DOI: 10.61782/fa.2025.0683

FORUM ACUSTICUM
EURONOISE 2025

POINT NEURON LEARNING FOR BROADBAND ARRAY PROCESSING

Amy Bastine **Thushara D. Abhayapala** **Prasanga N. Samarasinghe**
Audio & Acoustic Signal Processing Group,
Australian National University, Australia

ABSTRACT

Whilst Physics Informed Neural Networks (PINNs) solve certain limitations of traditional networks, they also have several drawbacks including inability to approximate PDEs that have sharp gradients, strong non-linearities and convergence to trivial solutions. Recently, we proposed the point neuron network by embedding the free space Green function into the network architecture enabling the learned model to strictly satisfy the physical law of sound propagation. The physical meaning of point neurons is equivalent to point sources or plane wave sources, and the weight, location (biases) and distribution of equivalent sources can be updated while training. In this paper, we extend the point neuron learning network for broadband signals. The proposed point neuron network can be implemented efficiently with fewer network parameters to model and estimate an arbitrary broadband sound field based on microphone observations without a pre-existing data set. As an example application, we use the proposed network to estimate Room Transfer Functions at locations with no measurements.

Keywords: *Physics Informed Neural Networks (PINN), Point Neuron Learning, Broadband Array Processing, Room Impulse Response Reconstruction*

1. INTRODUCTION

Advances in Machine Learning have revolutionized some applications of audio and acoustic signal processing. However, traditional machine learning (ML) algorithms typically require extensive datasets and prolonged training periods. Sampling sound over space using a large number of microphones is impractical and hence there is a lack of spatially rich data compared to temporal sampling of sound waves. Another drawback is in time-critical applications where hours of training is not possible. These facts limit the exploitation of both rich content of spatial sound as well as the power of learning algorithms. To cater for these limitations of traditional ML algorithms, Physics Informed Machine Learning (PIML) [1, 2] where physical constraints such as governing partial differential equations or boundary conditions have been added to the usual data driven loss function. There are recent PIML based solutions to some acoustic signal processing applications such as sound field estimation [3, 4] and reconstruction [5], nearfield acoustics holography [6] and room impulse response reconstruction [7, 8]. Nevertheless, PIML based techniques also have a number of limitations [9] such as being unable to approximate PDEs that have sharp gradients or strong non-nonlinearities, not being able to move away from local optimums, and convergence to trivial solutions.

Recently, we embed the fundamental solution to the wave equation, free space Green function, into the network architecture [10], enabling the learned model to strictly satisfy the physical law of sound propagation. In the proposed network, the basic processing unit is called a *point neuron* whose weight and biases can be learned by back propagation. The physical meaning of point neuron is equivalent to point sources or plane wave sources, and the weight, location (biases) and distribution of equivalent sources can be updated while training. The proposed point neuron network can be implemented to model and estimate an arbitrary sound field purely based on microphone observations without a preexisting data set. In this paper, we extend [10] for broadband array processing.

*__Corresponding author__: Amy.Bastine@anu.edu.au.

2. PROBLEM FORMULATION

Consider a source-free region $\Omega \subset \mathbb{R}^3$ surrounded by sound sources. The resulting sound field is observed at a finite set of spatial locations $\{\boldsymbol{x}_q = (x_q, y_q, z_q)\}_{q=1}^{Q} \subset \Omega$. The sound pressure at each observation point $\boldsymbol{x}_q$ is denoted by $P_m(\boldsymbol{x}_q, k) \in \mathbb{C}$ in the frequency-domain, where $k \in \mathcal{K} = \{k_1, \cdots, k_K\} \subset \mathbb{R}_+$ is a discrete set of wavenumbers defined by $k_{\mathrm{i}} = 2\pi f_{\mathrm{i}}/c$ for $\mathrm{i} = \{1, \cdots, K\}$, with f_{i} and c representing the i^{th} frequency and speed of sound, respectively.

