DOI: 10.61782/fa.2023.0804

forum acusticum 2023

ROOM IMPULSE RESPONSE RECONSTUCTION USING PHYSICS-CONSTRAINED NEURAL NETWORKS

Xenofon Karakonstantis[1*] **Efren Fernandez-Grande**[1]

[1] Acoustic Technology
Deparment of Electrical and Photonics Engineering
Technical University of Denmark

ABSTRACT

In this paper, we present a method for reconstructing the sound field in a room using physics-informed neural networks. Our approach employs a limited set of room impulse responses as training data for the network, while also incorporating the fundamental physical principles of sound propagation in space through the use of the wave equation. This allows the network to not only learn the underlying physics of sound propagation but also utilize the nonlinear mapping capabilities of neural networks to adjust for any inhomogeneities in the room and measurement artifacts. Furthermore, the network can determine particle velocity and intensity through the use of autodifferentiation. The results indicate the effectiveness of the approach in terms of reconstruction accuracy and computational efficiency. This work presents a promising approach for sound field reconstruction and has potential for improving the representation of sound fields in various acoustic settings, including rooms and other complex environments, particularly for the synthesis of room impulse responses.

Keywords: *: physics-informed neural network - deep learning - sound field - room impulse responses*

**Corresponding author*: *xenoka@dtu.dk.*

1. INTRODUCTION

The ability to accurately reconstruct a sound field in a given environment is an important task with diverse applications. This includes spatial audio reproduction for virtual and augmented reality systems [1, 2], sound field control for reproducing coveted acoustic fields [3, 4] and sound field analysis for acoustic environment characterisation, design and optimization [5, 6]. These applications can further aid in the design and optimization of concert halls, stadiums, office spaces, and other public spaces where acoustics are a critical component of the listener's experience.

One approach commonly used for sound field reconstruction involves the use of basis function expansions, where the sound field is represented as a linear combination of predefined functions. These basis functions can be spherical harmonics, plane waves, point and multipole sources, wavelets, [7] or a combination thereof. [8] The choice of basis function depends on the particular application and the desired trade-offs between accuracy and computational complexity. Recent literature has seen the use of various techniques, including elementary wave superposition, [5,9] room mode decomposition, [10,11] spherical harmonics decomposition, [6] and more recently, deep learning models such as U-nets [12] or generative adversarial networks (GANs). [13, 14]

Accurately representing the sound field in locations where no measurements are available poses a significant challenge in sound field reconstruction. To address this, a variety of methods have been proposed, such as regularisation techniques that impose sparsity constraints on the reconstructed sound field, [5, 15] and machine learning approaches that leverage data-driven models to predict

the sound field in unobserved locations. [12, 13] Recently, a class of models that interpret the heuristic decomposition of room impulse responses (RIRs) [8, 16] has shown promising results in this task.

Deep learning techniques have recently gained significant attention in sound field reconstruction and show promise as a promising direction. Convolutional neural networks (CNNs) and generative adversarial networks (GANs) are two types of deep learning models that can learn complex relationships between input data (measurements or samples of the sound field) and the output (reconstructed sound field). These models are well-suited for sound field reconstruction tasks that have complex relationships between the measurements and the sound field, often requiring prior knowledge of the room properties.

Both wave expansions and deep learning approaches often rely on complex algorithms and can be computationally intensive, which limits their scalability and applicability to larger problems. A promising alternative is the use of physics-informed neural networks (PINNs), which are neural networks trained to solve partial differential equations (PDEs), such as the wave equation governing the propagation of sound waves in a given environment. By incorporating physical principles into the training process, PINNs can learn to accurately reconstruct sound fields while also capturing the underlying physics of the problem [17]. This makes them a promising candidate model for sound field reconstruction, as they offer improved accuracy and computational efficiency compared to traditional methods.

In this paper, we explore the use of PINNs for reconstructing sound fields in various enclosures, including acoustically challenging scenarios. We evaluate the performance of PINNs on four datasets, each representing a different type of enclosure and sound field configuration. The results demonstrate the effectiveness of PINNs in accurately reconstructing sound fields, with potential applications in a wide range of scenarios.

