PARAMETRIC RESYNTHESIS OF MEASURED SPATIAL ROOM IMPULSE RESPONSES

Anthony Gallien, Benoit Alary, Markus Noisternig

Sorbonne Université,
STMS, IRCAM, CNRS, Ministère de la Culture
Paris, France
anthony.gallien@ircam.fr

ABSTRACT

Spatial Room Impulse Responses (SRIRs) are fundamental to immersive audio rendering and have become a key focus of recent machine learning research in acoustics and auralization. Due to the high computational cost of direct convolution, spatial audio systems commonly employ artificial reverberation algorithms. However, these approaches often fail to accurately reproduce the spatial, temporal, and spectral characteristics of early reflections, leading to notable deviations from measured SRIRs. This paper presents a comprehensive framework for the analysis and efficient resynthesis of SRIRs captured with Spherical Microphone Arrays (SMAs). The proposed method accounts for hardware-induced artifacts, including scattering and spatial aliasing. Early reflections are reconstructed using a parametric approach based on the Herglotz analysis method, while late reverberation is synthesized using a Directional Feedback Delay Network (DFDN) with optimized filter-attenuation and correlation-matching. The proposed framework produces signals whose spatial correlation and Energy Decay Relief (EDR) closely match those of measured SRIRs, demonstrating its effectiveness for both real-time spatial audio rendering and realistic dataset generation for machine learning applications.

1. INTRODUCTION

In immersive and spatial audio applications, Spatial Room Impulse Responses (SRIRs) are essential for the auralization of virtual sound sources. They encode the acoustic characteristics of the environment while conveying the spatial relationship between source and receiver. SRIRs are typically captured using Spherical Microphone Arrays (SMAs) and, to reproduce the measured sound field, are encoded in the Spherical Harmonic (SH) domain within the Higher-Order Ambisonic (HOA) framework [1].

However, direct convolution of long-duration SRIRs represented in high ambisonic order is computationally demanding. To enable real-time processing, parametric analysis and reproduction techniques are commonly employed [2, 3]. Late reverberation is traditionally modeled using Feedback Delay Networks (FDNs), with the concept of deriving delay-network parameters from reverberation time analysis first introduced by Jot [2]. Subsequent work has led to substantial improvements in both delay-network design [4, 5] and room impulse response analysis [6, 7], enabling increasingly accurate reproduction pipelines [3]. Accurate synthesis with delay networks critically depends on robust recursive filter design, which can be facilitated through optimized parallel equalization [8, 9] and multi-stage filtering strategies [10, 11]. Alternatively, velvet-noise-based methods have been proposed for RIR synthesis without explicit reverberation time analysis [4, 12]. Despite these advances, conventional reproduction approaches often neglect the complex spatial structure of early reflections, which remain critical for both perceptual and objective accuracy [13, 14, 15].

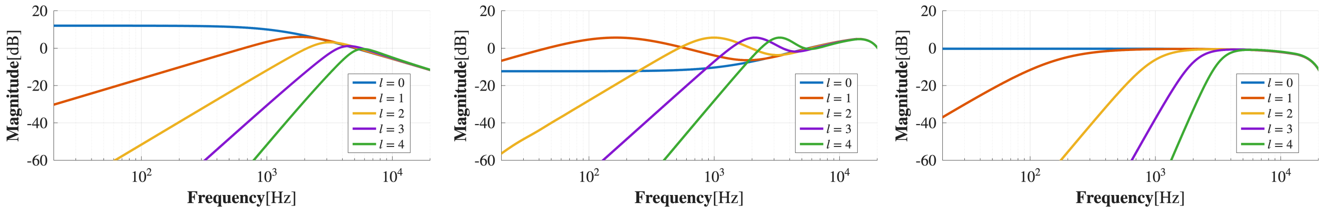
In immersive audio applications, reproducing directional characteristics is essential, as many real-world acoustic environments exhibit perceptible anisotropic reverberation [16]. Advanced architectures like Directional Feedback Delay Networks (DFDNs) address this by synthesizing direction-dependent decay rates [17, 18]. While the perceptual thresholds associated with anisotropy are not yet fully established, spatial simplifications can nevertheless be introduced to reduce rendering complexity [19, 20].

Recently, machine learning (ML) approaches have demonstrated promising results in spatial audio reproduction, including models designed to infer delay-network parameters or directly generate binaural and directional reverberation [21, 22, 23]. However, their practical deployment remains limited by a “sim-to-real” domain gap, as models are predominantly trained on idealized acoustic simulations, while inference is performed on real-world microphone array data.

This discrepancy is partly attributable to hardware-related effects. SMA-measured, HOA-encoded SRIRs exhibit spectral and spatial distortions, including rigid-sphere scattering and high-frequency spatial aliasing. Since conventional synthesis pipelines and training datasets typically neglect these measurement artifacts, the resulting mismatch can significantly degrade both auralization quality and downstream ML inference.

To address these limitations, this paper proposes a comprehensive analysis-resynthesis framework for reconstructing SMA-measured SRIRs encoded in HOA[1] (from direct sound and early reflections to late reverberation), while explicitly accounting for measurement artifacts such as remnant scattering and aliasing. The proposed method replaces conventional convolution with a highly optimized, direction-aware delay-network and a dedicated early-reflection synthesis stage, enabling computationally efficient real-time auralization for 6DoF applications. By faithfully modeling these physical imperfections, the framework not only improves spatial reproduction fidelity but also facilitates the generation of more realistic SRIR datasets that better match raw microphone signals, thereby providing a more robust basis for training next-generation ML models.

[1] Listening examples at https://dafx26-eac.ircam.fr/.



(a) Mode strengths functions $|b_\ell|$ **(b)** Encoding filters $|a_\ell|$ (Tikhonov reg. - 18 dB) **(c)** Resulting response $|c_\ell| = |a_\ell \cdot b_\ell|$

Figure 1: Magnitude frequency responses for a rigid sphere EM32 ($r_s = 4.2$ cm).

