset{!}{=} \left\langle \boldsymbol{\sigma}_{nm}^{\text{in}}(t), \boldsymbol{\sigma}_{nm}^{\text{in}}(t)\right\rangle. \tag{13}$$

Shown previously in [29], perfect reconstruction is achieved by

$$\boldsymbol{\sigma}_{nm}^{\mathrm{out}}(t) = \beta_A \boldsymbol{B}^{\mathrm{H}} \boldsymbol{s}_{\xi}(t), \tag{14}$$

and energy preservation reconstruction by

$$\tilde{\boldsymbol{\sigma}}_{nm}^{\mathrm{out}}(t) = \sqrt{\beta_E} \boldsymbol{B}^{\mathrm{H}} \tilde{\boldsymbol{s}}_{\xi}(t), \tag{15}$$

where $\boldsymbol{c}_n^{\mathrm{syn}} = \frac{1}{\boldsymbol{c}_n^{\mathrm{an}}/c_0^{\mathrm{an}}}$ is defined for both, according to Eq. (7). This formulation shows the advantage of the filter bank framework, because the preservation factors β_A and β_E are the only difference between both reconstruction topologies. For further details and derivations, see [21, 29].

Because Eq. (6) essentially constitutes a modified SHT, the transform to order N_{sph} requires spatially band-limited signals in the spatially discrete domain to avoid spatial aliasing. Because the signals $\boldsymbol{s}_\xi$ stem from a directional filtering operation, the inter-signal coherence of a diffuse, spatially band-limited signal may be formulated as the spatial coherence matrix

$$\boldsymbol{R}_{\xi,\xi'} = \boldsymbol{Y}_{\xi,nm} \operatorname{diag}_N(\boldsymbol{c}_n^{\mathrm{an}}) \boldsymbol{Y}_{\xi,nm}^{\mathrm{H}}. \tag{16}$$

Assuming decorrelated signals $\breve{\boldsymbol{s}}_\xi(t)$, such as from independent white noise realizations, the spatial band-limitation may be reestablished directly by

$$\tilde{\boldsymbol{s}}_\xi(t) = \boldsymbol{R}_{\xi,\xi'}/(\boldsymbol{1}\boldsymbol{R})\breve{\boldsymbol{s}}_\xi(t), \tag{17}$$

where the spatial coherence matrix is normalized by the sum over one axis (denoted as $\boldsymbol{1R}$), avoiding a change in total power. Note that frequency-dependent spatial coherence $\boldsymbol{R}_{\xi,\xi'}(f)$ may also be introduced, e.g., in order to mimic the measurement microphone array proprieties. In practice, SMAs lose their directivity toward low frequencies because of limitations of the radial filter gains. This could be matched by increasing the spatial coherence according to an a-priori measurement. However, in favor of generality, only the broadband constraint of Eq. (17) was applied.

2.2 Parameter Estimation

EDCs are representations of the sound energy decay of rooms. They can be obtained from RIRs by applying the backward integration procedure proposed by Schroeder [30]. The sound energy decay is frequently modeled as a single decaying exponential with one decay rate, commonly known as the reverberation time [23, 31]. However, it is becoming increasingly relevant to fit higher-order models to EDCs that include multiple exponentials and consequently multiple decay rates, as motivated earlier. In the following, EDCs are modeled with the decay model $d_\kappa(t)$ [32, 33]:

$$d_\kappa(t) = A_{\mathrm{noise}}(L - t) + \sum_{i=1}^{\kappa} A_i \left[e^{\frac{\ln(10^{-6})\cdot t}{f_{\mathrm{s}} T_i}} - e^{\frac{\ln(10^{-6})\cdot L}{f_{\mathrm{s}} T_i}} \right], \tag{18}$$

where κ is the model order, T_i and A_i are the decay rate and amplitude of the i^{th} exponential decay, A_{noise} is the amplitude of the noise term, L is the length of the EDC, $\ln(\cdot)$ denotes the natural logarithm, t is the sample index, and f_{s} is the sampling frequency of the RIR.

Various algorithms exist for estimating the model parameters T_i, A_i, and A_{noise} from noisy RIR measurements [22, 32, 34]. In the denoising approach, a recently proposed neural network architecture is used, which was shown to be robust, computationally efficient, and deterministic (as opposed to iterative), [22]. The neural network was trained with a purely synthetic EDC dataset, spanning EDCs with the following decay parameter ranges:

$$\begin{array}{lclcl} 1 & \leq & \kappa & \leq & 3 \\ 0.1\frac{L}{f_{\mathrm{s}}} & \leq & T_i & \leq & 1.5\frac{L}{f_{\mathrm{s}}} \\ -30\,\mathrm{dB} & \leq & A_i & \leq & 0\,\mathrm{dB} \\ -140\,\mathrm{dB} & \leq & A_{\mathrm{noise}} & \leq & -30\,\mathrm{dB} \end{array}. \tag{19}$$

The performance of the neural network was reliable when analyzing two measured datasets with more than 20,000 EDCs [22]. The evaluation datasets featured variable acoustic conditions, such as a considerably varying amount of furniture, room coupling, scattering, and diffraction from geometry.

In the denoising algorithm, the DecayFitNet is used as described in [22]. The network returns estimates for T_i, A_i, and A_{noise} for every analyzed octave band.

2.3 Denoising

The potentially noisy SRIR Ambisonic signals $\boldsymbol{h}_{nm}^{\mathrm{in}}(t)$ are first spatially segregated by the SFB analysis Eq. (4). The now spatially discrete signals $\boldsymbol{h}_\xi(t)$ may then undergo the decay analysis for each beamformer ξ, which provides the (multi-slope) reverberation tail parameterization and a noise floor estimate. These parameters inform a modal resynthesis of the late reverberation of each directional component. This study uses modal resynthesis because it allows for a direct resynthesis of signals that correspond to the model Eq. (18). Because SRIRs can exhibit very steep frequency roll-offs, in combination with short decay times toward high frequencies, the band rejection of time-frequency filter banks may be a limiting factor when designing the signals in the time domain, such as by band-passed noise sequences. Each ξ contains 2^{13} modes with logarithmically spaced randomized frequency $\omega \in [0, \pi]$ and randomized phase $\varphi \in [0, 2\pi]$. The modes are then modulated with the corresponding time derivative of the cumulative energy decay, i.e., $d'_\kappa(t)$, see Eq. (18) with $A_{\mathrm{noise}} = 0$. To vary smoothly between frequency bands, the decay envelope is linearly interpolated in log-scale between the neighboring bands' center frequencies, i.e., $d'_{\kappa,\omega}(t)$.