We aim to build a neural network model $\mathcal{P}(\boldsymbol{x}, k; \boldsymbol{\mu})$, parameterized by $\boldsymbol{\mu} \in \mathbb{C}^S$, to reconstruct the sound pressure $P(\boldsymbol{x}, k)$ for any $\boldsymbol{x} \in \Omega \; \forall k$ based on Q observation points, while adhering to the Helmholtz equation which governs the wave propagation over space. We formulate $\mathcal{P}(\boldsymbol{x}, k; \boldsymbol{\mu})$ by the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{\mu} \in \mathbb{C}^S}{\arg\min} \quad & \mathcal{L} = \mathcal{L}_{\mathrm{TRN}}\left(\hat{\boldsymbol{P}}, \boldsymbol{P}_m\right) + \lambda \mathcal{C}(\mu) && \text{(1a)} \\ \text{s.t.} \quad & \Delta^2 \mathcal{P}(\boldsymbol{x}, k; \boldsymbol{\mu}) + k^2 \mathcal{P}(\boldsymbol{x}, k; \boldsymbol{\mu}) = 0, && \text{(1b)} \\ & \boldsymbol{x} \in \Omega, \\ & \lambda \in [0, \infty), \end{aligned}$$

where $\mathcal{L}$ is the cost function and $\mathcal{L}_{\mathrm{TRN}}$ is the training loss that measures the supervised error between the model output $\hat{\boldsymbol{P}} \in \mathbb{C}^{Q \times K}$, given by $\mathcal{P}(\boldsymbol{x}_q, k_{\mathrm{i}}; \boldsymbol{\mu})$ evaluated at the observation points $\boldsymbol{x}_q$ for all $k_{\mathrm{i}} \in \mathcal{K}$, and the corresponding observed data $\boldsymbol{P}_m \in \mathbb{C}^{Q \times K}$. The hyperparameter λ controls the model complexity loss $\mathcal{C}(\boldsymbol{\mu})$, and Δ^2 is the Laplacian operator.

Standard physics-based learning approaches address the optimization problem by incorporating the physical constraint in Eqn. (1b) into the loss function defined in Eqn. (1a), thereby encouraging approximate satisfaction of the Helmholtz equation. However, most existing models lack the flexibility to enforce this constraint directly within their architectures. To address this limitation, a novel architecture called the point neuron network was proposed in [10], which inherently satisfies the Helmholtz equation and provides interpretability grounded in physical principles. In the following section, we extend this architecture to support broadband sound field modeling.

3. BROADBAND POINT NEURON LEARNING

To explicitly satisfy the Helmholtz wave propagation constraint in Eqn. (1b), the point neuron network embeds the fundamental solution of the wave equation given by the free-space Green's function into its architecture. This function,

$$G(\boldsymbol{x}, k|\boldsymbol{y}) = \frac{e^{ik\|\boldsymbol{x}-\boldsymbol{y}\|_2}}{4\pi\|\boldsymbol{x}-\boldsymbol{y}\|_2}, \tag{2}$$

represents the sound field at an observation point $\boldsymbol{x}$ due to a point source of unit strength located at $\boldsymbol{y}$, where $i = \sqrt{-1}$ and $\|\cdot\|_2$ denotes ℓ_2-norm.

To ensure unified treatment of both near-field and far-field sources, Eqn. (2) is normalized and reformulated as the point neuron unit, which serves as the core building block of the network. It is defined as

$$PN(\boldsymbol{x}, k|\boldsymbol{y}) = \|\boldsymbol{y}\|_2 \, e^{-ik\|\boldsymbol{y}\|_2} \frac{e^{ik\|\boldsymbol{x}-\boldsymbol{y}\|_2}}{4\pi\|\boldsymbol{x}-\boldsymbol{y}\|_2}. \tag{3}$$

3.1 Network architecture

Figure 1 illustrates the proposed broadband point-neuron network architecture. The model takes the spatial coordinate of an arbitrary point $\boldsymbol{x} = (x, y, z)$ and the wavenumber $k \in \mathcal{K} = \{k_1, \cdots, k_K\}$ as inputs, and outputs the estimated broadband sound pressure $\hat{P}(\boldsymbol{x}, k)$ at that coordinate, evaluated over all $k \in \mathcal{K}$. The network comprises V point-neuron units, each functioning as a virtual source that contributes to the reconstruction of the sound field.

As shown in Fig. 1a, each unit is parameterized by a spatial bias vector $(B_v^x, B_v^y, B_v^z) \in \mathbb{R}^3$ and frequency-dependent weights $w_v(k) \in \mathbb{C}^K$. The signal flow within each unit implements the point-neuron transfer function defined in Eqn. (3), weighted by $w_v(k)$, using the 0^{th} order spherical Hankel function of the first kind, $h_0^{(1)}(r) = e^{ir}/ir$.

The spatial bias and weights can be interpreted as the locations and strengths of virtual sources that can be optimized during training. By fully connecting and linearly combining the outputs of all V units, as depicted in Fig. 1b, the network reconstructs the sound field via the superposition of learned virtual sources, maintaining physical consistency across both space and frequency.