2. THEORETICAL BACKGROUND

2.1. Herglotz formalism

A key challenge in analyzing SMA-captured SRIRs lies in parameterizing early reflections, in particular, estimating the Direction of Arrival (DoA) of the direct sound and individual echoes. Classical methods such as Steered Response Power (SRP) often exhibit limited spatial resolution, while subspace-based methods such as MUSIC [24] can be adversely affected by the strong temporal correlation between the direct sound and its reflections. To address these limitations, in this paper, the analysis relies on the Herglotz wave function formalism [15] applied to HOA-encoded signals.

The Herglotz representation defines a bounded operator mapping a continuous angular density $g \in L^2(\mathbb{S}^2)$ to the measured acoustic field:

$$(Hg)(\mathbf{x}) = \int_{\mathbb{S}^2} g(\hat{\mathbf{s}})\, e^{ik\hat{\mathbf{s}}\cdot\mathbf{x}}\, d\hat{\mathbf{s}}.$$

Recovering g requires solving an underdetermined system over a dense directional grid ($D > (L+1)^2$), making it intrinsically an inverse problem. This contrasts fundamentally with standard plane-wave decomposition (PWD), which operates as a direct spatial projection problem (e.g., beamforming). To solve this ill-posed system, Tikhonov regularization is employed, acting as a spatial smoothing filter to stabilize the recovered continuous density map.

Finally, an iterative Gaussian fitting procedure is applied to parameterize this continuous density map g. By modeling the spatial spread of the acoustic energy, this approach robustly extracts both the actual number of incident echoes and their precise DoAs.

2.2. SMA Limitations

As introduced in Sec. 1, rigid-sphere SMAs (e.g., the mh acoustics Eigenmike® EM32 used in this work) inherently scatter the incident sound field. In the SH domain, this effect is characterized by the *mode strength* or *radial functions* $b_\ell(kr)$, which are defined in terms of spherical Bessel (j_ℓ) and Hankel (h_ℓ) functions [25]:

$$b_\ell(kr) = 4\pi i^\ell \left[j_\ell(kr) - \frac{j'_\ell(kr_s)}{h'_\ell(kr_s)} h_\ell(kr) \right], \tag{1}$$

where $k = \omega/c$ denotes the wavenumber and r_s is the sphere radius. Evaluating this expression on the array surface ($r = r_s$) yields a strictly frequency-dependent mode strength b_ℓ (Fig. 1a).

In principle, HOA encoding requires inversion of the mode strength, i.e., $1/b_\ell(kr_s)$. However, the vanishing magnitude of Bessel functions at low frequencies renders this inversion ill-conditioned, leading to significant noise amplification. To address this issue, regularized encoding filters $a_\ell(f)$ are designed (Fig. 1b) to approximate the inverse while balancing numerical stability and spatial accuracy [26]. The resulting effective response $c_\ell(f) = a_\ell(f) \cdot b_\ell(f)$ (Fig. 1c) exhibits a low-frequency roll-off that limits the usable range of higher SH orders and reflects residual low-frequency scattering effects.

At high frequencies, SMA performance is fundamentally constrained by spatial aliasing [27, 1]. Since acoustic fields such as plane waves admit infinite SH expansions, sampling them with a finite number of microphones causes higher-order components to be aliased into the highest representable SH order L of the SMA (e.g., $L = 4$ for the EM32).

The onset frequency of this spatial aliasing can be approximated as:

$$f_{\text{alias}} \approx \frac{cL}{2\pi r_s}, \tag{2}$$

which yields an upper limit of approximately $f_{\text{alias}} \approx 5.2$ kHz for the 4th-order EM32 array.

3. PROPOSED FRAMEWORK

As outlined in Sec. 1, the proposed framework operates in the HOA domain. The use of SH provides a continuous and orthogonal representation of the sound field that is, in principle, independent of the recording array geometry. A key advantage of this representation is the ability to derive directional views, such as PWD, through the inverse SH transform, enabling direct spatial analysis over the sphere. Accordingly, the processing pipeline begins by encoding the raw SMA signals into the HOA domain.

The pre-delay is first estimated and removed so that the direct sound defines the temporal origin of the analysis–resynthesis process. Subsequently, the mixing time, which marks the transition from spatially coherent early reflections to the stochastic late reverberation tail [28], is estimated to partition the SRIR for separate processing. This boundary can be interpreted as the time at which the spatial incoherence of the sound field converges to a steady state [29, 30]. Distinct analysis–resynthesis strategies are then applied to the early and the late components of the SRIR. Finally, the processed parts are temporally recombined using a hybrid reverberation approach [31]. The complete processing pipeline is illustrated in Fig. 2.

Throughout this section, the framework is illustrated using a single EM32 measurement (detailed in Sec. 4).

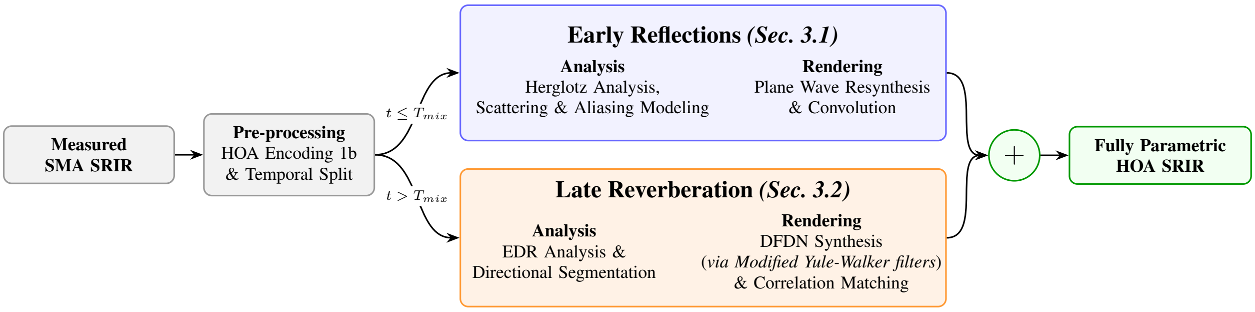


Figure 2: Flowchart of the proposed analysis and resynthesis framework.