The modal resynthesis is carried out as

$$\breve{h}_\xi(t) = \sum_\omega d'_{\kappa,\omega}(t) \sin(\omega t + \varphi)\sqrt{2\omega}, \tag{20}$$

where the latter factor compensates the energy distribution induced by the logarithmic spacing of ω.

The crossover point between the input SRIR and resynthesized tail is found based on the noise estimated from the parameterization while aiming to preserve as much of the original input as possible. Therefore, the squared input RIR is first smoothed with a 100th-order median filter. Then, the crossover point is determined by the first time sample 6 dB

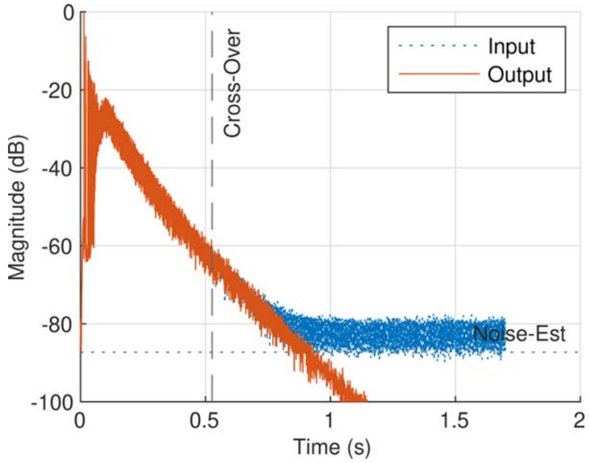

Fig. 6. Magnitude in decibels of the input $\boldsymbol{h}^{\mathrm{in}}_{nm}(t)$ and output $\boldsymbol{h}^{\mathrm{out}}_{nm}(t)$ signal instantaneous power of the presented algorithm using simulated spatial room impulse responses (SRIR$_{\mathrm{sim}}$) with 50-dB signal-to-noise–ratio (SNR) additive noise. The horizontal line indicates the highest noise estimate from the input parameterization, and the vertical line the determined crossover point between the input and denoising that results in the output.

above the highest estimated A_{noise} scaled by 2B because of the (simplified) band-pass proportion from B band-passes, and the orthonomormality error of the SFB analysis $\boldsymbol{A}$, determined as trace$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})/J$. Both the estimated noise floor determined from the directional parameterization and the resulting cross-over start, besides the SRIR instantaneous power, are shown in Fig. 6. Assuming decorrelated input and resynthesis tails, a 300-ms constant power fade parameterized with $a_{\mathrm{p}}(t)$ blends between both parts, such that

$$\hat{\mathbf{h}}^{\mathrm{out}}_{nm}(t) = \sqrt{a_{\mathrm{p}}(t)}\beta_A B^{\mathrm{H}} h_{\xi}(t) + \sqrt{(1-a_{\mathrm{p}}(t))}\sqrt{\beta_E}\mathbf{B}^{\mathrm{H}}\tilde{\mathbf{h}}_{\xi}(\mathbf{t}). \quad (21)$$

The unaltered early part of the SRIR reconstructs perfectly using Eq. (14). Replacing the beamformer outputs by independently randomized realizations, the spatial coherence and thus spatial bandwidth limitation needs to be re-established. Assuming fully decorrelated components $\breve{\boldsymbol{h}}_{\xi}(t)$, the normalized beam correlation matrix Eq. (17) may therefore simply be applied to obtain $\tilde{\boldsymbol{h}}_{\xi}(t)$. The signals of the resynthesized noise tail are re-encoded to the SHD using the energy-preserving topology of Eq. (15).

3 EVALUATION

3.1 Input Signals

The anisotropic multi-slope decay resynthesis algorithm was evaluated using two types of input signal. The first was a simulated SRIR, which provided a noise-free reference case to which varying levels of additive noise could be added. The second was a measured SRIR from a coupled room SRIR dataset [13], recorded with an Eigenmike and with audible noise present. The two SRIRs are detailed in the following and referred to as simulated SRIR (SRIR$_{\mathrm{sim}}$) and SRIR$_{\mathrm{meas}}$. Both SRIRs were order truncated to $N_{\mathrm{sph}} = 3$, which simplifies the presentation without any loss of generality. A waterfall plot of both scenarios is shown in Figs. 7(a) and 8(a).

The SRIR$_{\mathrm{sim}}$ was created using an image-source model for the early reflections and a stochastic decay part for the late reverberation. This hybrid model provided natural-sounding results in various previous applications and is considered a standard practice. The shoebox model simulated a room of dimensions 10.2 m x 7.1 m x 3.2 m, V = 231.7 m^3. The reverberation time was specified as 1 s, falling to 0.5 s at high frequencies. The source was positioned $[1, 0, 0]^{\mathrm{T}}$ m from the receiver. The reverb was faded linearly (over 50 ms) into an exponentially decaying stochastic reverb after 52.81 ms. The latter corresponds to $t_{\mathrm{mp95}} = 0.0117 \cdot V + 50.1$, a conservative estimation of the mixing time according to [35].

The procedure of utilizing a noise-free simulated SRIR allowed for the controlled addition of noise at a chosen level. According to the signal model Eq. (1), the experiment injected additive white noise of varying SNR. Therefore, spherically isotropic, but spatially band-limited, white noise was simulated as independent realizations of white noise in each SH component $\sigma^{\mathrm{noise}}_{nm}$ of equal power [36]. Fig. 6 depicts the instantaneous power of the resulting signal.

The second test signal was the measured SRIR$_{\mathrm{meas}}$, which includes multi-slope decays and directional reverberation, as well as non-Gaussian noise present in the measured impulse response. The measurement was taken according to a dataset of coupled SRIRs recorded using the mh Acoustics em32 Eigenmike, an SMA capable of fourth-order HOA capture[1] [13]. The chosen SRIR was measured at a distance of 250 cm, which is in the center of the coupling aperture between two rooms, of the Meeting Room to Hallway transition. The source is positioned inside the less reverberant room, and auralization of the SRIR shows clear directional elements to the decay and measured noise, making this SRIR appropriate as a test signal for the evaluation of the algorithm. Besides the microphone self-noise, there were multiple (directional) noise sources present. As needed for the evaluation, the average of six 10-s exponential sweep measurements achieved the maximum SNR in this scenario.

