A DENSELY SAMPLED DATASET OF ROOM IMPULSE RESPONSES FOR SOUND-FIELD RECONSTRUCTION

Mads Lang Matthesen[1,2], Rodrigo Ordoñez[1], Diego Caviedes Nozal[2], Henrik Gert Hassager[2]

[1] *Artificial Intelligence and Sound, Aalborg University, Denmark*
[2] *GN Advanced Science, Denmark*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Reconstructing sound-fields from a limited number of room impulse responses is a key challenge in spatial audio. Existing datasets are often spatially sparse or limited to a few rooms, hindering systematic benchmarking and training of sound-field reconstruction methods. This paper introduces Spatially Dense Room Impulse Responses (SpaDenRIR), a simulated dataset designed for developing and benchmarking sound-field reconstruction methods for early reflections. The dataset comprises 100 generated rooms, each containing a cubic microphone array of 22×22×22 microphones with 21.4 mm spacing, allowing spatial interpolation without aliasing up to 8 kHz. In every room, five sources are placed in distinct regions relative to the array. Room impulse responses are simulated using the image-source method and stored only for the first 100 ms, capturing the direct sound and early reflections. We describe the stochastic generation of room geometries, materials, source configurations, and array placement, and outline several application scenarios enabled by SpaDenRIR, including moving microphones and arbitrary array geometries. The dataset is intended as a shared training and evaluation resource for both model-based and machine-learning approaches to sound-field reconstruction.

The dataset with accompanying data handler and example code are publicly available [1].[1]

Keywords: *sound field estimation, dataset, geometrical acoustics, room impulse responses*

1. Introduction

Estimating a sound-field from a limited set of measurements is a central problem in acoustic signal processing, with applications spanning from immersive augmented reality and spatial audio reproduction to sound-field control and source separation. A sound-field can be represented by a set of room impulse responses (RIRs), which characterize the sound propagation in an enclosed space between a source and a receiver. The temporal structures of RIRs vary with minor changes in position. To characterize a sound-field accurately within a region of interest, densely spaced measurements are therefore necessary. Because detailed measurements are time-consuming, there has recently been a large focus on extrapolating and interpolating a sparse set of RIRs in an attempt to recreate a seemingly continuous sound-field that generalizes beyond the measured positions.

Several methods for estimating and reconstructing sound-fields have been proposed in the literature. Examples include exploiting spatial correlations to interpolate and extrapolate beyond a measured grid [2], system inversion approaches [3], or using moving microphones [4]. Recently, parts of the research field have moved towards deep learning based methods [5], which learn latent representations of the measured RIRs to generalize to unseen positions. Here, a variety of novel methods and architectures have been proposed, including multi-modal approaches [6], [7], convolutional neural networks [8], generative adversarial networks [9], [10], physics-informed machine learning [11], [12], [13], [14], [15], [16], transformer-based models [17] or wave based models [18].

The majority of these methods depend on datasets of sampled or simulated RIRs, either for training, for model updates, or as a ground truth for evaluating their performance. Several datasets of RIRs exist, including Treble10, AcousticRooms, and HiFi-HARP [7], [19], [20] presenting realistic simulated RIRs using a combination of wave-based and geometrical-acoustic simulations, where the latter includes 7th-order ambisonics. MeshRIR [21] introduced a recorded, spatially dense single-room dataset, and TrajectoRIR [22] a dataset of RIRs recorded with a moving microphone array.

Corresponding author: *mlmatthesen@gn.com.*

[1] `https://github.com/madslangmatthesen/SpaDenRIR`

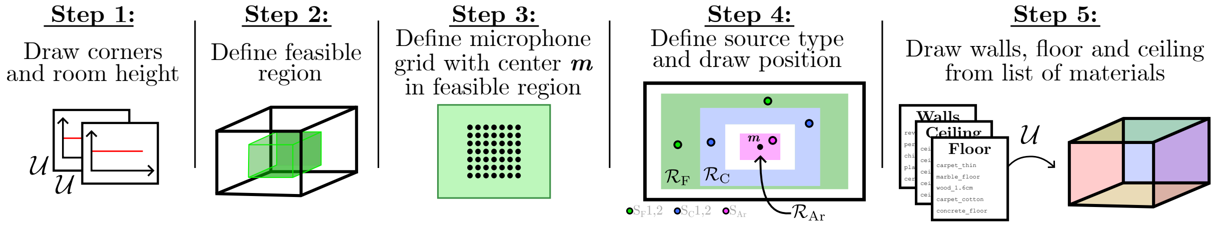

Figure 1: The steps of the room generation pipeline for the dataset.

