Data-Driven Physics-Informed Neural Network for Sound Field Estimation in Rooms of Arbitrary Size

Gen Sato* and Yusuke Ikeda*
* Tokyo Denki University, Japan
E-mail: 24udc03@ms.dendai.ac.jp, yusuke.ikeda@mail.dendai.ac.jp

***Abstract*—In recent years, many methods for simulating acoustic fields using deep learning have been proposed. In particular, neural networks using physical knowledge (PINN: Physics-Informed Neural Network) have been proposed for the purpose of solving various acoustic problems with higher accuracy. Many of the proposed PINN-based methods have achieved the estimation of a specific sound field. In this study, we propose a data-driven PINN for the estimation of simulated sound fields in rooms of arbitrary size with a single model. By using the coordinates and room size information as the input of PINN, the proposed method demonstrated a notable improvement in estimation accuracy by approximately 5.4 dB compared to the simple data-driven model.**

I. Introduction

Sound field simulation has applications in a variety of fields, including spatial acoustics, architectural acoustics [1], games, and sound rendering in mixed reality [2]. Sound field simulation methods can be divided into two main categories: simulations based on geometric acoustics [3] and simulations based on the wave equation [4]–[6]. Ray tracing is a well-established simulation method based on geometric acoustics [7]. By modeling sound propagation as rays, the sound field can be calculated at high speed. In contrast, numerical simulations that approximate the wave equation provide highly accurate simulations of sound fields, such as physical wave phenomena including diffraction and scattering. Previously, many methods have been proposed, including finite-difference time-domain method (FDTD) [4], finite-element method (FEM) [5], and boundary-element method (BEM) [6]. These methods can be computed with high accuracy, however, they are known to require a lot of computation time and computational resources.

Consequently, many studies have been conducted to improve processing speed and reduce the resources required for calculations. For example, Yoshida *et al.* achieved a reduction in the computational time by developing a method for minimizing the dispersion error that occurs in the time-domain FEM [8]. Wilkes *et al.* proposed a method for solving broadband sound fields in parallel using BEM based on fast multipole solutions [9].

In recent years, data-driven deep learning for sound field simulation has been proposed. These methods learn the properties of PDEs from a large number of sound fields computed using conventional numerical simulations based on the wave equation. A trained model that approximates the PDE allows for fast simulation of the sound field. Ziqi *et al.* proposed a method for estimating the sound pressure of a scattered sound field from the shape of the scattering object using convolutional neural networks, which are employed in tasks such as semantic segmentation. This method was demonstrated to be 100 times faster than conventional numerical simulations. [10]. Nikolas *et al.* have proposed a room sound-field simulation using a deep operator network [11]. In this method, the user can freely select the location of the sound source and the listener. Furthermore, this method can estimate sound pressure within 100 ms.

Furthermore, a deep learning method called a physics-informed neural network (PINN), which learns based on the governing equation of physical laws, has been proposed [12]–[14]. In particular, many PINN studies have been conducted in acoustics using the wave equation [15]–[19]. Xenofon *et al.* applied PINN to sound field reconstruction to achieve more accurate sound field interpolation than compressed sensing [15]. In addition, Xingyu *et al.* employed the PINN approach to estimate the acoustic scattering from measured signals with a limited number of microphones [18]. As shown above the PINN approach has proven effective in enhancing the precision of sound field estimation. From previous studies, the PINN approach is effective in improving the accuracy of sound field estimation. However, the model must be modified when the acoustic environment changes, such as when the shape of the room changes.

In this study, we propose a sound field simulation using data-driven deep learning with PINN. In particular, this study estimates the scattered sound field for a room of arbitrary size. The input of PINN is a single coordinate and a room size. The output is the sound pressure at the input coordinates in a room of the input size. For the training dataset, many sound fields with different room sizes are used. In the simulation experiments, we evaluate the estimation accuracy of the proposed method with a comparison of the model learned by data only.

The remainder of this paper is organized as follows. Section II introduces the problem statement and the proposed method. Section III presents the simulation experiments for evaluation in a two-dimensional sound field. Finally, we conclude the paper in Section IV.