3.2 Technical Evaluation

3.2.1 Test Metrics

Informal listening showed a clear and consistent SNR improvement with the proposed algorithm. In order to quantify the SNR increase, varying additive white diffuse noise was added to the noise-free SRIR$_{\mathrm{sim}}$. This allows a direct SNR comparison between the noise-free original SRIR and denoised SRIR obtained from an equivalent but noisy input.

The SNR is defined as the ratio between the measured power of a signal with negligible noise contribution and the measured power of a predominantly noisy signal. Therefore, the power of the first 5 ms is measured starting with the direct sound, comparing to the measured power of the last 5 ms of the IR. A metric to evaluate noise reduction based on the measured SNR improvement may then be

[1] Database available at http://doi.org/10.5281/zenodo.4095493.

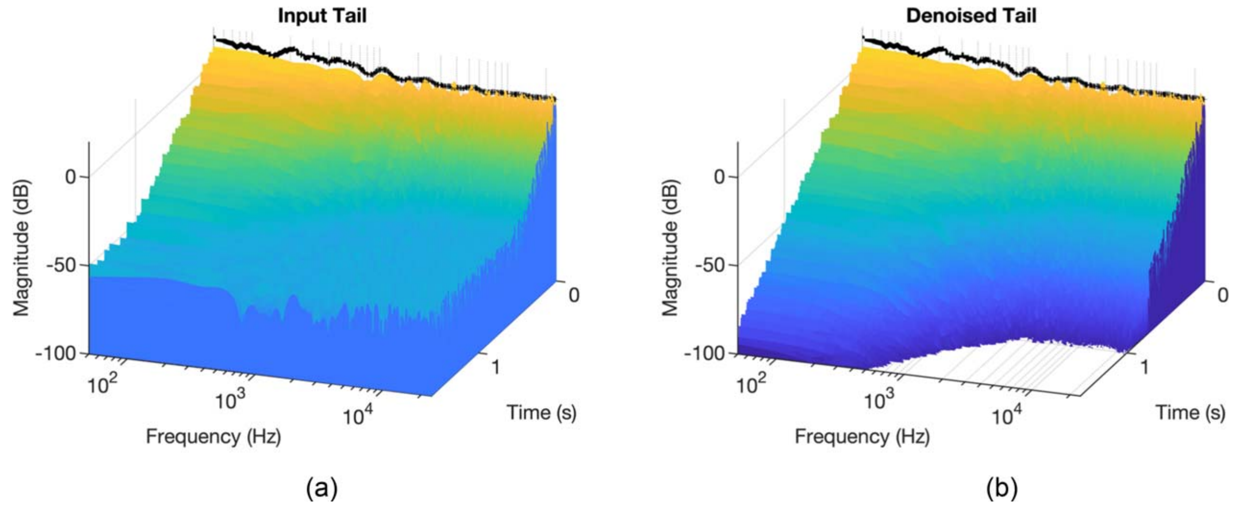

Fig. 7. Waterfall plots over the first beamformer output of the simulated spatial room impulse response ($\mathrm{SRIR}_{\mathrm{sim}}$) tail after 100 ms, between the noisy input (a) and denoised output (b). The input was simulated by adding white diffuse noise with 50-dB signal-to-noise ratio (SNR) to a noise-free impulse response (IR). The solid line visualizes the 1/3-octave smoothed IR tail magnitude (in decibels).

defined as

$$\begin{aligned}\mathrm{SNR}_{\mathrm{impv}} = \frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}} &= \frac{P^{\mathrm{out}}_{t_5} / P^{\mathrm{out}}_{t_{5,\,end}}}{P^{\mathrm{in}}_{t_5} / P^{\mathrm{in}}_{t_{5,\,end}}} \\ &= \frac{\mathrm{trace}(C^{\mathrm{out}}_{nm,t_5}) / \,\mathrm{trace}(C^{\mathrm{out}}_{nm,t_{5,\,end}})}{\mathrm{trace}(C^{\mathrm{in}}_{nm,t_5}) / \,\mathrm{trace}(C^{\mathrm{in}}_{nm,t_{5,\,end}})}, \end{aligned} \tag{22}$$

using

$$P = \frac{1}{(N+1)^2 T}\,\mathrm{trace}(C_{nm}), \tag{23}$$

and the SHD signal covariance C_{nm} over time interval T.

To quantify the spectral error across the sphere, the mean absolute difference of the directionally constrained spectra between the noise-free input and denoised output is compared as

$$\overline{E}_{\mathrm{spec}} = \frac{1}{JK} \sum_{\xi=1}^{J} \sum_{k=1}^{K} \left| |H^{\mathrm{late,out,smth}}_{\xi,k}| - |H^{\mathrm{late,in,smth}}_{\xi,k}| \right|. \tag{24}$$

Therefore, the log space (decibels) difference of the 1/3-octave–band smoothed spectra H is averaged over the frequency index k and beamformers ξ, essentially comparing the solid to dashed lines shown in Fig. 9. This error will also contain absolute level mismatches among directional components.