2. METHOD

2.1 Surrogate neural network for solving the wave equation

Surrogate neural networks have become increasingly popular in the field of partial differential equations (PDEs) due to their ability to handle high-dimensional input spaces and approximate complex functions effectively. These networks typically take in Cartesian coordinates $\mathbf{r} = (x, y, z)$ and a time instance t as input, and produce an output value representing the approximated eigenfunction $\Phi(\mathbf{r}, t)$ of the PDE at the corresponding coordinate within a predefined domain Ω_m. Their training involves fitting the forward problem to ensure that the network accurately predicts the function value at each input coordinate.

The neural networks fit the forward problem by satisfying a set of K constraints represented by $\mathcal{C}k$, which typically correspond to the PDEs being solved. This involves finding the value of $\Phi(\mathbf{r}, t)$ that satisfies the constraints for all coordinates $\mathbf{r}$ within each domain Ω_k and time instances $t \in \mathbb{R}^+$. In practice, these surrogate models are implemented as multi-layer perceptrons (MLPs) with non-linear activation functions and trained using a variant of stochastic gradient descent algorithm [18]. In this study, we employ sinusoidal activation functions for the neural network, as they are effective universal approximators when initialized properly. Sinusoidal activations are especially useful for modeling high frequency or periodic data, or data structures that require higher-order spatial derivatives, which other activation functions may not be able to handle. [19]

During training, these networks minimize the sum of constraints using the following loss function:

$$\mathcal{L} = \int_{\Omega} \sum_{k=1}^{K} \mathbb{1}_{\Omega_k}(\mathbf{r}, t) \left| \mathcal{C}_k(\Phi(\mathbf{r}, t), \nabla\Phi(\mathbf{r}, t), \ldots) \right| d\mathbf{r}dt. \tag{1}$$

Here, $\mathbb{1}_{\Omega_k}(\mathbf{r}, t)$ is a binary mask that specifies the constraints $\mathcal{C}_k(\cdot)$ that are applied to each point within the domain Ω. To sample points within the continuous space of Ω, we evaluate the loss function at various points during each iteration of the neural network training process.

Linear sound fields excited by a source can be described by the inhomogeneous wave equation as [20]

$$\nabla^2 p(t, \mathbf{r}) - \frac{1}{c^2} \frac{\partial^2 p(t, \mathbf{r})}{\partial t^2} = \delta(t, \mathbf{r}_s), \tag{2}$$

where $p(t, \mathbf{r})$ is the sound pressure as a function of time t and space $\mathbf{r}$ and a speed of sound c. The excitation term $\delta(t, \mathbf{r}_s)$ refers to the source at position $\mathbf{r}_s$. In a source-free region, we can ommit the right hand side of Eq. 2.

This paper aims to reconstruct and estimate the sound field $p(\mathbf{r}, t)$ of a room using a limited set of noisy measurements represented by $\tilde{\mathbf{p}} \in \mathbb{R}^{NM}$ of N time samples and M measurements, so that $\tilde{\mathbf{p}} = \mathbf{p} + \mathbf{n}$, where $\mathbf{n} \in \mathbb{R}^{NM}$ is additive measurement noise and $\mathbf{p}$ is the vector of the

unknown sound field $p(\mathbf{r}, t)$ evaluated at the measured locations. To achieve this, we propose a physics-informed neural network (PINN) approach that approximates the function of pressure over time and space. The PINN is trained to fit the measured pressure $\tilde{p}(\mathbf{r}_m, t_n)$ while fulfilling the wave equation.