3.1. Early reflections

To accurately estimate the Direction of Arrival (DoA) and Time of Arrival (ToA) using the Herglotz analysis (Sec. 2.1), the HOA signal is first segmented into short, overlapping 2.6-ms time windows, thereby limiting the number of simultaneous reflections within each frame. In addition, the analysis is spectrally constrained to the frequency range over which the SMA encoding remains physically reliable. This range is bounded above by the spatial aliasing frequency (Sec. 2.2), and below by the frequency at which the maximum directivity index (DI) [32] equals or exceeds its value at the aliasing frequency (i.e., $\mathrm{DI}(f_{\mathrm{low}}) \geq \mathrm{DI}(f_{\mathrm{alias}})$) [15]. Duplicate detections arising from window overlap are subsequently removed (Fig. 3).

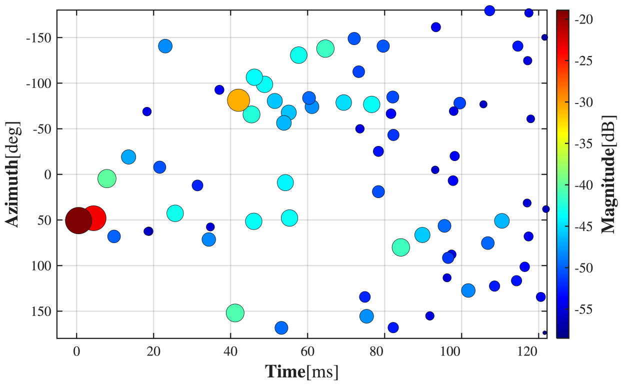


Figure 3: Direct sound (brown) and early reflections detected using the Herglotz analysis method.

Once the DoA and ToA of a reflection are estimated, a hypercardioid beam is steered toward the corresponding direction to extract its spectral content. At a sampling rate of 48 kHz, a 128-sample analysis window is used, starting 20 samples prior to the estimated ToA to preserve the transient onset.

For resynthesis, reflections are modeled as discrete plane waves. Each component is spatio-temporally aligned according to its estimated DoA and ToA, filtered using the extracted spectral content to preserve acoustic coloration, and finally re-encoded into the HOA domain at the maximum supported order.

3.1.1. Remnant Low-Frequency Scattering Modeling

While the previously described method enables near-perfect reconstruction in simulation, accurate resynthesis of real SMA measurements requires accounting for hardware-induced artifacts. As detailed in Sec. 2.2, exact inversion of the holographic function b_ℓ is not feasible, resulting in residual low-frequency scattering.

To reproduce the spectral and spatial characteristics of the SMA-measured sound field, a frequency-dependent beamforming strategy is proposed. By analyzing the effective encoded response $c_l(f) = a_l(f) \cdot b_l(f)$ (cf. Fig. 1c), cutoff frequencies $f_{\mathrm{ro}}^{(l)}$ are identified for each Ambisonic order l at the boundary of their flat-response region. Based on this observation, the beamformer adaptively selects the maximum reliable order per frequency band, resulting in progressively wider beam patterns at lower frequencies. The resulting broadband beamformer output, $\mathcal{B}_{tot}(f, \Omega_0)$, steered toward the estimated DoA Ω_0, is defined as a weighted sum of order-dependent beams:

$$\mathcal{B}_{tot}(f, \Omega_0) = \sum_{l=0}^{L} H_l(f) \mathcal{B}_l(f, \Omega_0), \tag{3}$$

where L denotes the maximum available Ambisonic order. To avoid the severe pre- and post-ringing associated with ideal brick-wall filters, relaxed crossover slopes are employed, similar to the approach in [26]. The bandpass crossover function $H_l(f)$ for each order l is defined as the difference between adjacent zero-phase low-pass filters H_{LP}:

$$H_l(f) = H_{\mathrm{LP}}^{(l+1)}(f) - H_{\mathrm{LP}}^{(l)}(f), \tag{4}$$

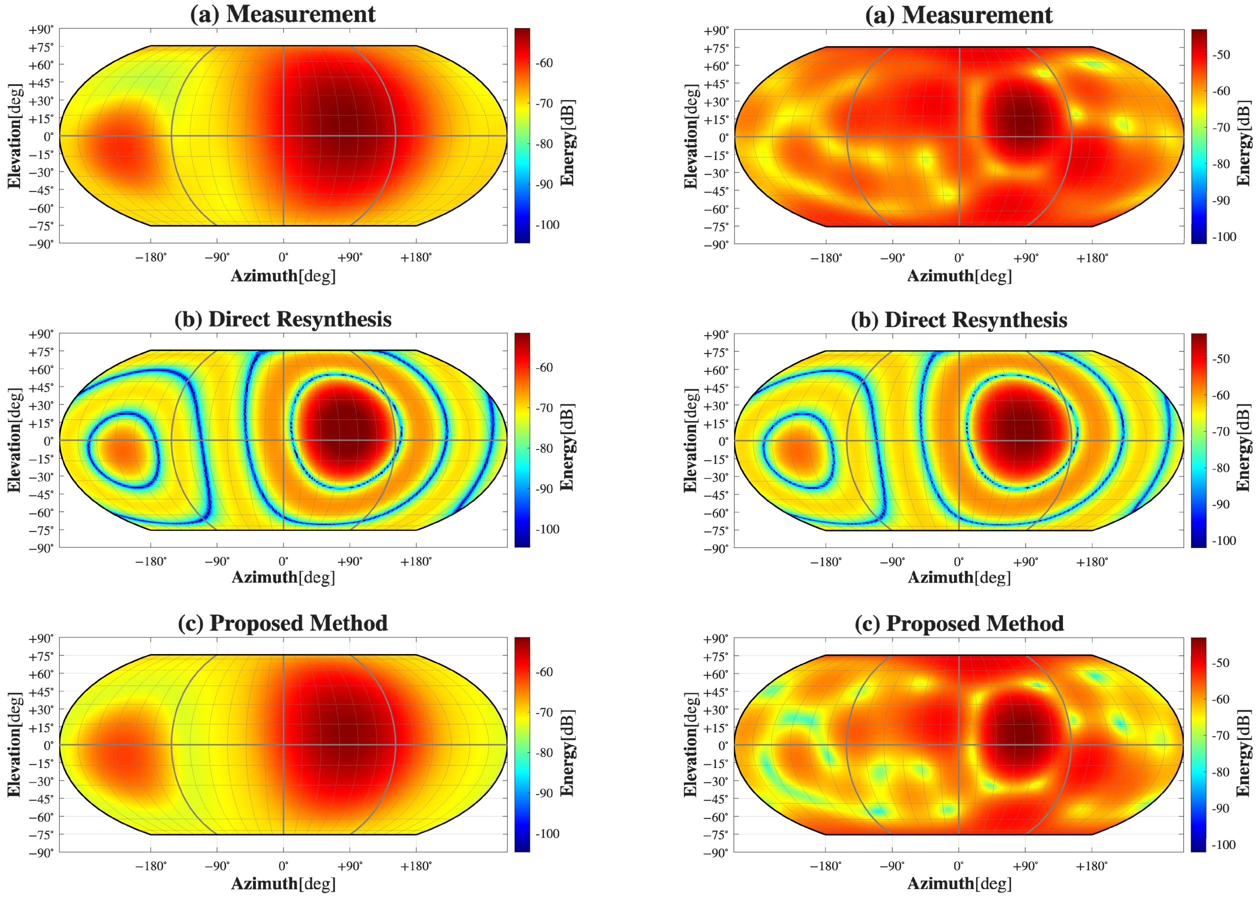
where $H_{\mathrm{LP}}^{(l)}(f)$ has a cutoff frequency of $f_{\mathrm{ro}}^{(l)}$. The boundary conditions are set to $H_{\mathrm{LP}}^{(0)}(f) = 0$ and $H_{\mathrm{LP}}^{(L+1)}(f) = 1$, ensuring perfect amplitude reconstruction.