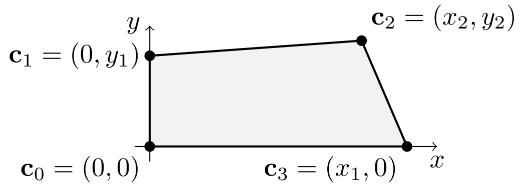

Figure 2: Illustration of a room base using four random variables.

Existing datasets are often tailored to specific scenarios and tasks and either lack spatial density or include a limited number of rooms. This motivates the addition of a more task-agnostic, spatially dense RIR dataset that can serve both as a high-resolution training resource and as a common benchmark for sound-field reconstruction methods, allowing for systematic investigation of small-scale positional dependencies accros many rooms.

We introduce a set of Spatially Dense RIRs (SpaDenRIR), a simulated multi-source data set of $22 \times 22 \times 22$ microphones placed 21.4 mm apart, to allow interpolation up to 8 kHz without spatial aliasing. The dataset comprises 100 rooms, each containing five sources. All RIRs are truncated to the first 100 ms, capturing the direct sound and early reflections.

This paper is structured as follows: Section 2 describes how the dataset was generated and concludes with an overview of the data set. Section 3 showcases select uses of the dataset. Lastly Section 4 concludes and outlines directions for future work building on the dataset.

2. Dataset Description

SpaDenRIR is a simulated dataset of 100 rooms, each populated with a spatially dense grid of RIRs generated using geometrical acoustics. The simulations were based on the image-source method implemented in Pyroomacoustics [23], [24]. All source and microphone positions, room geometries, and acoustical properties were based on stochastic sampling of random variables and redrawn for each room. Figure 1 shows individual steps of the data generation pipeline. The following section describes the step-wise generation of the dataset.

Step 1: Room Geometry

The floor plan of each room was defined by four corners:

$$\mathbf{c}_0 = (0,0), \quad \mathbf{c}_1 = (0,y_1),$$
$$\mathbf{c}_2 = (x_2,y_2), \quad \mathbf{c}_3 = (x_1,0), \tag{1}$$

in meters. Two scalar coordinates x_1 and y_1 were drawn independently from the same uniform distribution,

$$x_1, y_1 \overset{\text{i.i.d.}}{\sim} \mathcal{U}(3, 10). \tag{2}$$

The remaining scalars x_2 and y_2 were constrained to lie above the diagonal line spanned by $(x_1, 0)$ and $(0, y_1)$, so that the resulting polygon $\mathbf{c}_0, \ldots, \mathbf{c}_3$ forms a four-sided room base without self-intersections. Figure 2 shows an example of a room base constructed by the corners $\mathbf{c}_0, ..\mathbf{c}_3$. Lastly, the room height in meters was drawn from a separate uniform distribution:

$$H_{\text{room}} \sim \mathcal{U}(3, 6) \tag{3}$$

Step 2: Feasible Region

A feasible region for placing microphones and sources, $\mathcal{R}_{\text{feasible}}$, was defined as:

$$\begin{aligned}\mathcal{R}_{\text{feasible}} = &[x_{\min} + d_{margin},\ x_{\max} - d_{margin}] \\ &\times [y_{\min} + d_{margin},\ y_{\max} - d_{margin}] \\ &\times [z_{\min} + d_{margin},\ z_{\max} - d_{margin}],\end{aligned} \tag{4}$$

where $d_{\text{margin}} = 0.3\,\text{m}$ is a safety margin that ensures all microphones and sources are placed at least d_{margin} from the room boundaries.