The network is represented as $p_\bullet(\mathbf{r}, t) = \mathbf{W}_d \cdot \sin\left(\mathbf{W}_{d-1} \cdot \sin\left(\cdots \cdot \sin\left(\mathbf{W}_1 \cdot [\mathbf{r}, t]\right)\right)\right)$, where d is the number of hidden layers and $\mathbf{W}_1, \ldots, \mathbf{W}_d$ are the weight matrices. The pressure $p_\bullet(\mathbf{r}, t)$ refers to the predicted PINN pressure at position $\mathbf{r}$ and time t. The PINN does not assume explicit initial conditions for the pressure at time $t_0 = 0$ s, as the model is source agnostic. Furthermore, no boundary condition constraint is applied, to minimize the effect of strict assumptions made on the measurements. This leads to an objective function delineated by

$$\begin{aligned} &\arg\min_{p_\bullet} \sum_{m \in M} \sum_{n \in N} |(p_\bullet(\mathbf{r}_m, t_n) - \tilde{p}(\mathbf{r}_m, t_n))|^2 \\ &\text{s.t.} \\ &\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) p_\bullet(\mathbf{r}, t) = 0, \qquad \forall\, \mathbf{r} \in \Omega,\ t \in \mathbb{R}^+, \end{aligned} \tag{3}$$

where $\mathbf{r}_m$ and t_n refer to the positions and time instance of the measured pressure in a room. The Laplacian ∇^2 is obtained via autodifferentiation, a method used to efficiently calculate derivatives of functions, and is often associated with the underlying mechanism for backpropagation, the fundamental algorithm used in deep learning to train neural networks.

Once the network is trained, one can obtain the pressure field $p_\bullet(\mathbf{r}, t)\ \forall \mathbf{r} \in \Omega,\ t \in \mathbb{R}^+$, as well as the particle velocity with Euler's equation

$$\mathbf{u}_\bullet(\mathbf{r}, t) = -\frac{1}{\rho}\int_{t_0}^{t} \nabla p_\bullet(\mathbf{r}, t) \mathrm{d}t, \tag{4}$$

where the pressure gradient $\nabla p_\bullet(\mathbf{r}, t)$ is obtained with first order auto-differentation. Finally one can obtain the instantaneous intensity of the sound field in the domain as

$$\mathbf{I}_\bullet(\mathbf{r}, t) = \mathbf{u}_\bullet(\mathbf{r}, t) p_\bullet(\mathbf{r}, t), \tag{5}$$

allowing for the full characterisation of the measured sound field.

2.2 Simulated data set

For testing the proposed physics-informed neural network, we simulate a regular grid of room impulse responses with the method of image sources [21] by using the pyroomacoustics python package [22]. The simulated room is a 'shoebox' room with dimensions $6.2 \times 4.8 \times 3.8$ m^3, with the receiver positions situated in the middle of the room in a square grid of 1.2×1.2 m^2 parallel to the xy plane and situated at $z = 1.9$ m height, with a total of $36 \times 36 = 1296$ receiver positions. The source was placed 0.1 m from the wall parallel to the $y-$axis and at height $z_s = 1.9$ m (i.e. $[0, 0.1, 1.9]$ m). A uniform absorption was applied to the room boundaries in order to a acheive an average reverberation time of $T_{60} = 0.5$ s, using the simplification made in Sabine's formula ($T_{60} = 0.161\frac{V}{A}$ where V is the room volume and A the total absorption area) and the image source order was truncated to 13 to limit the computational time. Finally we add white noise scaled such that the the RIRs have a signal-to-noise ratio of 35 dB. The simulated configuration can be seen in Fig. 1. From the complete set of RIRs, we spatially subsample each dimension in the xy plane in a near-uniform manner, in order to retain $12 \times 12 = 144$ RIRs to be used for training the neural network, while the rest are used for validation. This way we emulate the challenges associated with obtaining a large number of real measurements, by employing a limited set of simulated RIRs. Consequently the spacing between each receiver is $d_{xy} = 0.102$ m for most reicever pairs and at most $d_{xy} = 0.137$ m, leading to a spatial Nyquist frequency of $f_{\text{Nyq}} = 1667$ Hz and $f_{\text{Nyq}} = 1250$ Hz respectively. This implies that perfect reconstruction can be obtained with a sinc function interpolation, up to the lower limit of the aforementioned frequencies, while the reconstructions are subject to spatial aliasing above this frequency. For the complete set of RIRs, the Nyquist frequency surmounts to $f_{\text{Nyq}} = 5000$ Hz. Although neural networks typically require data to be normalised within specific values, we found that the magnitude of the synthesised data was within a reasonable range, so this step was avoided.