During resynthesis, each spectral component is re-encoded into the HOA domain at its corresponding extraction order. Due to the inherently broader directivity of lower-order beams, spatial overlap between distinct early reflections may occur. This redundancy is mitigated in a subsequent processing stage. Finally, Fig. 4a illustrates the effectiveness of the proposed approach compared to a direct resynthesis that neglects residual scattering effects.

3.1.2. High-Frequency Spatial Aliasing Modeling

To accurately reproduce high-frequency spatial aliasing (Sec. 2.2), the physical response of the array to the incident plane wave reflections is explicitly modeled. Following Rafaely's framework [1], the spatial aliasing matrix $\mathbf{E}$ is defined as:

$$\mathbf{E} = \mathbf{Y}_{\mathrm{Meas}}^{\dagger} \mathbf{Y}_{\mathrm{PW}}, \tag{5}$$



(a) Low-frequency band (20 Hz to 2 kHz).

(b) High-frequency band ($f_{\text{alias}} \approx 5.2$ kHz to 20 kHz).

Figure 4: Measured and resynthesized direct sound energy by directions.

where $\mathbf{Y}^{\dagger}_{\text{Meas}}$ denotes the pseudoinverse of the SH sampling matrix corresponding to the SMAs geometry up to its maximum supported order L, and $\mathbf{Y}_{\text{PW}}$ is the SH matrix of an ideal plane wave expanded up to order $\tilde{L}$. To ensure coverage of the full audible bandwidth ($f_{\text{alias}} \geq 20$ kHz), an expansion order of $\tilde{L} \geq 16$ is required according to Eq. 2.

Applying $\mathbf{E}$ to the ideal plane-wave coefficients associated with the estimated DoA yields the aliased response. A spatial aliasing map is then obtained by subtracting the direct contribution of the true DoA, therby isolating the aliasing artifacts.This procedure is possible because the geometry of the specific SMA is known. By applying Eq. (5), we simulate how the array physically projects energy into false spatial directions.

Given the strong frequency dependence of spatial aliasing, this map is computed in third-octave bands spanning from the array's aliasing frequency up to 20 kHz. A peak-search algorithm is subsequently applied to identify dominant artifact directions (i.e., aliasing zones). The spectral content of these components is extracted by steering a hypercardioid beam toward each zone. Finally, these "phantom" sources are resynthesized as discrete, filtered plane waves, thereby completing the high-frequency reconstruction (Fig. 4b).

3.2. Late reverberation

This section addresses the analysis and resynthesis of late reverberation, defined as the portion of the SRIR occurring after the mixing time. The analysis is based on a time-frequency representation using the Energy Decay Relief (EDR) of the impulse response. For each frequency bin, the measurement noise floor and the early reflection segments are first estimated. A linear regression is then performed on the decay curve, approximated as an exponential, within the time interval defined by these bounds. The key parameters estimated from this process are the frequency-dependent reverberation time $\text{RT}_{60}(f)$, and the initial power spectrum $P_0(f)$. For resynthesis, an FDN is employed to model diffuse late reverberation, while a DFDN is used when anisotropic decay characteristics are present.

3.2.1. Improved Attenuation Filters

In an FDN architecture, control over the frequency-dependent RT_{60} is achieved by inserting attenuation filters G within the feedback loop [8]. Given an estimated $\text{RT}_{60}(f)$ profile, the required target gain (in dB) for the attenuation filter of the i-th delay line is derived as:

$$G_{i,\text{dB}}(f) = -60 \cdot \frac{M_i}{f_s \cdot \text{RT}_{60}(f)}, \tag{6}$$

where f_s denotes the sampling rate, and M_i is the length (in samples) of the i-th delay line. The factor M_i accounts for the fact that the attenuation filter is applied only once per loop traversal.

Before the filter design, the measured RT_{60} profiles are smoothed using a fractional 1/3-octave band filter to obtain a perceptually motivated frequency representation.

Low-order IIR filters based on the modified Yule-Walker method [33] are then used to approximate the target attenuation. This ARMA (auto regressive and moving average) based approach estimates the filter parameters from the signal's energy structure, interpreted as a power spectral density, and naturally yields minimum-phase realizations. It provides an effective trade-off between spectral accuracy and computational efficiency.

The recursive structure of the FDN produces reverberated signals matching the target frequency-dependent decay profile $RT_{60}(f)$, and an output spectral shaping filter is added to match the initial power spectrum of the decay $P_0(f)$. To this end, the modified Yule-Walker method is again employed to design the output filter.

