Audio Engineering Society

Convention Express Paper 327

Presented at the 158th Convention

2025 May 22-24, Warsaw, Poland

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Reconstructing Sound Fields with Physics-Informed Neural Networks: Applications in Real-World Acoustic Environments

Rigas Kotsakis[2], Sotiris Lois[1], Iordanis Thoidis[1], Nikolaos Vryzas[1], Lazaros Vrysis[1], and George Kalliris[1]

[1] *Aristotle University of Thessaloniki, Greece*

[2] *International Hellenic University, Greece*

Correspondence should be addressed to Nikolaos Vryzas: nvryzas@auth.gr

ABSTRACT

Reconstructing sound fields is an essential component in applications that aim to simulate sound in physical spaces to deliver immersive audio experiences, such as augmented, virtual, and mixed reality systems. Traditionally, sound field reconstruction is implemented via interpolation and probabilistic methods, which estimate the sound pressure field based on a limited number of acoustic measurements, combined with known room geometries. Recent advances have highlighted the potential of Physics-Informed Neural Networks (PINNs), which incorporate the governing physical laws of the problem, such as the acoustic wave equation, into the model training process. In this study, we investigate the application of PINNs for sound field reconstruction from real-world data, focusing on the estimation of room impulse responses at unobserved spatial locations. We train and evaluate the proposed PINN on a publicly available dataset of measured impulse responses recorded in the Great Hall of Queen Mary University of London. The model demonstrates accurate estimation of the impulse response envelope in unseen regions, particularly in low-frequency components of the sound field. However, challenges remain in capturing the detailed waveform structure in the mid- and high-frequency range. This work is conducted within the context of the SCENE research project.

1 Introduction

Sound field and impulse response reconstruction strategies are considered useful for acoustic modelling, while finding applicability in several aspects like immersive audio, virtual environments (VR/AR), room acoustics and architectural design. Traditionally, these reconstructions have usually relied on models exploiting techniques and properties deriving from wave propagation theory. One prominent example is the Compressed Singular Value Decomposition-based Equivalent Source Method, implicating sparsity in the sound field and applying matrix regularization [1]. When boundary conditions must be taken into account (especially for interior spaces and rooms), more complex mathematical tools have to be utilized, such as Green's functions and Kirsch-Kress implementations, leading into more complex acoustic settings [2].

In contrast to purely numerical methods, some heuristic techniques have been proposed based on spatio-temporal filtering of audio waves, without attempting to solve reverse engineering mathematical problems [3]. Analytical methods have also been extended to various geometries (i.e., orthogonal, cylindrical), in order to offer adaptability to acoustics perception in automotive interiors or non-symmetrical rooms via the helical wave spectrum framework [4]. Moreover, when noisy fragments appear, the efficiency of sound field

reconstruction is further hardened leading into supplementary mathematical modelling through generalized Fourier representations in order to diminish noisy data with least-square error strategies [5].

More recently, sound field reconstruction moves towards to time-domain aspects with the utilization of spherical microphone arrays and harmonic decomposition, allowing researchers to track the temporal evolution of sound fields with improved phase and amplitude accuracies based on precise impulse response estimation [6]. Furthermore, in more reverberant spaces, where spectral filtering acts more insufficiently, the time-domain approaches use regularization techniques to ensure both stability and temporal accuracy via spatial audio rendering [7].

Bayesian methods were also introduced because of their ability to incorporate dynamically prior knowledge and manage uncertainty. For instance, probabilistic frameworks have been developed to estimate room impulse responses from sparse microphone data. These methods offer better generalization and have been proven to be essential for applications like virtual concert hall simulations or studio design [8].
Another approach refers to blind estimation of room acoustic parameters, emphasizing on the effects of geometry information [9]. Machine and deep learning techniques have been significantly elaborated, thus, apart from typical sound classification applications to more sophisticated spatial audio design and generation [10]-[14]
Another common scenario when attempting sound field reconstruction refers to the existence of reduced measurements/ audio data. A treatment to this problem has been proposed with the Iterative Group Complex Orthogonal Matching Pursuit, which attempts to decomposes the acoustic environment into active and inactive audio space [15]. In addition, when combining Prolate Spheroidal Wave Functions with Orthogonal Matching Pursuit techniques, the sound field reconstruction goals can be effectively achieved even with limited audio data [16]. On the other hand, statistically oriented methods are also present, that attempt to reconstruct sound field based on the calculation or error functions, towards the detection and retainment of time domain amplitude peaks and also smooth potential sound artifacts especially in low frequencies [17].