2.3 Neural network architecture and training

For the PINN we use a multi-layer perceptron (MLP) network which receives at its input the collocation points $(\mathbf{r}, t) = (x, y, t)$, and is composed of 5 layers, each with 512 neurons and sinusoidal activation functions, apart from the final layer which is composed of a single neuron with a linear activation. The layers are initialised accord-

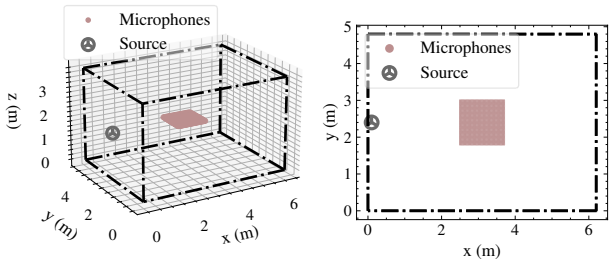

Figure 1: Image source method room and source-receiver configuration.

ingly for sinusoidal activation functions. [19] The network uses an adaptive loss function in order to "learn" how to balance the data and PDE terms in Eq. 3. [23] This loss function can be written as

$$L(\varepsilon;\theta;N) = \frac{1}{2\varepsilon_f^2} L_{PDE}\left(\theta;N_f\right) + \frac{1}{2\varepsilon_d^2} L_{data}\left(\theta;N_d\right) + \log \varepsilon_f \varepsilon_d, \tag{6}$$

with

$$L_{PDE}\left(\theta;N_f\right) = \frac{1}{|N_f|}\sum_{n_f=1}^{N_f} \left| f\left(\mathbf{r}_{n_f}, t_{n_f};\theta\right)\right|,$$

$$L_{\text{data}}\left(\theta;N_{\text{data}}\right) = \frac{1}{|N_d|}\sum_{n_d=1}^{N_d} \left|p_\bullet(\mathbf{r}_{n_d}, t_{n_d}) - \tilde{p}(\mathbf{r}_{n_d}, t_{n_d})\right|, \tag{7}$$

where θ are the MLP parameters, N_f, N_d are the number of collocation points where the PDE and data are evaluated respectively. The parameters ε_f and ε_d are adaptive weights which allow for the network to automatically assign the weights of individual loss terms by updating these parameters in each iteration based on maximum likelihood estimation.

We use the Adam optimiser for both the adaptive weights and the neural network parameters, with a learning rate of $\eta_\varepsilon = 2 \cdot 10^{-4}$ and $\eta_\theta = 2 \cdot 10^{-5}$ respectively. The parameters regarding the number of collocation points for each loss term are set to $N_f = N_d = 12000$, with points selected uniformly in both space and time for the PDE term, and the partial set of 144 RIRs at random spatio-temporal time instances for the data term. The training lasts for 10000 iterations on a single NVIDIA V100 GPU, which takes about 4 hours.

2.4 Evaluating the network performance

For evaluation of the reconstructed RIRs we use Pearson's correlation coefficient defined as

$$\rho(\mathbf{r},t) = \frac{\mathbb{E}[p(\mathbf{r},t)\hat{p}(\mathbf{r},t)] - \mathbb{E}[p(\mathbf{r},t)]\mathbb{E}[\hat{p}(\mathbf{r},t)]}{\sqrt{\mathbb{E}[p^2(\mathbf{r},t)]\mathbb{E}[p(\mathbf{r},t)]^2}\sqrt{\mathbb{E}[\hat{p}^2(\mathbf{r},t)] - \mathbb{E}[\hat{p}(\mathbf{r},t)]^2}}, \tag{8}$$

between any true (reference) $p(\mathbf{r},t)$ and reconstructed $\hat{p}(\mathbf{r},t)$ RIR on the reconstruction plane. This is a simple measure, showing how the reconstructed room reflections might covary with the experimental truth.