The reverberation tail reconstruction is apparent in the waterfall plots in Fig. 7. Quantifying the reverberation tail

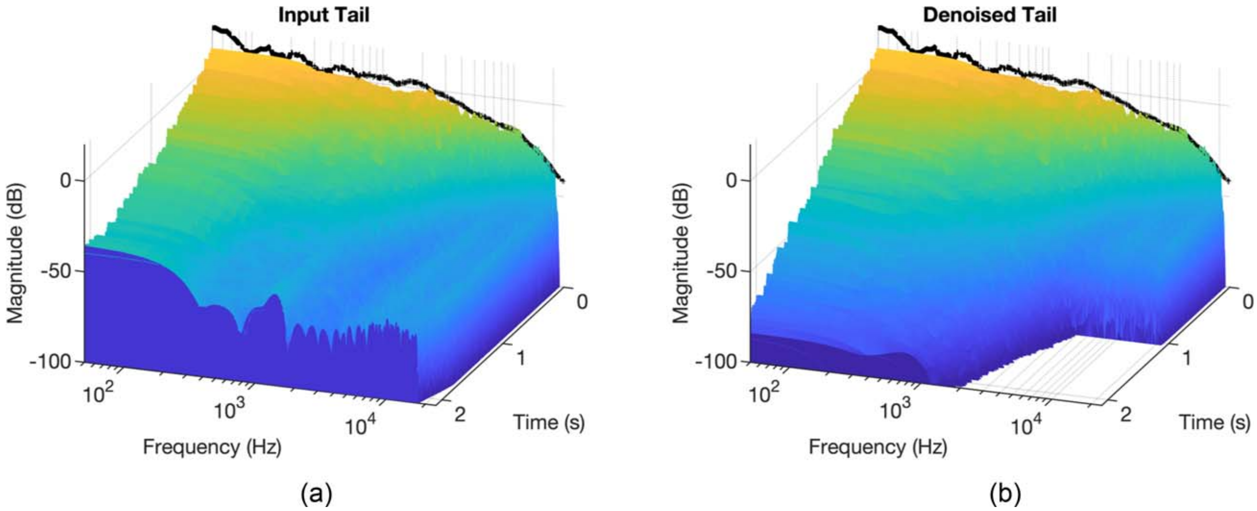

Fig. 8. Waterfall plots over the first beamformer output of the measured spatial room impulse response ($\mathrm{SRIR}_{\mathrm{meas}}$) tail after 100 ms, between the noisy input (a) and denoised output (b). Input measured by a spherical microphone array. The solid line visualizes the 1/3-octave smoothed impulse response (IR) tail magnitude (in decibels).

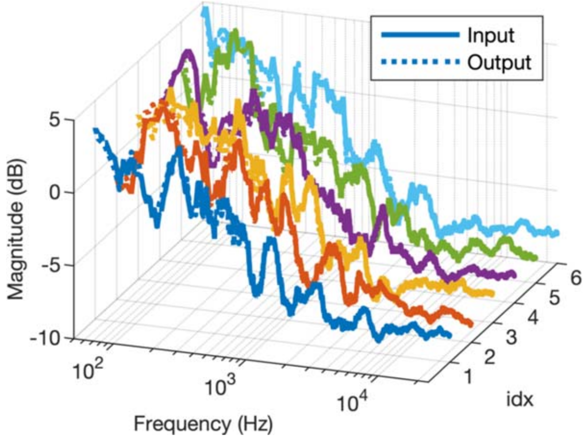

Fig. 9. Noise-free input (solid) and denoised output (dotted) reverberation tail spectra (after 100 ms). Output obtained from simulation with 30-dB signal-to-noise−ratio (SNR) additive noise and 1/3 octave smoothed. Shown are the beamformers ξ steering close to [+*x*, −*x*, +*y*, −*y*, +*z*, −*z*], numbered according to *idx*, as in Fig. 1.

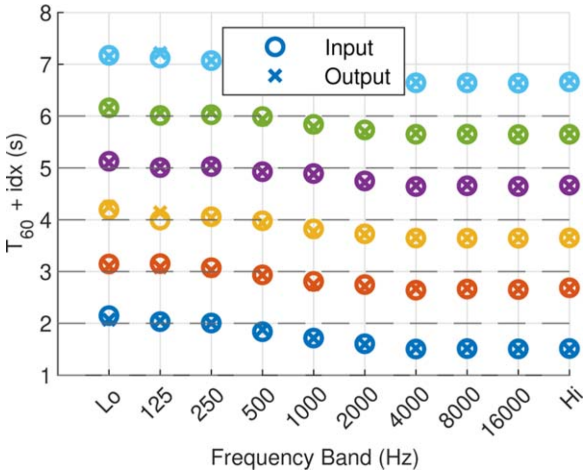

Fig. 10. Noise-free input (circles) and denoised output (crosses) $T_{60,\,\xi}(b)$. Output obtained from simulation with 30-dB signal-to-noise−ratio (SNR) additive noise. Shown are the beamformers ξ steering close to [+*x*, −*x*, +*y*, −*y*, +*z*, −*z*], numbered and offset according to *idx*, as in Fig. 1.

length error between the noise-free input and denoised output

$$\overline{E}_{T_{60}} = \frac{1}{JB}\sum_{\xi=1}^{J}\sum_{b=1}^{B}|T^{\text{out}}_{60,\xi}(b) - T^{\text{in}}_{60,\xi}(b)| \tag{25}$$

gives an indicator of the mean T_{60} error introduced by the processing. Extracted from backward-integrated EDCs in octave filter bands *b*, the time sample that first crosses −60 dB marks $T_{60}(b)$. In low SNR conditions, $T_{60}(b)$ is approximated by doubling the time to cross −30 dB. An example of the averaged values is shown in Fig. 10.

3.2.2 Results

Table 1 presents the results of the technical evaluation comparing the noise-free input SRIR_{sim} and denoised output, obtained from the same SRIR with varying levels of additive noise. When increasing the level of simulated additive diffuse white noise, i.e., decreasing $\text{SNR}_{\text{in, sim}}$, the reported algorithm generally performed strong denoising. Measured SNR_{impv} indicates steady SNR improvements, which saturates on one end because of no measurable differences and on the other end at around $\text{SNR}_{\text{in, sim}} = 30$. Both the mean spectral error $\overline{E}_{\text{spec}}$ and mean reverberation length error $\overline{E}_{T_{60}}$ stay very low up until $\text{SNR}_{\text{in, sim}} = 30$.

Table 1. Signal-to-noise−ratio (SNR) improvements [see Eq. (21)] and Mean Spectral Error $\overline{E}_{\text{spec}}$ [see Eq. (22)], between original and denoised simulated spatial room impulse responses (SRIR_{sim}) ($N_{\text{sph}} = 3$) with varying levels of added noise $\text{SNR}_{\text{in, sim}}$ in decibels, Mean Reverberation Time Error $\overline{E}_{T_{60}}$ in seconds [see Eq. (23)].