II. Method

A. Problem Statement

To simplify, we consider the two-dimensional sound fields with a Dirac impulse emitted from a line sound source at position $\mathbf{x}_{\mathrm{src}} = (x_{\mathrm{src}}, y_{\mathrm{src}})$. Our study focuses on estimating

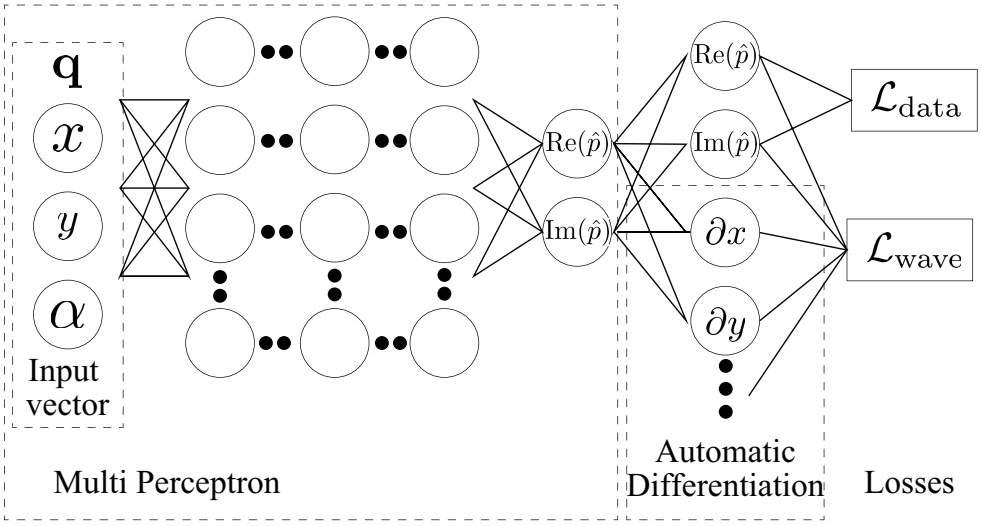

Fig. 1. Workflow of the proposed method. The real and imaginary parts of the complex sound pressure at the position (x, y) are estimated from the input vector $\mathbf{q}$ (x, y: two-dimensional plane coordinates, α: information on the room shape). $\mathcal{L}_{\mathrm{data}}$ calculates the mean squared error using the estimated complex sound pressure and teaching data. $\mathcal{L}_{\mathrm{wave}}$ calculates the Helmholtz equation using the Laplacian calculated by automatic differentiation and the estimated complex sound pressure. By using $\mathcal{L}_{\mathrm{data}}$ and $\mathcal{L}_{\mathrm{wave}}$, the backpropagation is conducted.

complex sound pressure $p_{\mathrm{all}}(\mathbf{x})$ at any position $\mathbf{x}$ in the sound field Ω of arbitrary room size. The sound pressure $p_{\mathrm{all}}(\mathbf{x})|\mathbf{x} \in \Omega$ at any position $\mathbf{x}$ in the room sound field Ω follows the Helmholtz equation holds:

$$(\nabla^2 + k^2)p_{\mathrm{all}}(\mathbf{x}) = -p_{\mathrm{src}}(\mathbf{x}), \tag{1}$$

where, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator, k is the wave number and p_{src} is line sound source.

From the Helmholtz equation in (1) and assumption of linearity, the sound pressure $p_{\mathrm{all}}(\mathbf{x})$ is represented by

$$p_{\mathrm{all}}(\mathbf{x}) = p_{\mathrm{src}}(\mathbf{x}) + p_{\mathrm{ref}}(\mathbf{x}) \tag{2}$$

$$p_{\mathrm{src}}(\mathbf{x}) = G(\mathbf{x}_{\mathrm{p}}, \mathbf{x}_{\mathrm{src}}) \tag{3}$$

where $p_{\mathrm{ref}}(\mathbf{x})$ is the reflected sound field. p_{src} is the direct sound which is represented by Green's function G. Furthermore, based on the Kirchhoff–Helmholtz integral, the reflected sound $p_{\mathrm{ref}}(\mathbf{x})$ is represented by

$$p_{\mathrm{ref}}(\mathbf{x}) = -\int_{\Gamma}(p(\mathbf{x}_\gamma)\frac{\partial G(\mathbf{x}, \mathbf{x}_\gamma)}{\partial n} - G(\mathbf{x}, \mathbf{x}_\gamma)\frac{\partial p(\mathbf{x}_\gamma)}{\partial n})d\Gamma, \tag{4}$$

where $\mathbf{x}_\gamma$ is the position of the boundary Γ of the sound field Ω and $\frac{\partial}{\partial n}$ represents the derivative in the outward normal direction to the boundary Γ.