Furthermore, we find the root mean square error (RMSE) a good measure of discrepancy between the experimental truth and the reconstructed RIR. This can be defined by

$$\text{RMSE} = \sqrt{\mathbb{E}\left[\left(p(\mathbf{r},t) - \hat{p}(\mathbf{r},t)\right)^2\right]} \tag{9}$$

3. RESULTS

3.1 Sound field reconstruction

With the network trained, the reconstructed pressure is obtained as a continuous function represented by the PINN. Fig. 2 displays the measured pressure (training data), as well as the ground truth and the reconstructed pressure in snapshots for $t = 14$, 20, 24 and 33 ms. Since the simulations assume an ideal low pass filter (ie. sinc function) as a point source, ripples in the form of (phantom) wave fronts can be observed in the domain when the delays do not align perfectly with the discrete aperture. For the most part the network ignores these equating them with the noise floor, however it reconstructs the direct part of the responses ($t = 14$ ms) along with an aliased wave superimposed on the actual wave front. This is also evident in the true sound field.

3.1.1 Sound field characterisation

The reconstructed particle velocity vector is shown in Fig. 3 for the same snapshots as the aforementioned pressure. Since the velocity is proportional to the gradient of the sound pressure, the high values (longer arrow heads) indicate rapid spatial changes. However its estimation is also quite prone to measurement noise and the epistemic uncertainty of the strict domain bounds where the PINN operates. The PINN seems to faithfully replicate these abrupt changes in the medium, however, time $t = 42$ ms

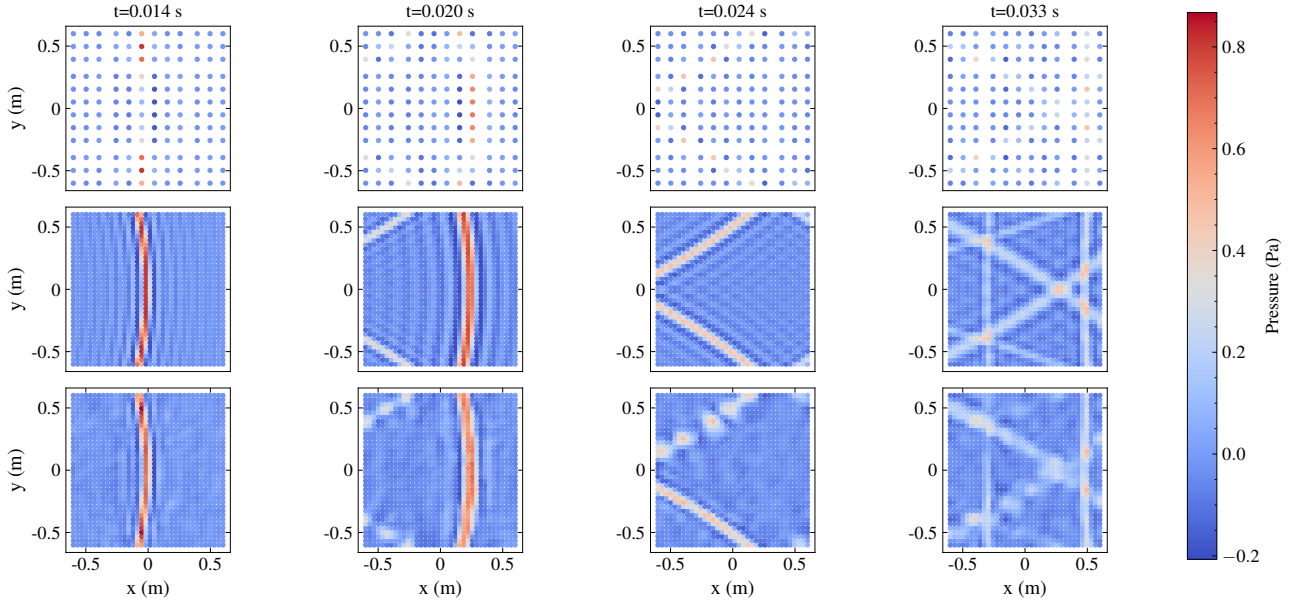

Figure 2: Pressure snapshots. From top to bottom: measured pressure, true pressure, reconstructed pressure (normalised to unit magnitude)

shows the discontinuities in the wavefront affect the particle velocity estimation also, with the wave front arriving from the top of the domain displaying erratic changes.