$\text{SNR}_{\text{in, sim}}$	SNR_{impv}	$\overline{E}_{\text{spec}}$	$\overline{E}_{T_{60}}$
100	1.04	≤0.01	≤0.01
80	19.96	≤0.01	≤0.01
60	38.12	≤0.01	≤0.01
50	48.69	0.01	0.01
40	57.31	0.02	0.02
30	62.18	0.05	0.04
20	53.95	0.20	0.34

3.3 Perceptual Evaluation

3.3.1 Test Design

The previously reported technical evaluation showed good results on the tested single-slope SRIR when modified with varying levels of additive white noise. However, in reality, SRIRs are more complex and can only be captured in all their aspects by measurements. These measurements cannot be achieved noise free because of limitations of the microphones and environmental noise. To quantify the real-world perceived performance of the proposed algorithm, a perceptual listening test was conducted with the measured SRIR detailed in Sec. 3.1.

The listening test subjects were recruited from the Aalto University Acoustics Lab and can be considered experienced listeners, designing and participating in multiple critical listening tests before. A total of 12 listeners (male/female), aged 24−38 years (mean: 29.5), rated the items in a single trial according to a Multiple Stimuli with Hidden Reference and Anchor–like test paradigm (using [37]) in terms of the presented spatial denoising quality (*poor−excellent*). Out of these 12 participants, nine identified the reference condition by rating it higher than 90, whereas the other results were excluded from the evaluation. The test took place in the listening booths of the Aalto University acoustics laboratories and took around 15 min in total.

All SRIRs were binaurally rendered by SH domain convolution, with a set of head-related impulse responses obtained from a KU100 dummy head [38]. The SH spectrum tapering approach with coloration compensation was applied in order to mitigate high-frequency loss by SH order truncation, as described in [39]. The equalization therein can be seen as a diffuse field equalization technique, where the resulting renderings contained enough high-frequency content for the judging of fine details, without relying on further non-linear processing.

The binaural SRIRs were convolved with a 5-s loop of a dry drumkit recording. The signal was chosen to excite a broadband response with enough transients. The reverberation decayed fully at the end of the loop, such that even small differences in the decay behavior could be detected. Participants were explicitly encouraged to limit the loop range during the test such that they can concentrate on small details, and the test conductor noticed that every participant focused on the last reverberation decay. Additionally, there was a training phase, where participants could get familiar with the interface and the type of performance differences. Although these instructions might result in an overly critical evaluation, it is assumed that these helped detect differences that might have otherwise been left undiscovered with a small sample size.

The reference to compare against (*REF*) is the unprocessed measurement with the maximum SNR achieved by averaging six SRIR measurements obtained from 10-s sweeps (measured broadband $\mathrm{SNR}_{\mathrm{in}} = 61.38\,\mathrm{dB}$). The items *C1*−*C6* are the proposed algorithm operating under decreasing SNR conditions, achieved by adding varying levels of recorded noise to the input signal. Besides the microphone self noise, the recorded noise included a quite prominent air conditioning noise and high-frequency inference from other electronic devices. The conditions were as follows: *C1*, 100-dB SNR; *C2*, 60-dB SNR; *C3*, 50-dB SNR; *C4*, 40-dB SNR; *C5*, 30-dB SNR; and *C6*, 20-dB SNR. Additionally, the test included two anchor-like conditions: *A1*, 30-dB SNR without any denoising, and *A2*, 30-dB SNR without spatial processing in the resynthesis (i.e., one omni-directional beamformer) and only a single estimated slope.

3.3.2 Results

The results of the listening test are presented in Fig. 11 as individual transparent dots and violin plots, whereby the width of the violin shows the density of data, median values are presented as a white point, the interquartile range is marked using a thick black line, and the range between the lower and upper adjacent values is marked using a thin black line. Comparing to the technical evaluation in the noise-free single-slope case, the participants could detect deviations from the reference earlier than the technical measures predicted from the single-slope case. However, the median follows an order as predicted and is in good agreement with the technical measures. The participants also rated the algorithm as *excellent* when operating in the typical measurement SNR range.

Because of the low number of participants, descriptive and non-parametric statistics were favored over parametric alternatives. The Wilcoxon signed-rank test evaluates a statistically significant difference between two conditions (H_0: the tested underlying distributions are the same). The results are presented in Fig. 12, reported at a 95% confidence level. With both medians of REF and C1 at 100, the ratings suggest that subjects were not able to confidently identify the condition where the proposed method was applied, in comparison to the reference. For decreasing SNRs, the test indicates a perceptual difference to no spatial denoising treatment. The effectiveness of the proposed algorithm is highlighted by the comparison between *A1* (no processing) and *C5* (with processing), in the 30-dB SNR scenario. Here, the proposed algorithm performs better than the unprocessed case, which is supported by their statistically significant difference. Similarly, the comparison *A2* (no spatial reconstruction) to *C5* (full spatial processing) shows an advantage of incorporating the directional reconstruction of the late SRIR tail.

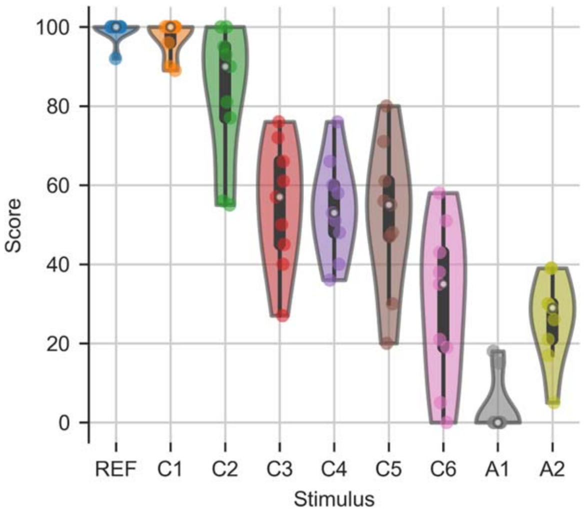

Fig. 11. Perceptual evaluation results, depicted as a violin plot. The inner part resembles a box plot, where the white point indicates the median. Each individual response is marked as a transparent dot.