From a computational perspective, a fourth-order IIR filter requires fewer operations than a second-order biquad graphic equalizer with two bands, while also avoiding the constraints of fixed center frequencies and the associated inter-band ripple.

3.2.2. Spatial Correlation Matching

Late reverberation is first decomposed into directional components using PWD. While ideal, fully diffuse reverberation can be efficiently synthesized with a single FDN generating uncorrelated outputs, real SMA measurements exhibit residual inter-channel correlation due to physical constraints [34, 35]. To achieve a spatially faithful rendering, this correlation must be extracted from the measurements and imposed into the synthesized FDN outputs.

To achieve this, both the PWD-derived measurement $\mathbf{x}$ and the raw FDN outputs $\mathbf{z}$ are decomposed into third-octave bands using a perfect reconstruction filter bank. For each band b, the spatial covariance matrix $\mathbf{C}_b = \text{Cov}(\mathbf{x}^{(b)})$ is computed to capture the measured spatial energy distribution. (Note that $\mathbf{x}$ and $\mathbf{z}$ are real matrices of dimension $[(L_{max}+1)^2 \times N]$, where N denotes the number of samples in the late SRIR.)

The spatial structure is then transferred via a mixing matrix $\mathbf{M}_b$. Using the eigenvalue decomposition of the positive semi-definite covariance matrix, $\mathbf{C}_b = \mathbf{V}_b \mathbf{D}_b \mathbf{V}_b^T$, we define:

$$\mathbf{M}_b = \mathbf{V}_b \mathbf{D}_b^{1/2} \mathbf{V}_b^T, \tag{7}$$

where $\mathbf{V}_b$ contains the orthogonal eigenvectors and $\mathbf{D}_b^{1/2} = \text{diag}(\sqrt{\lambda_i})$ is the diagonal matrix of square-rooted eigenvalues λ_i, yielding the principal square root of $\mathbf{C}_b$.

Concurrently, the raw FDN outputs $\mathbf{z}^{(b)}$ are normalized by their standard deviations to produce a whitened signal $\tilde{\mathbf{z}}^{(b)}$ with unit variance. The desired spatial correlation is imposed by:

$$\hat{\mathbf{z}}^{(b)} = \mathbf{M}_b \tilde{\mathbf{z}}^{(b)}. \tag{8}$$

This procedure is repeated across all frequency bands. The resulting spatially correlated components $\hat{\mathbf{z}}^{(b)}$ are recombined to form the broadband signal, which is then re-encoded into the HOA domain. A detailed mathematical derivation of the mixing matrix is detailed in Appendix A.

3.2.3. Efficient Anisotropic Reverberation via Directional Segmentation

Anisotropic late reverberation is resynthesized using DFDNs [17], which reproduce direction-dependent decay profiles while producing inherently uncorrelated outputs. To ensure spatial fidelity, these outputs are subsequently processed with the correlation-matching procedure described in the previous section.

Although DFDNs are more efficient than direct convolution for rendering medium to long reverberation tails, running 25 parallel delay networks to cover the full 4th-order Ambisonics remains computationally demanding for resource-constrained devices. To address this, we refine a frequency-dependent directional segmentation strategy that clusters perceptually similar acoustic regions, thereby reducing CPU load while preserving the overall spatial accuracy.

Extending the segmentation algorithm proposed in [36, 20], our approach considers frequency-dependent reverberation time ($\text{RT}_{60}(f)$) across the Directional Room Impulse Responses (DRIRs).

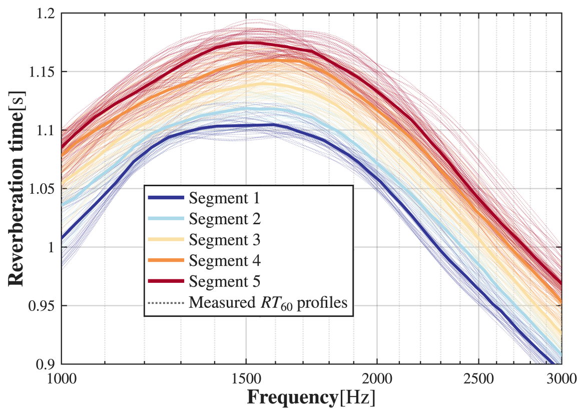


Figure 5: The decay profiles of individual directions (dotted lines) and the median curves of the five segments (thick lines).

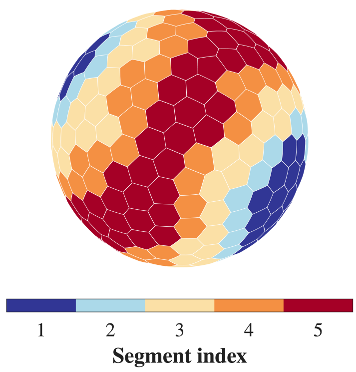


Figure 6: Distribution of the five segments around the sphere, where the indexes increase with the reverberation time.

A spectrally weighted sum, emphasizing perceptually critical frequencies via an exponential weighting, is used to iteratively partition the spatial data until the desired number of segments is achieved. Each segment is then assigned a single independent

FDN to synthesize all enclosed PWD directions, significantly lowering computational cost.

Fig. 5 shows the median RT_{60} profiles of each segment (solid lines) overlaid on the individual directional profiles (dotted lines), while Fig. 6 depicts their spatial distribution. For high-resolution visualization, 289 PWD directions are displayed.

Importantly, this segmentation applies only to the reverberation time. The initial decay power P_0 remains fully direction-dependent for every PWD channel. Since applying a per-direction output filter is computationally equivalent to using a shared segment filter, preserving the unsegmented P_0 maintains the exact anisotropic energy distribution of the original measurement without compromising real-time performance.

4. EVALUATION

The proposed framework has been evaluated on multiple datasets; here, results are presented for EM32 SRIR measurements conducted in the variable-acoustics IRCAM Espro hall. The late reverberation was synthesized using a 5-segment DFDN, where each FDN is composed of 32 delay lines with lengths randomly chosen between 30 and 70 ms and the recirculating matrices are generated as random orthogonal.