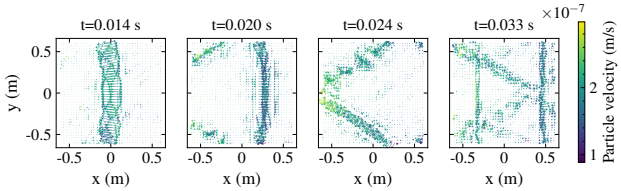

Figure 3: Reconstructed particle velocity snapshots.

Similarly the intensity in Fig. 4 indicates the direction of propagation of wavefronts. At time $t = 14$ ms, the snapshot displays a spherical spread and in snapshots $t = 20$, 24 and 30 ms we see more planar waves. Some of the wave fronts are not perfectly reconstructed, as becomes evident in $t = 24$ ms, since the aperture was placed symmetrically in the room with respect to the source, so both wavefronts should be almost identical. It seems that the PINN might have a slight bias towards the bottom half of the aperture, as the reconstructed fields are represented more accurately there.

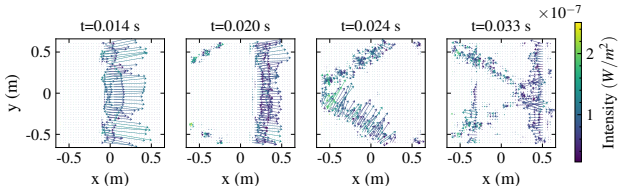

Figure 4: Reconstructed instantaneous intensity snapshots.

3.1.2 Evaluation

In order to carry out a comparison, we interpolate the sound field with a simple sinc interpolation. At a intermicrophone spacing between $d_{Nyq} = 0.102$ m and $d_{Nyq} = 0.137$ m, perfect reconstruction should be possible up to a frequency of $f_{\text{Nyq}} = 1250$ Hz, after which spatial aliasing occurs. The responses are low-pass filtered with a cut-off frequency of 4 kHz, well-beyond the Nyquist rate. Figure 5 displays the pressure root mean square error (RMSE) and correlation coefficient for the interpolated pressure. Based on the evaluation, the error is situated around -25 dB, while the correlation is below 0.3 for most of the interpolated sound field. In contrast, the PINN RMSE and correlation displayed in Fig. 6 shows

an RMSE of about -35 dB with most of the sound field showing a degree of correlation of around 0.7.

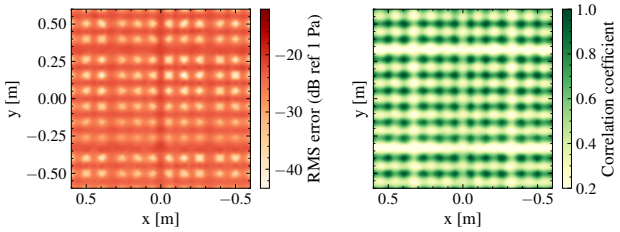

Figure 5: Sinc interpolated pressure RMSE and correlation coefficent over aperture.

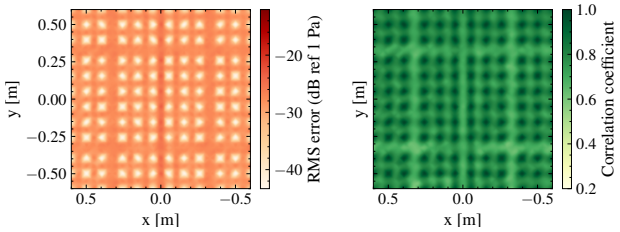

Figure 6: PINN pressure RMSE and correlation coefficent over reconstructed aperture.

4. DISCUSSION

Given that spatial aliasing can hinder the capabilities of sound field control applications and sound field aurlisation algorithms among other funcionalities of microphone array processing, a PINN seems to show substantial capabilities at "super-resolving" sound fields, while being interpretable as a model to a certain degree. When compared to a simple sinc interpolation, it is clearly noteworthy. It is able to reconstruct simulated sound fields which emulate real-world conditions with limited resources and subsequently, experimental data is scarce. Noteably, it is able to fully characterise a sound field by reconstructing pressure, particle velocity and intensity. Given that autodifferentation is a fairly cheap operation, this becomes attractive feature for most near-field acoustic holography (NAH) also.