It has to be highlighted that besides the aforementioned and more strict processing implementations, machine learning techniques unavoidably entered the scene of sound field reconstruction. Deep Neural Network topologies along with the Gaussian probabilistic modeling lead into infrastructures that are able to learn spatially adaptive kernels on-the-fly, allowing for better generalization and reduced overfitting in complex acoustic scenes [18]. Complementary parametric models in the spherical harmonic domain are also being used to segment sound fields, supporting tasks like source separation and spatial filtering when using compact microphone arrays [19].

2 Physics-Informed Neural Networks

Although neural networks are widely adopted across scientific and engineering domains due to their flexibility and expressiveness, encoding the physical laws that govern acoustic wave propagation into such models remains a significant challenge. This difficulty arises from the complexity and high dimensionality of the underlying physical principles involved in sound propagation. Instead of solely relying on training data like traditional machine learning strategies, PINNs take into consideration the actual physical problem and respective governing physical laws [23].

That is, PINNs have introduced a promising framework by embedding the wave equation directly into the learning objective. By doing this, they can estimate measurements that are not only data-driven but also aligned with what is known to be true from physics. This makes them useful when there is not enough data to work with, or when the available data is noisy or incomplete, as they can still give accurate predictions. PINNs have been proven to be valuable in areas where traditional methods might not be able to handle the complexity of the problems, or even when trying to solve the corresponding physical equations is especially tough [24].

By incorporating the laws of physics directly into their training process, PINNs do not just learn the underlying patterns but face the rules that govern the physical world. This means they can handle tricky situations like solving problems with missing information or estimating unknown parameters in a model. Essentially, PINNs can offer better results with less data, which is something that sets them apart from purely data-driven methods or traditional machine learning approaches [25]. This approach enables networks to preserve physical consistency

while remaining effective even with sparse or limited training data.

Previous studies have demonstrated that PINNs can accurately reconstruct acoustic sound fields, including speech, using only a small number of microphone measurements [20]. When extended with more advanced, physics-guided architectures, these models have shown robustness to noise and have enabled field reconstruction even in acoustically complex environments [21], [22]. A recent study by Karakonstantis et al. [29] further demonstrated the effectiveness of PINNs for reconstructing room impulse responses from simulated and measured data, by incorporating both the physical constraints of the wave equation and real-world recordings collected in a reverberant room.

The current research explores the applicability of Physically Informed Neural Networks for impulse response estimation utilizing a combination of experimental measurements and simulated data for sound field reconstruction [26].

3 Methodology

The first step towards audio field reconstruction experiments was to formulate an initial audio data collection from impulse responses. We utilized the publicly available database of impulse responses measured at the Great Hall of Queen Mary University of London's Mile End campus [27]. The recordings were captured using the sine sweep method [28], positioning a loudspeaker source while microphones were placed in the seating area on the floor. As Figure 1 depicts there is approximately a 23m x 16m area, which was firstly cleared of chairs. The microphone positions were scattered over a 12 m x 12m area, creating a x-y grid starting from 0-0 position to 12-12, therefore 169 impulse response measurements [29].

All impulse responses were resampled to 8 kHz to reduce computational cost during training and inference, while preserving the temporal structure relevant for perceptual and acoustic evaluation. From the total of 169 recordings, 100 measurements were selected and used for training and validation, while the remaining 69 recordings were held out for testing and evaluation.

We follow a physics-informed learning approach to train a neural network for acoustic field reconstruction. The network receives as input the spatio-temporal coordinates (x, y, t), and is trained to output as $\hat{p}$, the corresponding sound pressure value $p(x, y, t)$. Supervision is performed using a combination of a data-driven loss and a physics-based loss. Specifically, the signal-to-distortion ratio loss (L_{SDR}) is applied on the measured impulse response points to directly supervise the model against the ground truth recordings.