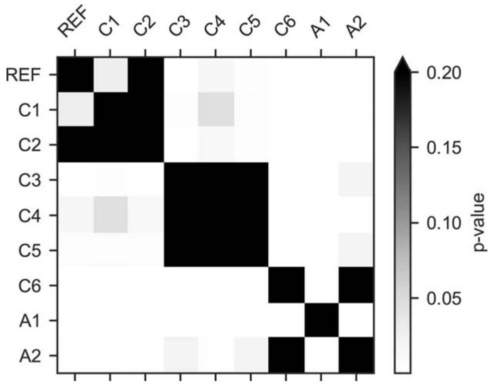

Fig. 12. Difference in condition rating scores, indicated by a Wilcoxon signed-rank test. *P*-values are clipped for better visualization.

4 DISCUSSION

This paper presented a method for SRIR tail resynthesis, including directional dependency and multi-slope decays. The importance of directional processing was justified by observing the anisotropic behavior in Figs. 1 and 2 and the multi-slope parameterization in Fig. 4. The results presented in SEC. 3.2.2 highlight promising performance. The parameter extraction, reverb tail resynthesis, and SH re-encoding delivered predictable outcomes.

The measured SRIRs obtained from an SMA exhibited prominent noise, with a distinct high-frequency component. Additionally, there were environmental noise sources, such as an air conditioning unit, in the room. The noise is audible even when listening to the SRIR alone and is not spectrally white as shown in Fig. 8(a), which means that a Gaussian noise assumption cannot be made. It also shows a frequency-dependent decay time, with rapidly vanishing high-frequency components. After the denoising procedure presented in SEC. 2, the noise was not perceivable anymore. This was confirmed by a formal listening test, which showed the timbre and tail decay can match convincingly, something that is supported by comparing the waterfall plots Fig. 8(b) and decay slopes in Fig. 4.

When measuring the input to output on the noise-free input $\mathrm{SRIR}_{\mathrm{sim}}$, the performance could also be quantified. It showed that the SNR could consistently be improved upon, up to the measurable limit. However, the denoising process was carried out frequency-dependently; hence, the mean spectral error $\overline{E}_{\mathrm{spec}}$ quantified changes in timbre. It shows that $\overline{E}_{\mathrm{spec}}$ stays relatively low until around −30-dB noise-floor SNR. This corresponded well to the point where a slight shift in timbre was audible when listening to the reverberation tails. Compared to Fig. 8, it shows that the high frequencies drop rapidly in the same range. At such low SNR conditions, the tail is buried in noise, and a meaningful parameter estimation is not possible. Nevertheless, Fig. 9 also shows the excellent matching between input and output spectra, where the deviations in the lower-frequency region might originate from different mode frequencies during resynthesis.

The effect of mismatched reverberation times becomes audible in the same range of additive noise. Around −30-dB SNR from the noise floor, a slight double decay starts to become audible. Again, this affects high frequencies first. The parameter estimation becomes harder on flattened EDCs, generally resulting in overestimating the corresponding T_{60}, even though the parameterization detects the noise floor.

The findings of the experiment support the informal findings from the measured $\mathrm{SRIR}_{\mathrm{meas}}$, where a controlled evaluation is not possible because there is no noise-free comparison SRIR. The broadband SNR of this SRIR is the range in which the proposed denoising algorithm could be justified by the previous systematic evaluation.

Ultimately, the denoising algorithm replaces parts of the SRIR. The lower the input SNR, the larger the interference with the original SRIR. The perceptual evaluation showed that the algorithm can operate almost transparently and that denoising may improve the perceived audio quality. However, replacing significant portions of the impulse response is audible when compared to a reference. The test also highlighted the importance of following the spatial characteristics of the SRIR and showed that the spatial aspect of the described method improves the perceptual performance.

It should be noted that the performance is dependent on the actual input conditions. Nevertheless, the results indicate reliable performance, as long as the input EDC can be considered sufficiently above the noise floor. The experienced performance limitations in this work seemed to be related to the parameter extraction. Further information on the temporal evolution of an impulse response was not yet incorporated. The current framework only acts on the measured input parameters, without inference or convergence assumptions. The latter additional assumptions may improve the performance on critical input SNR conditions, where parameters are not retractable anymore.

Although the modal resynthesis allowed for a straightforward implementation according to the multi-slope IR model, the resynthesis to the SHD introduced additional complexity. A formulation directly in the SHD could provide an appropriate alternative here.

Compared to previous work (e.g., [19]), using fewer beamformers might improve directional independence of the parameterization and synthesis computation time. However, the presented SFB relies on a uniform spatial sampling and quadrature weights, accordingly. Future formulations may allow non-uniform sampling, supporting more minimal spatial sampling strategies. The spatial weighting of the chosen spatial Butterworth filter pattern, however, seemed to provide sufficient suppression for all examined SRIRs in this work so that the spatial interaction was not identified as a limiting factor.

5 CONCLUSION

This paper has presented a method for resynthesizing the late decay of an SRIR, whereby an SFB extracts directionally constrained signals that are then analyzed and described as multiple exponential decays and a noise floor. The late reverberation is then resynthesized for each directional component signal using modal synthesis without the noise floor, and the SRIR is reconstructed with restored late reverberation.

The method has been evaluated both technically and perceptually, using both synthesized and measured SRIRs. The technical evaluation showed that the resynthesis technique is effective for SNRs as low as 40 dB, but at 30 dB and lower, it produced limited results because significant portions of the evaluated IR disappeared in the noise floor. The perceptual evaluation showed that the method was successful, although artefacts were audible for denoised SRIRs with low SNRs. Future work could include more elaborate parameterization in very low SNR conditions, i.e., by inferring parts that disappear within the noise floor from parts that are retractable.

Further resources of this study are available online.[2]

6 ACKNOWLEDGMENT

This research was supported by the Academy of Finland and the Human Optimised XR (HumOR) Project. The authors thank Pierre Massé for discussing and sharing his work on directional SRIR denoising. The authors also thank all listening test participants. In particular, the authors would like to thank the anonymous reviewers for their effort and the feedback they provided to improve this article.

7 REFERENCES

[1] C. F. Eyring, "Reverberation Time Measurements in Coupled Rooms," *J. Acoust. Soc. Am.*, vol. 3, no. 2A, pp. 181–206 (1931 Oct.). https://doi.org/10.1121/1.1915555.

[2] H. Kuttruff, *Room Acoustics* (CRC Press, Boca Raton, FL, 2000), 4th ed.