Furthermore, given that the training time is not extensive, it can easily be compared to other methods for sound field reconstruction, with the benefit of representing a continuous pressure field rather than its discretised form, allowing for a gridless reconstruction. This is quite useful for real-time sound field aurilisation applications, since inference time is negligible.

5. CONCLUSION

In this paper, we presented a novel approach for reconstructing sound fields in a room using physics-informed neural networks. The method shows promising results in terms of reconstruction accuracy and computational efficiency, making it a valuable alternative to traditional methods. By incorporating the fundamental physical principles of sound propagation, the approach provides a good understanding of deep learning models for physical processes, which are often associated with black-boxes. Furthermore, their is clear outlook on using a PINN with experimental data, for a broad range of applications where the acoustic holography principle is used. Our work contributes to the ongoing research on sound field reconstruction and has potential for improving the representation of sound fields in various applications.

6. ACKNOWLEDGMENTS

This work was supported by the VILLUM Foundation, under grant number 19179, "Large-scale acoustic holography." and the Danish Sound Cluster.

7. REFERENCES

[1] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1584–1593, 1996.

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

[3] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 81–91, 2015.

[4] M. B. Møller and M. Olsen, "Sound zones: On performance prediction of contrast control methods," in

Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control, Audio Engineering Society, 2016.

[5] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3770–3779, 2018.

[6] M. Nolan, S. A. Verburg, J. Brunskog, and E. Fernandez-Grande, "Experimental characterization of the sound field in a reverberation room," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2237–2246, 2019.

[7] E. Zea, "Compressed sensing of impulse responses in rooms of unknown properties and contents," *Journal of Sound and Vibration*, vol. 459, p. 114871, 2019.

[8] E. Fernandez-Grande, D. Caviedes-Nozal, M. Hahmann, X. Karakonstantis, and S. A. Verburg, "Reconstruction of room impulse responses over extended domains for navigable sound field reproduction," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–8, IEEE, 2021.

[9] D. Caviedes-Nozal, N. A. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," *The Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 1107–1119, 2021.

[10] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205–216, 2013.

[11] O. Das, P. Calamia, and S. V. A. Gari, "Room impulse response interpolation from a sparse set of measurements using a modal architecture," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 960–964, IEEE, 2021.

[12] F. Lluís, P. Martínez-Nuevo, M. Bo Møller, and S. Ewan Shepstone, "Sound field reconstruction in rooms: Inpainting meets super-resolution," *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. 649–659, 2020.

[13] X. Karakonstantis and E. Fernandez Grande, "Sound field reconstruction in rooms with deep generative models," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 263, pp. 1527–1538, Institute of Noise Control Engineering, 2021.

[14] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, "Generative models for sound field reconstruction," *The Journal of the Acoustical Society of America*, vol. 153, no. 2, pp. 1179–1190, 2023.

[15] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 10, pp. 1929–1941, 2017.

[16] A. A. F. Durán and E. F. Grande, "Reconstruction of room impulse responses over an extended spatial domain using block-sparse and kernel regression methods," in *24th International Congress on Acoustics*, 2022.

[17] N. Borrel-Jensen, A. P. Engsig-Karup, and C.-H. Jeong, "Physics-informed neural networks for one-dimensional sound field predictions with parameterized sources and impedance boundaries," *JASA Express Letters*, vol. 1, no. 12, p. 122402, 2021.

[18] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol. 378, pp. 686–707, 2019.

[19] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7462–7473, 2020.

[20] F. Jacobsen and P. M. Juhl, *Fundamentals of general linear acoustics*. John Wiley & Sons, 2013.

[21] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[22] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation

and array processing algorithms," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 351–355, IEEE, 2018.

[23] Z. Xiang, W. Peng, X. Liu, and W. Yao, "Self-adaptive loss balanced physics-informed neural networks," *Neurocomputing*, vol. 496, pp. 11–34, 2022.