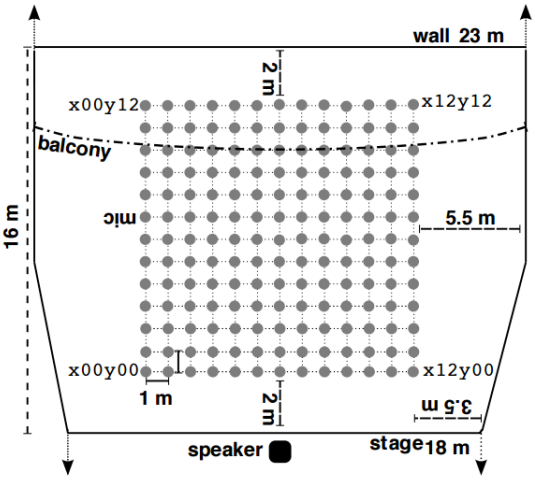


Figure 1. Great Hall of Queen Mary University

An additional loss term is used based on the wave equation, to incorporate physical constraints and constrain the network to produce outputs that are consistent with the underlying physical dynamics of acoustic wave propagation, following the approach of Karakonstantis *et al.* [23]. Spatio-temporal points are densely resampled within the measurement domain, with a resolution higher than that of the original microphone grid. These interpolated points are used solely for enforcing the partial differential equation loss term (L_{PDE}), without requiring corresponding ground truth pressure values. Finally, the total loss function can be expressed as:

$$L_{total} = L_{SDR} + L_{PDE} \quad (1)$$

where:

$$L_{SDR} = -10 \cdot \left(\frac{\|p\|^2}{\|p - \hat{p}\|^2} \right) \quad (2)$$

$$L_{PDE} = \frac{1}{N} \sum \left\| \frac{\partial^2 \hat{p}}{\partial t^2} - c^2 \cdot \left(\frac{\partial^2 \hat{p}}{\partial x^2} + \frac{\partial^2 \hat{p}}{\partial y^2} \right) \right\| \quad (3)$$

The PINN consisted of a deep neural network with a total of 460k trainable parameters. It comprised eight hidden layers, each containing 256 units and the sine activation function. The model was trained for 200 epochs using the Adam optimizer, with each epoch involving random sampling of time windows across the training dataset.

4 Results

The performance of the model was evaluated on the Great Hall test set, consisting of spatial positions that are uniformly distributed across the measurement grid. Visual and quantitative evaluation shows that the model successfully captures the low-frequency structure of the impulse response, particularly around the onset and early reflections, where the signal energy is concentrated. This is reflected in the low mean squared error (MSE = 0.0007) and mean absolute error (MAE = 0.005) between the estimated and ground truth waveform responses. Figure 2 illustrates this effect, comparing the normalized energy envelope of the estimated and target impulse responses at selected test locations.

Additionally, the model produces accurate estimates of reverberation time, with RT20 and RT30 absolute errors averaging 0.04 s and 0.06 s, respectively. However, the model struggles to fully reconstruct the diffuse tail of the impulse response, especially in the mid and high-frequency regions, up to 4 kHz. This limitation is evident in the negative SI-SDR score (-9.7 dB) and relatively low normalized cross-correlation (NCC = 0.2), indicating a mismatch in fine temporal structure beyond the early reflections.

5 Conclusions and Future Work

The current work examined the application of PINNs towards impulse response estimation for the sound field reconstruction problem. For this reason, an open data collection of audio impulse responses was utilized, deriving from the Great Hall of Queen Mary University of London and quite promising results emerged from the training process. Subsequent experiments are yet to be conducted for the acoustic behaviour of other rooms with different geometries and aim of scope, in order to evaluate and generalize the preliminary results of current research. The main scope of SCENE project is to create a dynamic and constantly evolving audio dataset that would facilitate efficient transferability of room acoustics based on a restricted number of measurements of the corresponding physical space.