Step 3: Cubic Microphone Array

Let $\mathcal{M} = \{\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_{n_{\text{mics}}}\}$ denote the cubic array of microphone position, with $\mathcal{M} \subset \mathcal{R}_{\text{feasible}}$.

A total of 10 648 microphones were arranged in a cubic grid of size $(0.45\,\text{m}, 0.45\,\text{m}, 0.45\,\text{m})$, resulting in 22 mics per dimension. This entails a distance between adjacent microphones of 21.4 mm. The spacing satisfies the spatial Nyquist criterion, so that the sound-field within the grid can be reconstructed without spatial aliasing up to 8 kHz at a speed of sound of $343\,\text{m}\,\text{s}^{-1}$.

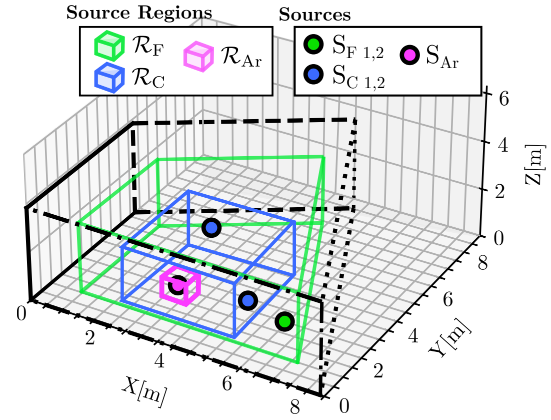

Figure 3: Illustration of a room with five sources and a microphone grid. $S_{\mathrm{F}1,2}$ denote the sources placed in the `'far'`-region, $S_{\mathrm{C}1,2}$ the sources in the `'close'`-region, and S_{Ar} the source placed within the array. The colored boxes show the corresponding regions, and the black box shows the room.

The center of the microphone array, $\mathbf{m}$, was drawn from another uniform distribution within the feasible region, $\mathcal{R}_{\mathrm{feasible}}$, found by:

$$\mathbf{m} \sim \mathcal{U}(\mathcal{R}_{\mathrm{feasible}}) \tag{5}$$

All microphones were modeled as independent, ideal omnidirectional point receivers.

Step 4: Sources

Three regions were defined with respect to the microphone array. Figure 1, Step 4 illustrates how these were defined. $\mathcal{R}_{\mathrm{Ar}}$ is the axis-aligned box with side lengths $\mathbf{L}_{\mathrm{array}}$ and center $\mathbf{m}$, corresponding to the microphone grid. The `'close'`-region $\mathcal{R}_{\mathrm{C}}$ is a rectangular shell around the array, ensuring sources placed in this area are between 0.1 m and 0.5 m away from the box bounding the array. Finally, the `'far'`-region, $\mathcal{R}_{\mathrm{F}}$ is the part of the feasible room $\mathcal{R}_{\mathrm{feasible}}$ lying outside $\mathcal{R}_{\mathrm{C}}$.

These regions were then used to place five sources in each room: Two in the `'far'`-region ($S_{\mathrm{F},1}, S_{\mathrm{F},2} \in \mathcal{R}_{\mathrm{F}}$), two in the `'close'`-region ($S_{\mathrm{C},1}, S_{\mathrm{C},2} \in \mathcal{R}_{\mathrm{C}}$), and one inside the array volume ($S_{Ar} \in \mathcal{R}_{\mathrm{Ar}}$). Figure 3 shows an illustration of a room with five sources and a microphone array.

Step 5: Materials and Absorption Coefficients

The materials used for the ceiling, floor, and each individual wall were sampled from predefined lists using uniform distributions. The lists contain twelve wall materials, ten floor materials, and five ceiling materials. All materials were modeled as frequency-dependent with absorption values specified in 7 octave bands with center frequencies 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz. In the Pyroomacoustics implementation of the image-source method, some materials lack a separate 8 kHz band; in these cases, the value in the 4 kHz band is reused. Figure 4 shows the per-band distributions of the absorption coefficients.