[3] A. Billon, V. Valeau, A. Sakout, and J. Picaut, "On the Use of a Diffusion Model for Acoustically Coupled Rooms," *J. Acoust. Soc. Am.*, vol. 120, no. 4, pp. 2043–2054 (2006 Oct.). https://doi.org/10.1121/1.2338814.

[4] M. Ermann, "Double Sloped Decay: Subjective Listening Test to Determine Perceptibility and Preference," *Build. Acoust.*, vol. 14, no. 2, pp. 91–107 (2007 Jun.). https://doi.org/10.1260/135101007781448055.

[5] P. Luizard, B. F. G. Katz, and C. Guastavino, "Perceptual Thresholds for Realistic Double-Slope Decay Reverberation in Large Coupled Spaces," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. 75–84 (2015 Jan.). https://doi.org/10.1121/1.4904515.

[6] B. Alary, P. Massé, S. J. Schlecht, M. Noisternig, and V. Välimäki, "Perceptual Analysis of Directional Late Reverberation," *J. Acoust. Soc. Am.*, vol. 149, no. 5, pp. 3189–3199 (2021 May). https://doi.org/10.1121/10.0004770.

[7] C. M. Harris and H. Feshbach, "On the Acoustics of Coupled Rooms," *J. Acoust. Soc. Am.*, vol. 22, no. 5, pp. 572–578 (1950 Sep.). https://doi.org/10.1121/1.1906653.

[8] N. Xiang, J. Escolano, J. M. Navarro, and Y. Jing, "Investigation on the Effect of Aperture Sizes and Receiver Positions in Coupled Rooms," *J. Acoust. Soc. Am.*, vol. 133, no. 6, pp. 3975–3985 (2013 Jun.). https://doi.org/10.1121/1.4802740.

[9] J. Catic, S. Santurette, and T. Dau, "The Role of Reverberation-Related Binaural Cues in the Externalization of Speech," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 1154–1167 (2015 Aug.). https://doi.org/10.1121/1.4928132.

[10] A. Kuusinen and T. Lokki, "Wheel of Concert Hall Acoustics," *Acta Acust. united Acust.*, vol. 103, no. 2, pp. 185–188 (2017 Feb.). https://doi.org/10.3813/AAA.919046.

[11] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10 (1973 Feb.).

[12] M. A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," presented at the *50th Convention of the Audio Engineering Society* (1975 Mar.), paper L-20.

[13] T. McKenzie, S. J. Schlecht, and V. Pulkki, "Acoustic Analysis and Dataset of Transitions Between Coupled Rooms," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 481–485 (Toronto, Canada) (2021 Jun.). https://doi.org/10.1109/ICASSP39728.2021.9415122.

[14] J.-D. Polack, X. Meynial, and V. Grillon, "Auralization in Scale Models: Processing of Impulse Response," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 939–945 (1993 Nov.).

[15] J.-M. Jot, G. Vandernoot, and O. Warusfel, "Analysis and Synthesis of Room Reverberation in the Time and Frequency Domains—Application to the Restoration of Room Impulse Responses Corrupted by Measurement Noise," *J. Acoust. Soc. Am.*, vol. 99, no. 4, pp. 2530–2574 (1996 Apr.). https://doi.org/10.1121/1.415793.

[16] J.-M. Jot, L. Cerveau, and O. Warusfel, "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model," presented at the *103rd Convention of the Audio Engineering Society* (1997 Sep.), paper 4629.

[17] M. Noisternig, T. Carpentier, T. Szpruch, and O. Warusfel, "Denoising of Directional Room Impulse Responses Measured With Spherical Microphone Arrays," in *Proceedings of the 40th Fortschritte der Akustik (DAGA)*, pp. 600–601 (Oldenburg, Germany) (2014 Mar.).

[18] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Denoising Directional Room Impulse Responses With Spatially Anisotropic Late Reverberation Tails," *Appl. Sci.*, vol. 10, no. 3, paper 1033 (2020 Feb.). https://doi.org/10.3390/app10031033.

[19] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "A Robust Denoising Process for Spatial Room Impulse Responses With Diffuse Reverberation Tails," *J. Acoust. Soc. Am.*, vol. 147, no. 4, pp. 2250–2260 (2020 Apr.). https://doi.org/10.1121/10.0001070.

[20] P. Massé, T. Carpentier, O. Warusfel, and M. Noisternig, "Measurement, Analysis, and Denoising of Directional Room Impulse Responses in Complex Spaces," in *Proceedings of the Forum Acusticum* (Lyon, France), pp. 137–144 (2020 Dec.).

[21] C. Hold, A. Politis, L. McCormack, and V. Pulkki, "Spatial Filter Bank Design in the Spherical Harmonic Domain," in *Proceedings of the 29th European Signal Processing Conference (EUSIPCO)*, pp. 106–110 (Dublin, Ireland) (2021 Aug.). https://doi.org/10.23919/EUSIPCO54536.2021.9616091.

[22] G. Götz, R. Falcón Pérez, S. J. Schlecht, and V. Pulkki, "Neural Network for Multi-Exponential Sound Energy Secay Analysis," https://arxiv.org/abs/2205.09644 (2021).

[2] Some audio samples and supplementary materials are included on a companion page: http://research.spa.aalto.fi/publications/papers/jaes-anisotropic-multislope-SRIR-resynthesis/.

[23] ISO, "Acoustics—Measurement of Room Acoustic Parameters—Part 1: Performance Spaces," *Standard 3382-1* (2009 Jun.).

[24] J. Balint, F. Muralter, M. Nolan, and C.-H. Jeong, "Bayesian Decay Time Estimation in a Reverberation Chamber for Absorption Measurements," *J. Acoust. Soc. Am.*, vol. 146, no. 3, pp. 1641–1649 (2019 Sep.). https://doi.org/10.1121/1.5125132.

[25] E. G. Williams, "Cylindrical Waves," in E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, pp. 115–148 (Academic Press, London, UK, 1999). https://doi.org/10.1016/B978-0-12-753960-7.X5000-1.

[26] B. Rafaely, *Fundamentals of Spherical Array Processing*, Springer Topics in Signal Processing, vol. 16 (Springer, Cham, Switzerland, 2019). https://doi.org/10.1007/978-3-319-99561-8.