In general, the direct sound field can be obtained analytically with a model such as the Green's function, for example, in order to use a simple, well-defined model of the source in the simulation. In contrast, the reflected sound field depends on the sound pressure and particle velocity at the boundary Γ. The boundaries of the sound field considered in this study correspond to the boundaries between the air and the material, such as the walls of a room. Therefore, the reflected sound field depends on the boundary conditions such as the shape and size of the room and the sound absorption coefficient. Hence, in this study, only the reflected sound field is the object of estimation, so p_{ref} will be denoted as p in the following sections.

B. Workflow and Input Vector

In this study, based on PINN, we proposed the estimation method for the complex sound pressure $p(\mathbf{x})$ of the reflected sound field at coordinate $\mathbf{x}$ in a room of arbitrary size. Fig. 1 shows the workflow in the proposed method.

In the proposed method, a multilayer perceptron (MLP) estimates the real $\mathrm{Re}(\hat{p}(\mathbf{q}))$ and imaginary $\mathrm{Im}(\hat{p}(\mathbf{q}))$ parts of the complex sound pressure $\hat{p}(\mathbf{q})$ corresponding to the input vector $\mathbf{q}$. PINN employs a PDE for the loss function and learns to consider the physical characteristics of the output. In this study, since the sound pressure is estimated in a frequency domain sound field, the Helmholtz equation in (1) is used as the PDE. Therefore, MLP is learned using $\mathcal{L}_{\mathrm{data}}$, which evaluates the error of the estimated sound pressure, and $\mathcal{L}_{\mathrm{wave}}$, which evaluates the Helmholtz equation using the estimated sound pressure. This workflow is repeated to create a PINN model that estimates the complex sound pressure at an arbitrary point for an arbitrary room size.

In general, PINN is well known for estimating physical quantities such as sound pressure corresponding to input coordinates. The input coordinates are not restricted to specific coordinates such as grid points. Thus, since the model learns the mapping between the coordinates and the data, if the relationship between the two changes, a different model will be required. However, in this study, it is necessary to learn the reflected sound fields of different room sizes. Thus, in the proposed method, the room size information is added to the input vector $\mathbf{q} = (x, y, \alpha)$. The scalar α is the information of room size for the estimated sound field.

C. PINN Architecture

In this study, the MLP for estimating the sound complex pressure $\hat{p}(\mathbf{q})$ at any point in any reflected sound fields can be represented as follows using the function $f(\cdot)$.

$$\hat{p}(\mathbf{q}) = f(\mathbf{w}, \mathbf{q}) = (\phi_N \circ \phi_{N-1} \circ \cdots \phi_n \circ \cdots \phi_1)(\mathbf{q}), \quad (5)$$

where, $\mathbf{q}$ represents the input vector and $\mathbf{w}$ represents the learnable weights. ϕ is a function in the N-th layer. In other words, MLP is represented by a composition function of ϕ.

The function ϕ in the n-th layer is

$$\phi_n(\mathbf{q}_n) = \sin(\omega_0(\mathbf{q}_n \mathbf{w}_n + \mathbf{b}_n)), \quad (6)$$

where, the input vector in the n-th layer is represented by $\mathbf{q}_n$, while the learnable weights and biases in this layer are represented by $\mathbf{w}_n$ and $\mathbf{b}_n$, respectively. ω_0 is a hyperparameter when a sine function is used as the activation function [20]. Since this network can calculate the second-order derivative through automatic differentiation, the Helmholtz equation is employed for the PINN loss function. Compared to other activation functions, such as hyperbolic tangent and rectified linear units, the Sine function is known to be more accurate in high-order partial derivatives. Thus, using a sine function for the activation function, PINN improves the estimation accuracy [15]–[17].

The loss function $\mathcal{L}$ for each α is given by

$$\mathcal{L} = \mathcal{L}_{\mathrm{data}} + \lambda \mathcal{L}_{\mathrm{wave}}, \quad (7)$$

$$\mathcal{L}_{\mathrm{data}} = \frac{1}{M_{\mathrm{data}}} \sum_{m \in \mathcal{M}_{\mathrm{data}}} |\hat{p}_m - p_m|^2, \quad (8)$$

$$\mathcal{L}_{\mathrm{wave}} = \frac{1}{M_{\mathrm{data}}} \sum_{m \in \mathcal{M}_{\mathrm{data}}} |(\nabla^2 + k^2)\hat{p}_m| + \frac{1}{M_{\mathrm{wave}}} \sum_{m \in \mathcal{M}_{\mathrm{wave}}} |(\nabla^2 + k^2)\hat{p}_m|, \quad (9)$$