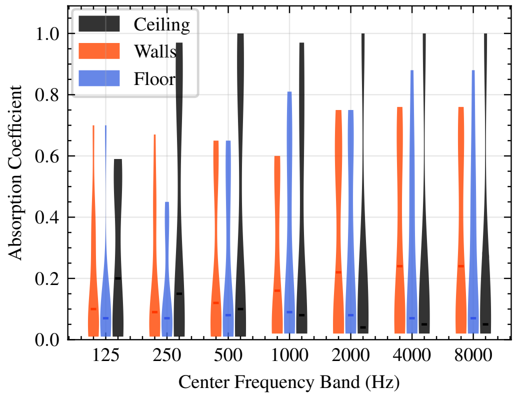

Figure 4: Distributions of absorption coefficients of the boundary materials.

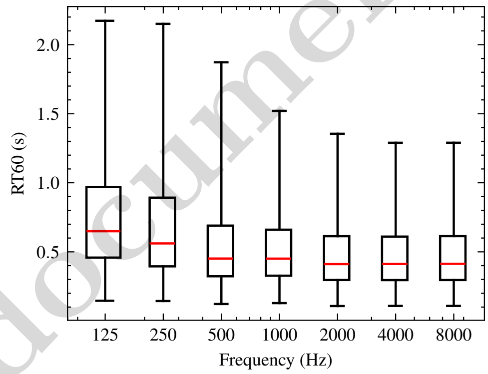

Figure 5: RT60 distributions per frequency band computed using weighted absorption coefficients. The red lines show the medians, and outliers (top 5 % reverberant bands) were not included for ease of readability.

In order to estimate the RT60 for each room using the Sabine equation, the total surface area S_{tot} and the volume, V_{room} of each room were found.

The RT60 for each frequency band per room was estimated using the absorption coefficients weighted by the area of the boundary, given by:

$$\bar{\alpha} = \frac{\sum_{j=1}^{J} A_j\, \alpha_j}{\sum_{j=1}^{J} A_j}, \tag{6}$$

where $\bar{\alpha}$ denotes the area-weighted mean absorption coefficient for each frequency band, and A_j denotes the area of each planar boundary j. This is used to estimate the RT60s for each room:

$$\mathrm{RT60} = 0.161 \cdot \frac{V_{\mathrm{room}}}{S_{\mathrm{tot}}\, \bar{\alpha}}. \tag{7}$$

The resulting estimated RT60s are similar to available statistics on natural reverberation [25], except that in our case the lowest frequency band is generally the most reverberant. This is likely a consequence of simulating fully empty rooms.

The maximum reflection order i_0 required to cover a decay over a time window T was approximated by

$$i_0 = \frac{cT}{L_{\text{room,min}}}, \tag{8}$$

where $L_{\text{room,min}}$ is the smallest room dimension and c the speed of sound. This is a simple path–length argument derived for shoebox rooms with smooth boundaries and should be regarded as an order-of-magnitude estimate.

The computational complexity of the image-source method grows roughly as

$$N(i_0) = \mathcal{O}\big((N-1)^{i_0}\big), \tag{9}$$

where i_0 is the maximum reflection order and N is the number of planar boundaries of the enclosure [26]. The minimum room dimension in the dataset is 3 m, and only the first 100 ms of each RIR are stored. Using Equation (8) with $T = 100\,\text{ms}$ and $L_{\text{room,min}} = 3\,\text{m}$ gives a required order of approximately $i_0 \approx 11$. Although this estimate does not account for the polygonal floor plans, it serves as a useful scale. We therefore set the maximum image order to $i_0 = 15$ as a compromise between computational feasibility and representative RIRs.

Lastly, the RIRs were written to an HDF5 file together with room descriptors following the formatting shown in Table 1. Both the full dataset and the single-room dataset files contain the `room_idx`. Most importantly, the RIRs are stored in `['room_idx']['rir_data']['rirs']` with shape $(M, S, L) = (10648, 5, 4800)$, where $M = \mathbf{n_{mics}}$ is the number of microphones, S the number of sources, and L the RIR length in samples. All RIRs were sampled at $f_s = 48\,\text{kHz}$ and truncated to the first 100 ms.