[27] B. Devaraju, *Understanding Filtering on the Sphere*, Ph.D. thesis, University of Stuttgart, Stuttgart, Germany (2015 Aug.). https://doi.org/10.18419/opus-3985.

[28] R. H. Hardin and N. J. Sloane, "McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions," *Discrete Comput. Geom.*, vol. 15, no. 4, pp. 429–441 (1996 Apr.). https://doi.org/10.1007/BF02711518.

[29] C. Hold, S. J. Schlecht, A. Politis, and V. Pulkki, "Spatial Filter Bank in the Spherical Harmonic Domain: Reconstruction and Application," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 361–365 (New Paltz, NY) (2021 Oct.). https://doi.org/10.1109/WASPAA52581.2021.9632709.

[30] M. R. Schroeder, "New Method of Measuring Reverberation Time," *J. Acoust. Soc. Am.*, vol. 37, no. 3, pp. 409–412 (1965 Mar.). https://doi.org/10.1121/1.1909343.

[31] ISO, "Acoustics—Measurement of Room Acoustic Parameters—Part 2: Reverberation Time in Ordinary Rooms," *Standard 3382-2* (2008 Jun.).

[32] N. Xiang and P. M. Goggans, "Evaluation of Decay Times in Coupled Spaces: Bayesian Parameter Estimation," *J. Acoust. Soc. Am.*, vol. 110, no. 3, pp. 1415–1424 (2001 Aug.). https://doi.org/10.1121/1.1390334.

[33] N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of Decay Times in Coupled Spaces: Reliability Analysis of Bayesian Decay Time Estimation," *J. Acoust. Soc. Am.*, vol. 117, no. 6, pp. 3707–3715 (2005 May). https://doi.org/10.1121/1.1903845.

[34] M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of Modal Decay Parameters From Noisy Response Measurements," *J. Audio Eng. Soc.*, vol. 50, no. 11, pp. 867–878 (2002 Nov.).

[35] A. Lindau, L. Kosanke, and S. Weinzierl, "Perceptual Evaluation of Physical Predictors of the Mixing Time in Binaural Room Impulse Responses," presented at the *128th Convention of the Audio Engineering Society* (2010 Jan.), paper 8089.

[36] N. Epain and C. T. Jin, "Spherical Harmonic Signal Covariance and Sound Field Diffuseness," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 10, pp. 1796–1807 (2016 Oct.). https://doi.org/10.1109/TASLP.2016.2585862.

[37] M. Schoeffler, S. Bartoschek, F.-R. Stöter, et al., "webMUSHRA — A Comprehensive Framework for Web-Based Listening Tests," *J. Open Res. Softw.*, vol. 6, no. 1, paper 8 (2018 Feb.). https://doi.org/10.5334/jors.187.

[38] B. Bernschütz, "A Spherical Far-Field HRIR Compilation of the Neumann KU100," in *Proceedings of the 39th Fortschritte der Akustik (DAGA)*, pp. 592–595 (Merano, Italy) (2013 Mar.). https://doi.org/10.5281/zenodo.3928297.

[39] C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, and I. J. Tashev, "Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 261–265 (Brighton, UK) (2019 May). https://doi.org/10.1109/ICASSP.2019.8683751.

THE AUTHORS

Christoph Hold

Thomas McKenzie

Georg Götz

Sebastian J. Schlecht

Ville Pulkki

Christoph Hold is a doctoral candidate in the Department of Signal Processing and Acoustics at Aalto University, Finland, focusing on spatial audio processing. He received an M.Sc. in audio communication technology in 2019 and B.Sc. in electrical engineering from the Technische Universität Berlin (TU Berlin), where he specialized in signal processing and virtual acoustics. From 2015 to 2017, he was a research assistant at TU Berlin, followed by two research internships (2017 and 2018) at Microsoft Research in Redmond, WA. He is interested in high-quality audio and its perception. For the Audio Engineering Society, he was the chair of the Berlin Student Section and part of the 142nd AES Convention committee.

Thomas McKenzie is a postdoctoral researcher in the Department of Signal Processing and Acoustics at Aalto University, where he studies room acoustics and six degrees-of-freedom spatial audio. He completed a B.Sc. in Music, Multimedia, and Electronics at the University of Leeds, UK, in 2013, before completing his M.Sc. in Postproduction with Sound Design and Ph.D. in Music Technology at the University of York, UK, in 2015 and 2020, respectively. His research interests include spatial audio and psychoacoustics.

Georg Götz is a doctoral candidate at the Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Finland. He received his B.Sc. (2016) and M.Sc. (2019) degrees in Media Technology from Ilmenau University of Technology, specializing in spatial audio and psychoacoustics. He was the deputy chair of the AES Student Section Ilmenau from 2013 to 2016. The research of his doctoral thesis is about machine-learning–based virtual acoustics rendering. His other research interests include room acoustics, psychoacoustics, and virtual reality/augmented reality technology.

Sebastian J. Schlecht is a Professor of Practice for Sound in Virtual Reality at the Acoustics Lab, Department of Signal Processing and Acoustics, and Media Lab, Department of Art and Media, of Aalto University, Espoo, Finland. He received a Diploma in Applied Mathematics from the University of Trier, Germany, in 2010 and M.Sc. degree in Digital Music Processing from the School of Electronic Engineering and Computer Science at Queen Mary University of London, UK, in 2011. In 2017, he received a Doctoral degree at the International Audio Laboratories Erlangen, Germany, on artificial spatial reverberation and reverberation enhancement systems. From 2012 to 2019, Dr. Schlecht was also external research and development consultant and lead developer of the 3D Reverb algorithm at the Fraunhofer IIS, Erlangen, Germany.

Ville Pulkki is a professor in the Department of Signal Processing and Acoustics at Aalto University, Helsinki, Finland. He has been working in the field of spatial audio for over 20 years. He developed the vector-base amplitude panning (VBAP) method in his Ph.D. (2001) and directional audio coding after the Ph.D. with his research group. He also has contributions in perception of spatial sound, laser-based measurement of room responses, and binaural auditory models. He has received the Samuel L. Warner Memorial Medal Award from the Society of Motion Picture and Television Engineers and the AES Silver Medal Award. He enjoys being with his family, building his summer house, and performing in musical ensembles.