where λ is the balance parameter. $\mathcal{M}_{\mathrm{data}}$ is a set of sound pressure indices of training data for each α, $\mathcal{M}_{\mathrm{wave}}$ is a set of sound pressure indices that only calculates $\mathcal{L}_{\mathrm{wave}}$ for each α. p_m and $\hat{p}_m$ denote m-th ground truth and estimated complex sound pressure, respectively. M_{data} and M_{wave} are the number of elements in each set. $\mathcal{L}_{\mathrm{data}}$ calculates the mean squared error using training data and estimated data of sound pressure ($\hat{p}_m \in \mathcal{M}_{\mathrm{data}}$). $\mathcal{L}_{\mathrm{wave}}$ is a loss function for learning so that the sound pressure satisfies the laws of physics, based on the homogeneous Helmholtz equation.

III. Experiments

In this experiment, we compare the estimation accuracies with the proposed method and simple MLP only using the loss function in (8).

TABLE I
Condition of BEM

Sound speed [m/s]	340
Frequency [Hz]	500
Sound source coordinates	$(0, 0)$
Normal incidence sound absorption coefficient	0.5

A. Experiment Condition

The pair data of input vector $\mathbf{q}$ and complex sounds pressure $p(\mathbf{q})$ were used as the training dataset. For the training data, the reflected sound fields at 500 Hz were calculated by the BEM [21]. Table I shows the simulation conditions of the BEM. The spacing between boundary elements was 0.1 m (approximately $1/6$ of the wavelength) or less. To simplify the problem, the experiment was conducted with the sound fields in square rooms. The length of one side of the square room was 2–4.5 m. Note that the target region was limited to a 2 m square region centered on the origin, which is part of the sound field of the square room. The information α is the side length of the room. A total of 251 sound fields for training and evaluation were simulated by modifying α in increments of 0.01 m. Three different datasets were constructed, with α increments d being set to 0.03 m, 0.04 m, and 0.05 m, respectively. Experiments were then conducted to compare these data sets, which contained 84, 63, and 51 data numbers, respectively. The test dataset for evaluation is other than the data used for training.

For datasets, the input coordinates were chosen randomly, the number $M_{\mathrm{data}}(\alpha)$ was $(\alpha/0.2)^2$ such that the density was 1/0.04 m^2 per the room side length α. In other words, the each sound field was discretized by about $1/3$ of a wavelength. For the purpose of comparing methods of determining input coordinates, the other test dataset was created with a grid of 0.1 m intervals for the input coordinates. The number of coordinates M_{wave} was the same as M_{data}. The coordinates for calculation of the loss $\mathcal{L}_{\mathrm{wave}}$ were randomly determined for each iteration. The input coordinates, complex sound pressure, and α were normalized within the range of -1 to 1.

The network was a five-layer MLP comprising 256 neurons in each layer. The optimization method employed Adam, with a learning rate of 1.0×10^{-5}. The hyperparameter ω_0 was set to 30, and the balance parameter λ was set to 1.0×10^{-3}. The batch processing was conducted for each room size α. The loss for validation was calculated for every 10 epochs on the test dataset with random coordinates. The model with the smallest loss value within 200 epochs was used for evaluation.

The accuracy of the estimated sound field was evaluated using the following signal-to-noise ratio (SNR).

$$\mathrm{SNR} = 10 \log_{10} \frac{|p(\mathbf{x})|^2}{|p(\mathbf{x}) - \hat{p}(\mathbf{x})|^2} \quad (10)$$

where, $p(\mathbf{x})$ is ground truth at position $\mathbf{x}$. $\hat{p}(\mathbf{x})$ is the estimated sound pressure.