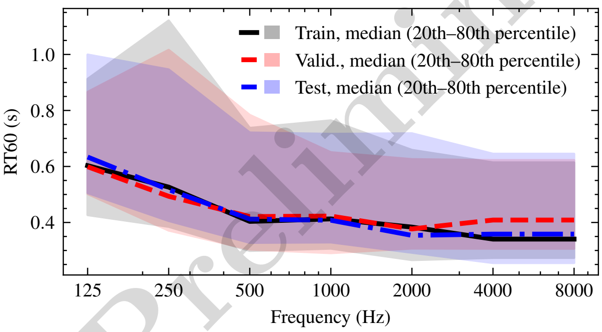

Figure 6: Median and 20-80 % percentiles for the data splits.

2.1. Train, Validation and Test Splits

As the dataset can be used as a benchmark for model-training and evaluation, a predefined train/validation/test-split has been made. However, due to the limited data size ($R = 15$ for test and validation splits, and $R = 70$ for the train split), conclusive statistical testing has not been made. Figure 6 shows the median and 20-80th percentiles of the RT60s for each split.

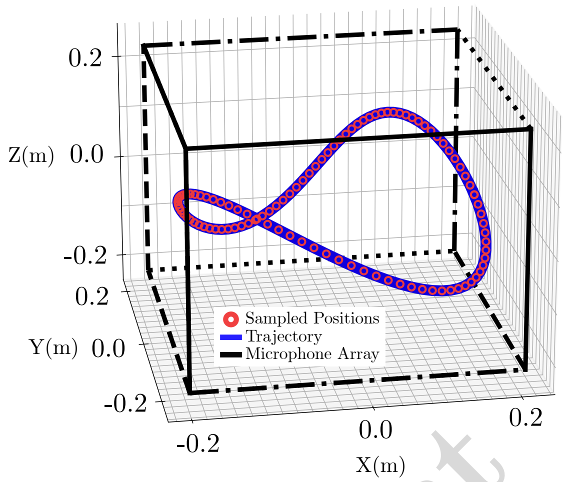

Figure 7: Illustration of a single microphone moving within a microphone array. The bounding box illustrates the array. The path shows the continuous path of the microphone, which is sampled with f_T Hz.

3. Application Examples

The dense sampling and multi-source structure of SpaDenRIR enables a range of controlled simulation scenarios. This section covers three examples.

3.1. Moving Microphone

As the sound-field ideally can be interpolated without spatial aliasing up to 8 kHz, it is possible to define arbitrary, continuous trajectories within the grid. As illustrated in Figure 7, a trajectory, $\mathbf{p}(t)$, is sampled at rate f_T, and the RIR at each sampled position is obtained by interpolating either from the surrounding local microphones, or from the complete set of measurements. A note on interpolation methods and quality can be found in the GitHub repository associated to the dataset.

Each interpolated RIR along $\mathbf{p}(t)$ can be convolved with the same anechoic signal, for example a speech recording. This yields a sequence of reverberant signals $\{y_k[n]\}$, where index k follows the trajectory. By applying overlapping Hann windows and cross-fading between $y_k[n]$ and $y_{k+1}[n]$, the sequence is combined into a single output that approximates a continuous recording from a microphone moving along $\mathbf{p}(t)$ in a static room. The procedure can be repeated for all sources in the room, and for arbitrary trajectories, which enables controlled simulations of moving receivers under multi-source conditions.

3.2. Arbitrary Arrays

The same interpolation framework allows arbitrary array geometries inside the grid. Instead of a single moving receiver, a set of virtual microphones with fixed relative positions is defined within the grid, and the RIR at each virtual element is interpolated from the dense simulations. This makes it possible to compare array layouts, or to generate training data for array-specific processing algorithms under identical room and source conditions, both in static-microphone

Table 1: Overview of stored room and RIR dataset information.