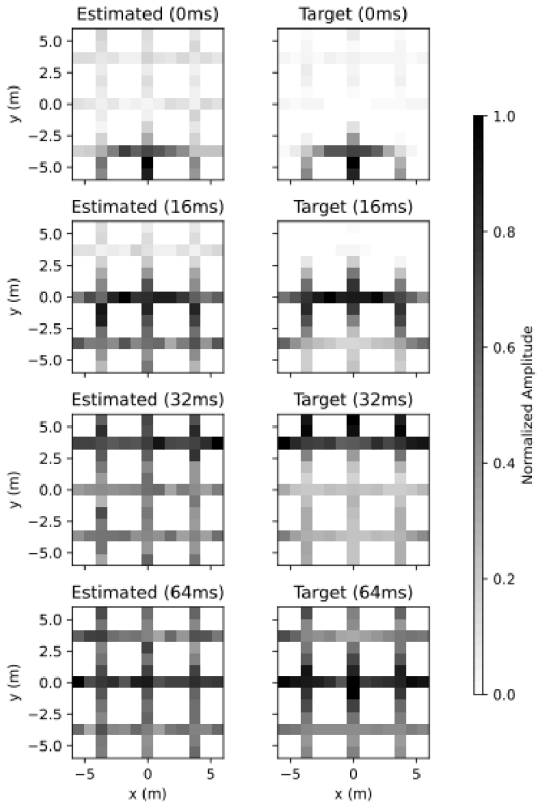


Figure 2. Estimated and ground truth normalized envelope levels of the room impulse response at selected time points. Grey regions correspond to test grid positions used for evaluation, while training positions are shown in white.

6 Acknowledgement

This research is part of the SCENE project and has received funding from the European Union's Horizon research and innovation programme under grant agreement No 101095303.

References

[1] L. Jiang, Z. Huang, Y. Xi and J. Liu, "Sound Field Reconstruction of Plate Using Compressed Singular Value Decomposition Equivalent Source Method Combined with Generalized Inverse of Matrix," in Proceedings of 2024 OES China Ocean

Acoustics (COA) Conference, 29-31 May, 2024, Harbin, China.

[2] R. Potthast, F.M. Fazi, P.A. Nelson and J. Seo, "Two Sound Field Reconstruction Techniques Based on Integral Equations," in Proceedings of Hands-Free Speech Communication and Microphone Arrays Conference, 06-08 May, 2008, Trento, Italy.

[3] S. Koyama, K. Furuya, Y. Hiwasaki and Y. Haneda, "Analytical Approach to Wave Field Reconstruction Filtering in Spatio-Temporal Frequency Domain," IEEE Transactions on Audio, Speech, and Language Processing, Vol 21, no 4, 2013.

[4] S. Koyama, K. Furuya, Y. Hiwasaki and Y. Suzuki, "Sound Field Reproduction Using Multiple Linear Arrays Based on Wave Field Reconstruction Filtering in Helical Wave Spectrum Domain," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 26-31 May, 2013, Vancouver, Canada.

[5] Fazi, Filippo M.; Nelson, Philip A., "The Ill-Conditioning Problem in Sound Field Reconstruction," in Proceedings of 123rd AES Convention, 5-8 October, New York, USA.

[6] P. Jiang, Z. Chu and P. Li, "Time-domain Sound Field Reconstruction Based on Spherical Acoustic Holography," in Proceedings of 5th International Conference on Information Communication and Signal Processing (ICICSP), 26-28 November, 2022, Shenzhen, China .

[7] M. Kolundzija, C. Faller and M. Vetterli, "Designing Practical Filters for Sound Field Reconstruction," in Proceedings of 127th AES Convention, Berkeley, USA, 2009.

[8] A. Figueroa-Duran, X. Karakonstantis and E. Fernandez-Grande, "Bayesian Framework for Room Impulse Response Reconstruction Using Explicit Frequency Regularisation," in Proceedings of 156th AES Convention, 15-17 June, 2024, Madrid, Spain.

[9] N. Vryzas, L. Vrysis, M.E. Stamatiadou, R. Kotsakis and C. Dimoulas, "The effect of geometry information in blind estimation of room acoustic parameters" In *Audio Engineering Society Convention 154*. Audio Engineering Society, 2023.

[10] R. Kotsakis, M. Matsiola, G. Kalliris and C. Dimoulas, "Investigation of spoken-language detection and classification in broadcasted audio content", *Information*, *11*(4), 211, 2020.

[11] N. Vryzas, L. Vrysis, R. Kotsakis and C. Dimoulas, "A web crowdsourcing framework for transfer learning and personalized speech emotion recognition" *Machine Learning with Applications*, *6*, 100132, 2021.