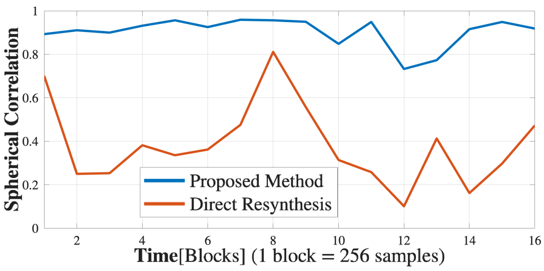
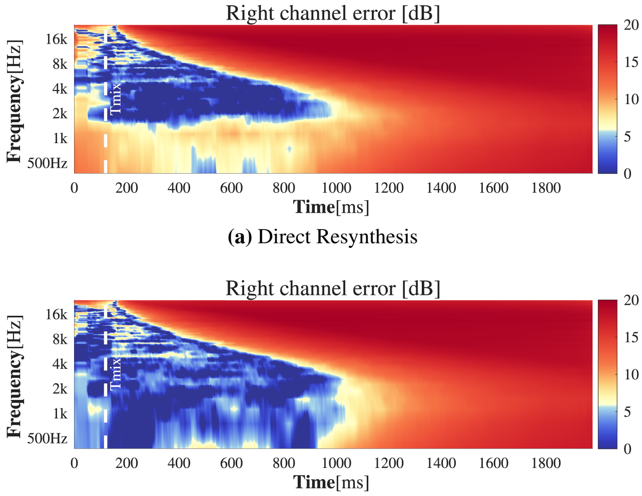


Figure 7: Spherical correlation analysis of early reflections between measurement and resynthesis over time (block size = 5.3 ms).

Fig. 7 evaluates the spatial accuracy of reconstructed early reflections by tracking the spherical correlation [37] between the original measurement and the resynthesized signals in 5.3 ms time blocks. As depicted, the direct resynthesis, using ideal plane waves without modeling rigid sphere scattering and spatial aliasing, exhibits substantial drops in correlation across multiple time blocks, indicating a significant deviation from the true spatial structure. In contrast, the proposed method maintains consistently high correlation, closely matching the ground truth. This comparison highlights the importance of explicitly modeling low-frequency scattering and high-frequency aliasing (Sec. 3.1) to preserve the spatial characteristics of SMA-measured early reflections.

Fig. 8 further evaluates the spatio-spectral fidelity across the full SRIR using EDR differences in dB between the binaurally encoded measurement and two resynthesis conditions. Fig. 8a shows the direct resynthesis baseline (see also Fig. 7-a), highlighting significant errors (yellow and orange regions) throughout the decay.

In contrast, Fig. 8b illustrates the improved performance of the proposed method. For early reflections ($t \leq T_{mix}$), the error is markedly reduced, confirming that low-frequency scattering and high-frequency aliasing emulation restore the acoustic coloration of the HOA-encoded measurement. In the late reverberation tail



(a) Direct Resynthesis

(b) Proposed Method

Figure 8: EDR difference [dB] between binaurally encoded measurement and (a) direct resynthesis (early reflections + uncorrelated FDNs), versus (b) proposed method.

($t > T_{mix}$), the error is similarly minimized, demonstrating that applying the covariance matrix to DFDN outputs effectively reproduces the measured (target) spatial correlation. Low errors across valid time-frequency regions also confirm that the denoising procedure successfully preserves the integrity of the acoustic decay. Large deviations (red regions) appear only where the measurement reaches the noise floor, whereas the proposed framework synthesizes a noise-free exponential tail.

5. DISCUSSION

The proposed framework substantially improves the spatial and spectral fidelity of resynthesized SRIRs, yet certain limitations remain. For instance, the Herglotz analysis cannot identify every individual reflection due to the finite maximum encoding order of the SMA. Furthermore, the late reverberation model assumes a single-slope exponential decay per directional band. Complex acoustic environments exhibiting multi-slope decays, such as coupled rooms, are not addressed in this study. Nonetheless, the modular nature of the framework allows straightforward integration of specialized multi-slope delay networks [38] in future works. By coupling the reproduction of physical SMA artifacts with flexible parametric control over the full SRIR, the framework shows strong potential for generating large-scale, realistic datasets for ML applications.

Beyond resynthesis, the proposed pipeline can act as a spatial and spectral SMA enhancement tool. By explicitly estimating the DoAs of early reflections, modeling scattering and spatial aliasing, and reassigning uncorrupted spectral content to its theoretical spherical coordinates, the system effectively refines the measured data. Furthermore, the late reverberation analysis is decoupled from the physical order of the SMA, allowing reverberation time extraction to be performed over a much denser spatial grid [20]. However, future work must address the energy overlap between adjacent beams to prevent power overestimation. It must also be acknowledged that while this representation is refined, the intrinsic spatial resolution limit of the original SMA measurement is not artificially increased. Ultimately, this parametric approach allows the resulting SRIR to be rendered for any arbitrary loudspeaker layout. Conceptually, this aligns with the framework proposed in [39].

6. CONCLUSION

This paper presents a complete framework for the analysis and efficient resynthesis of SRIRs captured by SMAs, explicitly compensating for hardware-induced artifacts such as low-frequency scattering and high-frequency spatial aliasing. Early reflections are parametrically reconstructed, and late reverberation is rendered via optimized, segment-based DFDNs. This approach significantly reduces computational cost while preserving the anisotropic characteristics of the acoustic space. The framework enables both highly accurate resynthesis and real-time 6DoF auralization. It is also suitable for generating ML training datasets, offering greater parametric flexibility than traditional geometric acoustic simulations. Models trained on this data are expected to achieve improved inference performance on real-world SMA measurements. Future work will include perceptual listening tests to formalize the thresholds of spatial approximations and validate the auralization quality.

7. ACKNOWLEDGMENTS

This work was partially supported by a national grant from the *Agence Nationale de la Recherche* named France 2030 ANR-23-IACL-0007 and by the France 2030 CONTINUUM project. The authors thank Olivier Warusfel and Nicolas Obin for insightful discussions, as well as all the acoustic team members who contributed to the measurement campaigns.

8. REFERENCES

[1] B. Rafaely, *Fundamentals of Spherical Array Processing*, 2nd ed. Cham, Switzerland: Springer Topics in Signal Processing, 2019.