Entry	Description	Shape
Group: `room_information`		
`room_idx`	Integer index identifying the room instance.	( )
`fs`	Sampling frequency.	( )
`room_corners`	Room corner coordinates (x, y).	(4,2)
`room_height`	Room height.	( )
`volume`	Room volume.	( )
`wall_abs_coeffs`	Wall absorption coefficients per band.	(4,**B**)
`wall_materials`	Wall material names.	(4,)
`floor_material`	Floor material name.	( )
`floor_abs_coeffs`	Floor absorption coefficients per band.	(**B**)
`ceiling_material`	Ceiling material name.	( )
`ceiling_abs_coeffs`	Ceiling absorption coefficients per band.	(**B**)
`center_freqs`	Center frequencies of absorption bands.	(**B**)
`rt60_freq_dependent`	Frequency-dependent RT60 estimates.	(**B**)
`rt60`	Broadband RT60 estimate.	( )
Group: `rir_data`		
`ism_order`	ISM order used for simulation.	( )
`max_order_estimation`	Maximum feasible ISM order (Equation (8)).	( )
`mic_array_size`	Physical size of microphone array.	(3,)
`mic_coords`	Microphone coordinates.	(**M**,3)
`source_coords`	Source coordinates.	(**S**,3)
`rirs`	Room impulse responses.	(**M**,**S**,**L**)

or moving microphone configurations, and for both single-source or multiple source-setups.
As the dataset is based on idealized sources and receivers using the image-source method, effects from scattering, near-fields, or reactive surfaces are not included.

3.3. Sound-field Extrapolation and Interpolation

SpaDenRIR lends itself well to developing and testing sound-field interpolation and extrapolation methods. Different microphone subsets can be defined as training configurations, either using fixed selection patterns (e.g. specific array layouts) or stochastic sampling strategies. This allows construction of multiple, diverse training sets within a single room, or across many rooms, while reserving other microphones as evaluation points.
In this way, SpaDenRIR supports systematic studies of interpolation and extrapolation accuracy for the early reflections as a function of measurement density, array geometry, frequency range, and room configuration, and is particularly well suited as a common benchmark for machine-learning based sound-field reconstruction methods.

4. Further Work

The present release of SpaDenRIR focuses on early reflections simulated with geometrical acoustics in empty rooms. Further work includes modeling the late decay of the RIRs efficiently. A first direction is to augment the existing RIRs with late-reverberant tails modeled by parametric or statistical reverberation models fitted to the room parameters [27].
A second, related extension is to increase the maximum image-source order, and possibly the stored time window, for a subset of rooms. This would make it possible to quantify the accuracy of the hybrid early-plus-tail models and to study the transition region between deterministic reflections and diffuse decay as a function of room size and frequency. In parallel, more realistic room responses could be obtained by introducing scattering objects and more complex rooms and coupled rooms, increasing the complexity of the late reverberation.
Finally, the simulated data should be complemented with measurements. This could be realized either with a physically implemented dense array, or via moving-microphone measurements that emulate dense sampling, enabling direct comparison between simulated and measured early-reflection fields and late-reverberant tails.

5. Funding

The work presented in this project is a collaborative effort between Aalborg University and GN Advanced Science. The project was partially funded by Innovation Fund Denmark under grant 5304-00071B.

6. Acknowledgments

The authors gratefully acknowledge CLAAUDIA for providing access to high-performance computing resources, which enabled the large-scale data generation used in this study.

References

[1] M. L. Matthesen et al., *SpaDenRIR: Spatially dense simulated room impulse responses dataset*, 2026. DOI: 10.5281/zenodo.21618947.

[2] E. Fernandez-Grande et al., "Reconstruction of room impulse responses over extended domains for navigable sound field reproduction," in *Proc. Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 2021, pp. 1–8.

[3] A. Emthyas et al., "Spatial interpolation and extrapolation of binaural room impulse responses via system inversion," in *Proc. AES 5th Int. Conf. on Audio for Virtual and Augmented Reality*, 2024.