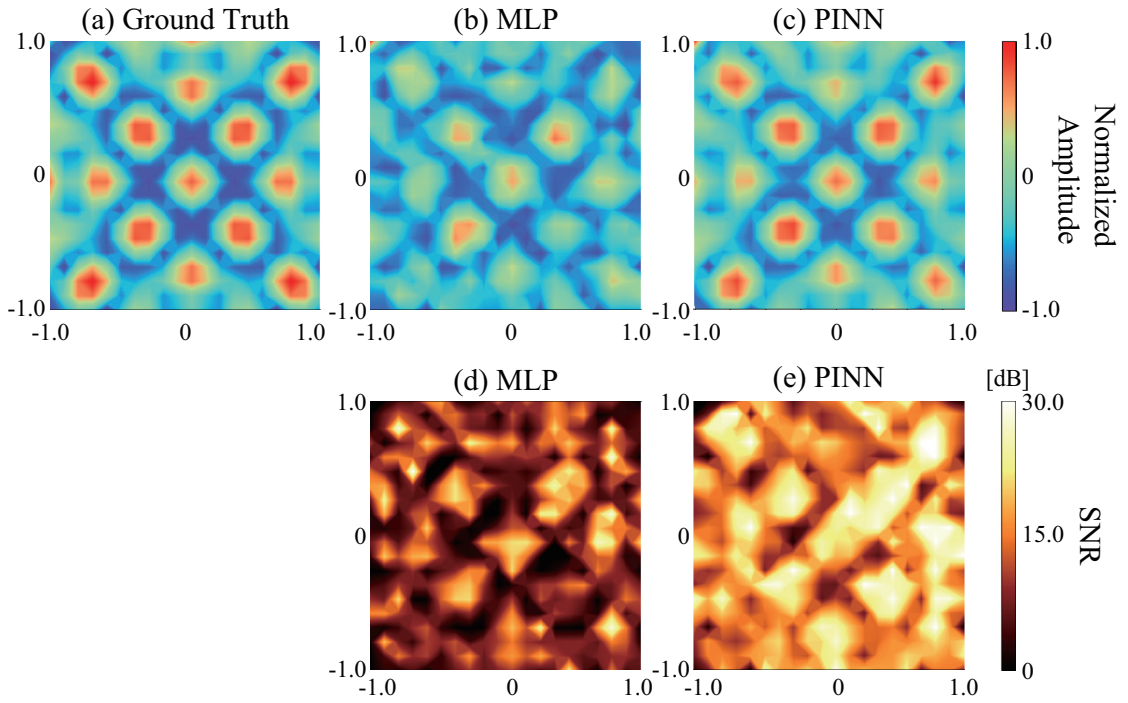

Fig. 2. Comparison of estimated sound fields at $\alpha = 2.350$ m with a model trained on a dataset with spacing $d = 0.04$ m. The estimation region is $[-1 1]$ m. (a) Ground truth calculated by BEM, (b)–(c) and Estimated sound fields by MLP and PINN (proposed method), (d) and (e) SNR distributions by MLP and PINN

TABLE II
MEAN SNR OF THE MODEL TRAINED BY RANDOM COORDINATE DATASET

Increment d	0.03 m	0.04 m	0.05 m
MLP	9.6 dB	7.4 dB	5.8 dB
PINN(Proposal)	13.5 dB	12.8 dB	11.2 dB

TABLE III
MEAN SNR OF THE MODEL TRAINED BY GRID COORDINATE DATASET

Increment d	0.03 m	0.04 m	0.05 m
MLP	9.0 dB	7.1 dB	5.4 dB
PINN(Proposal)	12.9 dB	12.2 dB	10.6 dB

B. Results

Models were created using three training datasets with different increments d of α. Table II and Table III show the mean SNR for the test dataset.

From Table. II, the proposed method improved the estimation accuracy with all three models compared to MLP. When d = 0.04 and 0.05 m, the proposed method improved the estimation accuracy by 5.4 dB. When d = 0.03 m, the difference in estimation accuracy between MPL and PINN was 3.9 dB smaller compared to d = 0.04 m and 0.05 m. Thus, the difference in estimation accuracy between MLP and PINN tended to become smaller as d became smaller, i.e., as the number of training data increased.

From Table. III, the proposed method with the model trained by grid coordinates slightly decreased the estimation accuracy by 0.6 dB compared to the random coordinates in Table. II. However, the proposed method improved the accuracy by 3.9 dB, 5.1 dB, and 5.2 dB at d = 0.03, 0.04, and 0.05, respectively.

Fig. 2 shows the estimated sound field at $\alpha = 2.350$ m utilizing a model trained by the dataset with $d = 0.04$ m. From Figs. 2(a)(b), the estimated sound field by MLP, data-driven models, was significantly degraded over the estimation region. From Figs. 2(a)(c), the sound field estimated by PINN closely approximates the ground truth, compared to MLP. From Figs. 2(d)(e), the proposed method significantly improved the SNR distribution compared to MLP. In this condition, the mean SNRs of MLP and PINN were 7.4 dB and 14.7 dB, respectively. Thus, the PINN improved mean SNR by 7.3 dB than MLP.