[2] J.-M. Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, San Francisco, CA, USA, Mar. 1992, pp. 221–224.

[3] G. D. Santo, B. Alary, K. Prawda, S. J. Schlecht, and V. Välimäki, "RIR2FDN: An improved room impulse response analysis and synthesis," in *Proc. 27th Int. Conf. Digit. Audio Effects (DAFx24)*, Guildford, UK, Sep. 2024.

[4] V. Välimäki, B. Holm-Rasmussen, B. Alary, and H.-M. Lehtonen, "Late reverberation synthesis using filtered velvet noise," *Appl. Sci.*, vol. 7, no. 5, p. 483, 2017.

[5] G. D. Santo, K. Prawda, S. Schlecht, and V. Välimäki, "Differentiable feedback delay network for colorless reverberation," in *Proc. 26th Int. Conf. Digit. Audio Effects (DAFx23)*, Aalborg, Denmark, Sep. 2023, pp. 244–251.

[6] T. Jasa and N. Xiang, "Efficient estimation of decay parameters in acoustically coupled-spaces using slice sampling," *J. Acoust. Soc. Am.*, vol. 126, no. 3, pp. 1269–1279, 2009.

[7] G. Götz, R. F. Pérez, S. J. Schlecht *et al.*, "Neural network for multi-exponential sound energy decay analysis," *J. Acoust. Soc. Am.*, vol. 152, no. 2, pp. 942–953, Aug. 2022.

[8] J. M. Jot, "Proportional parametric equalizers application to digital reverberation and environmental audio processing," in *Proc. 139th Audio Eng. Soc. Conv.*, New York, USA, Oct. 2015.

[9] K. Prawda, S. J. Schlecht, and V. Välimäki, "Improved reverberation time control for feedback delay networks," in *Proc. Int. Conf. Digit. Audio Effects (DAFx)*, Birmingham, UK, Sep. 2019, pp. 247–254.

[10] K. Prawda, V. Välimäki, and S. Serafin, "Evaluation of accurate artificial reverberation algorithm," in *Proc. Sound Music Comput. Conf. (SMC)*, Torino, Italy, Jun. 2020, pp. 247–254.

[11] V. Välimäki, K. Prawda, and S. J. Schlecht, "Two-stage attenuation filter for artificial reverberation," *IEEE Signal Processing Letters*, vol. 31, pp. 391–395, 2024.

[12] V. Välimäki and K. Prawda, "Late-reverberation synthesis using interleaved velvet-noise sequences," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1149–1160, 2021.

[13] T. J. Terbock, F. Brinkmann, H. Gamper, N. Raghuvanshi, and S. Weinzierl, "Spatio-temporal windowing for encoding perceptually salient early reflections in parametric spatial audio rendering," *J. Audio Eng. Soc.*, vol. 71, no. 10, pp. 664–678, 2023.

[14] P. Sharma and C. Kehling, "How machines perceive rooms – regions of relevance in room impulse responses," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2025, pp. 1–5.

[15] P. Massé, A. Gallien, W. Kreuzer, and M. Noisternig, "Echo detection using the Herglotz wavefunction in spatial room impulse responses measured with spherical microphone arrays," submitted for publication (preprint available on ArXiv), 2026.

[16] 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, May 2021.

[17] B. Alary, A. Politis, S. Schlecht, and V. Välimäki, "Directional feedback delay network," *J. Audio Eng. Soc.*, vol. 67, no. 10, pp. 752–762, Oct. 2019.

[18] B. Alary and A. Politis, "Frequency-dependent directional feedback delay network," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2020, pp. 176–180.

[19] G. Götz, S. J. Schlecht, and V. Pulkki, "Common-slope modeling of late reverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 3945–3957, 2023.

[20] B. Alary, A. Politis, and S. Bilbao, "Designing directional reverberators for spatial sound reproduction," in *Proc. AES 5th Int. Conf. Audio Virtual Augment. Reality*, Redmond, WA, USA, Aug. 2024.

[21] S. Lee, H.-S. Choi, and K. Lee, "Differentiable artificial reverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 2541–2556, Jul. 2022.

[22] A. I. Mezza, R. Giampiccolo, E. D. Sena, and A. Bernardini, "Data-driven room acoustic modeling via differentiable feedback delay networks with learnable delay lines," *EURASIP J. Audio, Speech, Music Process.*, vol. 2024, no. 1, p. 51, Oct. 2024.

[23] O. Das, S. Schlecht, G. D. Santo, and Z. Cvetkovic, "Differentiable grouped feedback delay networks for learning direction and position-dependent late reverberation," *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Jan. 2026.

[24] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[25] J. Daniel and S. Moreau, "Further study of sound field coding with higher order ambisonics," in *Proc. 116th Audio Eng. Soc. Conv.*, Berlin, Germany, May 2004.

[26] F. Zotter, "Filterbank rule for spherical microphone array encoding and eqs for above aliasing, transducers, and binaural playback," in *Proc. 156th Audio Eng. Soc. Conv.*, Madrid, Spain, Jun. 2024.

[27] H. Morgenstern, B. Rafaely, and M. Noisternig, "Design framework for spherical microphone and loudspeaker arrays in a multiple-input multiple-output system," *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 2024–2038, Mar. 2017.

[28] J.-D. Polack, "Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics," *Appl. Acoust.*, vol. 38, no. 2–4, pp. 235–244, 1993.

[29] P. Götz, K. Kowalczyk, A. Silzle, and E. A. P. Habets, "Mixing time prediction using spherical microphone arrays," *J. Acoust. Soc. Am.*, vol. 137, no. 2, pp. 206–212, Feb. 2015.

[30] 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, p. 1033, Feb. 2020.

[31] T. Carpentier, M. Noisternig, and O. Warusfel, "Hybrid reverberation processor with perceptual control," in *Proc. 17th Int. Conf. Digit. Audio Effects (DAFx)*, Erlangen, Germany, Sep. 2014.

[32] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 2, pp. 702–714, 2007.

[33] B. Friedlander and B. Porat, "The modified Yule-Walker method of ARMA spectral estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-20, no. 2, pp. 158–173, 1984.

[34] H. Lee, "Multichannel 3d microphone arrays: A review," *J. Audio Eng. Soc.*, vol. 69, no. 1/2, pp. 5–26, 2021.