[4] T. Deppisch et al., "Spatial room impulse response estimation from a moving microphone array," in *Proc. 33rd Eur. Signal Process. Conf. (EUSIPCO)*, 2025, pp. 91–95.

[5] T. Van Waterschoot, "Deep, data-driven modeling of room acoustics," in *Proc. Forum Acusticum / EuroNoise*, European Acoustics Association, 2025, pp. 4065–4072.

[6] A. Ratnarajah et al., "AV-RIR: Audio-visual room impulse response estimation," *arXiv preprint arXiv:2312.00834*, 2024.

[7] X. Liu et al., "xRIR: Hearing anywhere in any environment," *arXiv preprint arXiv:2504.10746*, 2025.

[8] M. Pezzoli et al., "Deep prior approach for room impulse response reconstruction," *Sensors*, vol. 22, no. 7, 2022.

[9] X. Karakonstantis and E. Fernandez-Grande, "PWGAN – generative adversarial networks with physical sound field priors," *arXiv preprint arXiv:2308.00426*, 2023.

[10] E. Fernandez-Grande et al., "Generative models for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 153, no. 2, pp. 1179–1190, 2023.

[11] S. A. Verburg, E. Fernandez-Grande, and P. Gerstoft, "Differentiable physics for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 158, no. 5, pp. 4059–4069, 2025.

[12] X. Karakonstantis et al., "Room impulse response reconstruction with physics-informed deep learning," *J. Acoust. Soc. Am.*, vol. 155, no. 2, pp. 1048–1059, 2024.

[13] M. Olivieri et al., "Physics-informed neural network for volumetric sound field reconstruction of speech signals," *arXiv preprint arXiv:2403.09524*, 2024.

[14] S. Koyama et al., "Physics-informed machine learning for sound field estimation: Fundamentals, state of the art, and challenges," *IEEE Signal Process. Mag.*, vol. 41, no. 6, pp. 60–71, 2024.

[15] K. Kurata et al., "Spatial interpolation of room impulse responses based on deeper physics-informed neural networks with residual connections," *arXiv preprint arXiv:2512.22915*, 2025.

[16] K. Horikoshi et al., "Physics-informed deep prior for reconstruction of early-part room impulse responses," *jXiv*, 2026.

[17] S. Xu et al., "RIR-Former: Coordinate-guided transformer for continuous reconstruction of room impulse responses," *arXiv preprint arXiv:2602.01861*, 2026.

[18] E. Fernandez-Grande, A. Figueroa-Duran, and S. A. Verburg, "Spatially extended reconstruction of room impulse responses using a generalizable wave model," in *Proc. Forum Acusticum / EuroNoise*, 2025, pp. 4135–4138.

[19] S. S. Mullins et al., "Treble10: A high-quality dataset for far-field speech recognition, dereverberation, and enhancement," *arXiv preprint arXiv:2510.23141*, 2025.

[20] S. Saini and J. Peissig, "HiFi-HARP: A high-fidelity 7th-order ambisonic room impulse response dataset," *arXiv preprint arXiv:2510.21257*, 2025.

[21] S. Koyama et al., "MESHRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, 2021, pp. 1–5.

[22] S. Damiano, K. MacWilliam, V. Lorenzoni, T. Dietzen, and T. v. Waterschoot, "The trajectoRIR database: Room acoustic recordings along a trajectory of moving microphones," *J. Audio Speech Music Process.*, vol. 2026, no. 1, p. 19, Mar. 3, 2026, ISSN: 3091-4523.

[23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, 1979.

[24] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulations and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2018, pp. 351–355.

[25] J. Traer and J. H. McDermott, "Statistics of natural reverberation enable perceptual separation of sound and space," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 113, no. 48, E7856–E7865, 2016.

[26] H. Kuttruff, *Room Acoustics*. Taylor & Francis, 2006, ISBN: 0-203-18623-0.

[27] S. D. Ewert et al., "Computationally-efficient and perceptually-motivated rendering of diffuse reflections in room acoustics simulation," 2023.