IV. CONCLUSIONS

In this study, we proposed the estimation method for the sound field of arbitrary room size based on PINN and data-driven MLP. From the experiments, it is clear that the proposed method can estimate the sound field of a room of a size not included in the training with higher estimation accuracy than a model learned from data only. We also found that PINN improves estimation accuracy compared to MLP with smaller numbers of training data. Furthermore, the utilization of coordinates as input vectors enables the estimation of the sound field without the constraints imposed by grids.

In future work, to model more general sound fields, we will investigate PINN using the other characteristics of the sound field as input, such as the sound absorption coefficients and the position of the sound source.

Acknowledgment

This work was partially supported by Tokyo Denki University Special Academy Research Fund 224S40.

References

[1] T. Yoshida, T. Okuzono, and K. Sakagami, "A parallel dissipation-free and dispersion-optimized explicit time-domain fem for large-scale room acoustics simulation," *Buildings*, vol. 12, no. 2, 2022.

[2] N. Raghuvanshi and J. Snyder, "Parametric directional coding for precomputed sound propagation," *ACM Trans. Graph.*, vol. 37, no. 4, Jul. 2018.

[3] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 708–730, Aug. 2015.

[4] T. Tsuchiya, "Recent techniques on sound field simulation," *Japanese Journal of Applied Physics*, vol. 61, no. SG, SG0801, 2022.

[5] L. L. Thompson, "A review of finite-element methods for time-harmonic acoustics," *The Journal of the Acoustical Society of America*, vol. 119, no. 3, pp. 1315–1330, Mar. 2006.

[6] S. Kirkup, "The boundary element method in acoustics: A survey," *Applied Sciences*, vol. 9, no. 8, 2019.

[7] A. Krokstad, S. Strom, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, 1968.

[8] T. Yoshida, T. Okuzono, and K. Sakagami, "Dissipation-free and dispersion-optimized explicit time-domain finite element method for room acoustic modeling," *Acoustical Science and Technology*, vol. 42, no. 5, pp. 270–281, 2021.

[9] D. R. Wilkes, A. J. Duncan, and S. Marburg, "A parallel and broadband helmholtz fmbem model for large-scale target strength modeling," *Journal of Theoretical and Computational Acoustics*, vol. 28, no. 03, p. 2 050 001, 2020.

[10] Z. Fan, V. Vineet, H. Gamper, and N. Raghuvanshi, "Fast acoustic scattering using convolutional neural networks," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 171–175.

[11] N. Borrel-Jensen, S. Goswami, A. P. Engsig-Karup, G. E. Karniadakis, and C.-H. Jeong, "Sound propagation in realistic interactive 3d scenes with parameterized sources using deep neural operators," *Proceedings of the National Academy of Sciences*, vol. 121, no. 2, e2312159120, 2024.

[12] 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.

[13] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations," *arXiv preprint arXiv:1711.10561*, 2017.

[14] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part ii): Data-driven discovery of nonlinear partial differential equations," *arXiv preprint arXiv:1711.10566*, 2017.

[15] X. Karakonstantis, D. Caviedes-Nozal, A. Richard, and E. Fernandez-Grande, "Room impulse response reconstruction with physics-informed deep learning," *The Journal of the Acoustical Society of America*, vol. 155, no. 2, pp. 1048–1059, Feb. 2024.

[16] M. Pezzoli, M. Cobos, F. Antonacci, and A. Sarti, "Sparsity-based sound field separation in the spherical harmonics domain," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 1051–1055.

[17] I. Tsunokuni, G. Sato, Y. Ikeda, and Y. Oikawa, "Spatial extrapolation of early room impulse responses with noise-robust physics-informed neural network," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. advpub, 2024EAL2015, 2024.

[18] X. Chen, F. Ma, A. Bastine, P. Samarasinghe, and H. Sun, "Sound field estimation around a rigid sphere with physics-informed neural network," in *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2023, pp. 1984–1989.

[19] K. Shigemi, S. Koyama, T. Nakamura, and H. Saruwatari, "Physics-informed convolutional neural network with bicubic spline interpolation for sound field estimation," in *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2022, pp. 1–5.

[20] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proc. NeurIPS*, 2020.

[21] "OpenAcoustics,"
http : / / www . openacoustics . org / sample - page/. Last accessed 2024-8-5.