[35] S. J. Schlecht, J. Fagerström, and V. Välimäki, "Decorrelation in feedback delay networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. PP, no. 99, pp. 1–11, 2023.

[36] P. Heckbert, "Color image quantization for frame buffer display," *ACM Siggraph Computer Graphics*, vol. 16, no. 3, pp. 297–307, 1982.

[37] T. Carpentier and A. Einbond, "Spherical correlation as a similarity measure for 3-d radiation patterns of musical instruments," *Acta Acust.*, vol. 7, p. 40, 2023.

[38] O. Das and J. S. Abel, "Grouped feedback delay networks for modeling of coupled spaces," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 486–496, Jul. 2021.

[39] 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," *J. Audio Eng. Soc.*, vol. 68, no. 5, pp. 338–354, 2020.

A. APPENDIX: MIXING MATRIX

This appendix derives the mixing matrix $\mathbf{M}_b$ used to impose the target spatial covariance $\mathbf{C}_b$ on the FDN or DFDN outputs.

Problem Formulation. Let $\tilde{\mathbf{z}}^{(b)}$ be the whitened DFDN outputs for frequency band b. These signals are zero-mean, mutually uncorrelated, and have unit variance, so that their spatial covariance is the identity:

$$\mathrm{Cov}\left(\tilde{\mathbf{z}}^{(b)}\right) = \mathbb{E}\left[\tilde{\mathbf{z}}^{(b)}\left(\tilde{\mathbf{z}}^{(b)}\right)^T\right] = \mathbf{I}.$$

The objective is to find a matrix $\mathbf{M}_b$ such that the synthesized signal $\hat{\mathbf{z}}^{(b)} = \mathbf{M}_b\tilde{\mathbf{z}}^{(b)}$ strictly exhibits the measured target covariance $\mathbf{C}_b = \mathrm{Cov}(\mathbf{x}^{(b)})$.

Covariance of the Synthesized Signal. The covariance of $\hat{\mathbf{z}}^{(b)}$ is:

$$\mathrm{Cov}\left(\hat{\mathbf{z}}^{(b)}\right) = \mathbb{E}\left[\hat{\mathbf{z}}^{(b)}\left(\hat{\mathbf{z}}^{(b)}\right)^T\right] = \mathbb{E}\left[\left(\mathbf{M}_b\tilde{\mathbf{z}}^{(b)}\right)\left(\mathbf{M}_b\tilde{\mathbf{z}}^{(b)}\right)^T\right].$$

Using the matrix transposition property $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ and factoring out the deterministic matrix $\mathbf{M}_b$, we obtain:

$$\mathrm{Cov}\left(\hat{\mathbf{z}}^{(b)}\right) = \mathbf{M}_b\mathbb{E}\left[\tilde{\mathbf{z}}^{(b)}\left(\tilde{\mathbf{z}}^{(b)}\right)^T\right]\mathbf{M}_b^T = \mathbf{M}_b\mathbf{I}\mathbf{M}_b^T = \mathbf{M}_b\mathbf{M}_b^T.$$

To satisfy the target condition $\mathrm{Cov}(\hat{\mathbf{z}}^{(b)}) = \mathbf{C}_b$, we must solve:

$$\mathbf{M}_b\mathbf{M}_b^T = \mathbf{C}_b.$$

Factorization via Eigenvalue Decomposition. Since $\mathbf{C}_b$ is real, symmetric, and positive semi-definite, its eigenvalue decomposition is $\mathbf{C}_b = \mathbf{V}_b\mathbf{D}_b\mathbf{V}_b^T$, where $\mathbf{V}_b$ is orthogonal ($\mathbf{V}_b^T\mathbf{V}_b = \mathbf{I}$), and $\mathbf{D}_b = \mathrm{diag}(\lambda_i)$ is diagonal with non-negative eigenvalues. Substituting into the target equation:

$$\mathbf{M}_b\mathbf{M}_b^T = \mathbf{V}_b\mathbf{D}_b\mathbf{V}_b^T.$$

Because $\mathbf{D}_b$ is diagonal and non-negative, a valid factorization is obtained by taking its principal square root, defined element-wise as $\mathbf{D}_b^{1/2} = \mathrm{diag}(\sqrt{\lambda_i})$. This allows us to write $\mathbf{D}_b = \mathbf{D}_b^{1/2}\mathbf{D}_b^{1/2}$. Inserting the identity matrix $\mathbf{I}$ between these terms gives:

$$\mathbf{M}_b\mathbf{M}_b^T = \mathbf{V}_b(\mathbf{D}_b^{1/2}\mathbf{I}\mathbf{D}_b^{1/2})\mathbf{V}_b^T.$$

Substituting $\mathbf{I}$ with the orthogonal property $\mathbf{V}_b^T\mathbf{V}_b$:

$$\mathbf{M}_b\mathbf{M}_b^T = \mathbf{V}_b\mathbf{D}_b^{1/2}(\mathbf{V}_b^T\mathbf{V}_b)\mathbf{D}_b^{1/2}\mathbf{V}_b^T.$$

Regrouping the terms reveals two identical blocks:

$$\mathbf{M}_b\mathbf{M}_b^T = (\mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T)(\mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T).$$

Resulting Mixing Matrix. Let the candidate solution be $\mathbf{M}_b = \mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T$. We can easily verify that this matrix is symmetric ($\mathbf{M}_b = \mathbf{M}_b^T$):

$$\mathbf{M}_b^T = (\mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T)^T = (\mathbf{V}_b^T)^T(\mathbf{D}_b^{1/2})^T\mathbf{V}_b^T = \mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T = \mathbf{M}_b.$$

Therefore, the equation $(\mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T)(\mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T) = \mathbf{M}_b\mathbf{M}_b^T$ holds true. The matrix $\mathbf{M}_b = \mathbf{V}_b\mathbf{D}_b^{1/2}\mathbf{V}_b^T$ (Eq. 7) is the unique symmetric positive semi-definite square root of $\mathbf{C}_b$. Applying this matrix to the whitened FDN or DFDN outputs imposes the exact target spatial correlation without introducing directional bias.