[12] M. Heydari, M. Souden, B. Conejo, B. and J. Atkins, "ImmerseDiffusion: A Generative Spatial Audio Latent Diffusion Model", In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE, 2025.

[13] N. Vryzas, L. Vrysis, R. Kotsakis and C. Dimoulas, "Speech emotion recognition adapted to multimodal semantic repositories", In *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)* (pp. 31-35). IEEE, 2018.

[14] C.A. Dimoulas and G.M. Kalliris, "Investigation of wavelet approaches for joint temporal, spectral and cepstral features in audio semantics" In *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.

[15] S. Xu, J. A. Zhang, T. Abhayapala, A. Bastine and P. Samarasinghe, "Iterative and Complex Orthogonal Matching Pursuit for Broadband Sparse Sound Field Reconstruction," in Proceedings of 18th International Workshop on Acoustic Signal Enhancement (IWAENC), 09-12 September, 2024, Aalborg, Denmark.

[16] X. Zhang, J. Lou, S. Zhu, J. Lu and R. Li, "Sound Field Reconstruction Using Prolate Spheroidal Wave Functions and Sparse Regularization," Sensors, Vol 23, no 19,

2023.

[17] Linsen Huang and Rui Zeng, “Sound Field Reconstruction Using Error Function Method,” IEEE Sensors Journal, Vol 24, no 24, 2024.

[18] Z. Liang, W. Zhang and T.D. Abhayapala, “Sound Field Reconstruction Using Neural Processes with Dynamic Kernels,” Journal of Audio Speech and Music Processing, Vol 12, 2024.

[19] A. Politis, J. Vilkamo and V. Pulkki, “Sector-Based Parametric Sound Field Reproduction in the Spherical Harmonic Domain,” IEEE Journal of Selected Topics in Signal Processing, Vol 9, no 5, 2015.

[20] M. Olivieri, X. Karakonstantis, M. Pezzoli, F. Antonacci, A. Sarti and E. Fernandez-Grande, “Physics-Informed Neural Network for Volumetric Sound Field Reconstruction of Speech Signals,” Eurasip Journal on Audio, Speech, and Music Processing, Vol 1, no 42, 2024.

[21] K Shigemi, S. Koyama, T. Nakamura and H. Saruwatari, “Physics-Informed Convolutional Neural Network with Bicubic Spline Interpolation for Sound Field Estimation,” in Proceedings of 2022 International Workshop on Acoustic Signal Enhancement (IWAENC), 05-08 September, 2022, Bamberg, Germany.

[22] S. Papadimitropoulos, C. Tsogka and M. Hasan, “Synthetic Aperture Imaging Using Physically Informed Convolutional Neural Networks,” in Proceedings of 2024 IEEE Conference on Computational Imaging Using Synthetic Apertures (CISA), May 20-23, 2024, Boulder, USA.

[23] S. Cuomo, V.S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi and F. Piccialli, “Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What’s Next,” Journal of Scientific Computing, Vol. 92, no 88, 2022.

[24] A. Farea, O. Yli-Harja and F. Emmert-Streib, “Understanding Physics-Informed Neural Networks: Techniques, Applications, Trends, and Challenges,” AI, Vol. 5, no 3, pp. 1534-1557, 2024.

[25] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear PDEs,” Journal of Computational Physics, Vol 378, pp. 686-707, 2019.

[26] S. Lois, R. Kotsakis, C. Sevastiadis, N. Vryzas, L. Vrysis, C. Dimoulas, G. Kalliris, “Simulation of room acoustic response for arbitrary selection of receiver-source position”, presented in National Acoustics Conference 2024, 17-19 October, Rethymno, Greece.

[27] R. Stewart and M. Sandler, "Database of Omnidirectional and B-Format Impulse Responses", in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010), Dallas, Texas, March 2010.

[28] A. Farina, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique,” in Proceedings of 108th AES Convention, Paris, France, February 2000.

[29] X. Karakonstantis, D. Caviedes-Nozal, A. Richard and E. Fernandez-Grande, “Room impulse response reconstruction with physics-informed deep learning,” Journal of Acoustical Society of America, Vol. 155, no 2, pp. 1048–1059, 2024.