While the architecture is derived from the normalized Green's function, it remains fully compatible with standard neural network training procedures. Embedding the Green's function into the network grounds the model in physical principles, reducing reliance on purely data-driven learning and the need for large datasets.

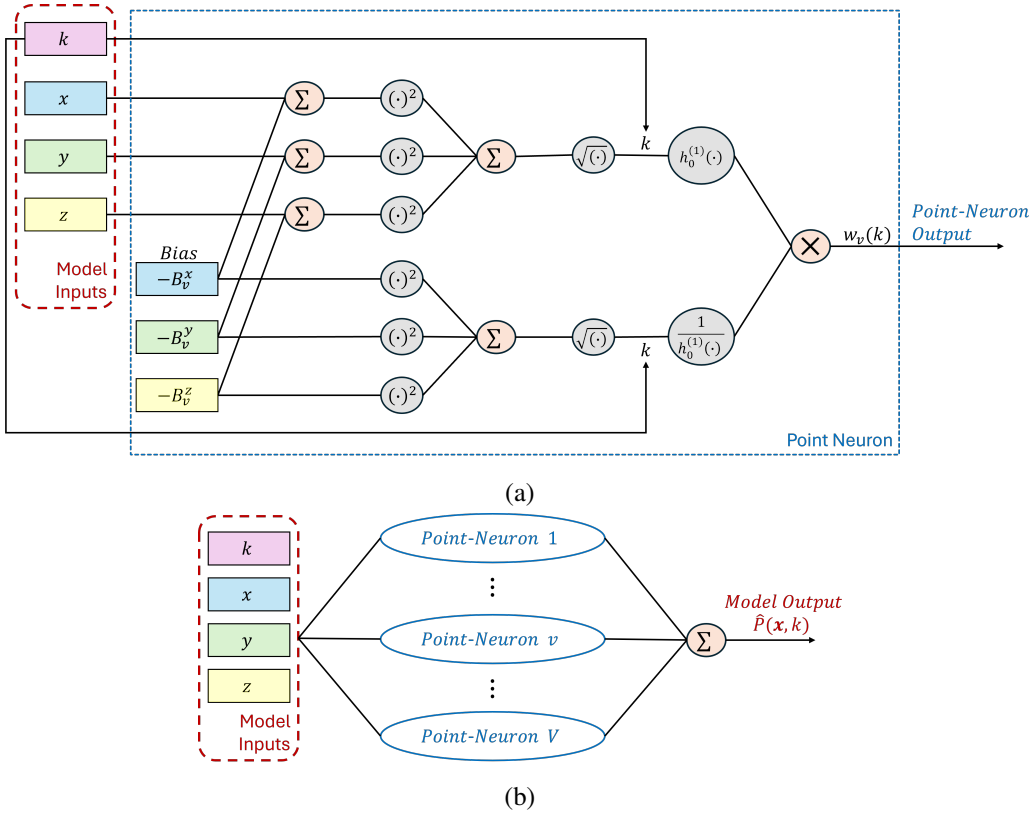

Figure 1: Network architecture of point neuron learning for broadband array processing. (a) Building block of v^{th} point neuron. (b) Full model architecture with V number of point neurons.

3.2 Back propagation and training

Based on the network architecture and optimization problem Eqn. (1), we define the system cost function as

$$\arg\min_{\boldsymbol{\mu}\in\mathbb{C}^S} \quad \mathcal{L} = \sum_{k=k_1}^{k_K}\sum_{q=1}^{Q} \|\hat{P}(\boldsymbol{x}_q,k) - P_m(\boldsymbol{x}_q,k)\|_2^2 + \lambda \sum_{k=k_1}^{k_K} \|\boldsymbol{w}(k)\|_1 \tag{4}$$

s.t. Eqn. (1b),

where

$$\hat{P}(\boldsymbol{x}_q,k) = \sum_{v=1}^{V} w_v(k) \left(\frac{D_v}{D_q^v} e^{ik(D_q^v - D_v)}\right), \tag{5}$$

$$D_v = \sqrt{(B_v^x)^2 + (B_v^y)^2 + (B_v^z)^2}, \tag{6}$$

$$D_q^v = \sqrt{(B_v^x - x_q)^2 + (B_v^y - y_q)^2 + (B_v^z - z_q)^2}, \tag{7}$$

$\boldsymbol{w}(k) = [w_1(k), w_2(k), \cdots, w_V(k)]^T \in \mathbb{C}^{V\times 1}$ and $\|\cdot\|_1$ denotes ℓ_1-norm. We apply the ℓ_1-norm regularization to control the model complexity by limiting the number of active point neurons and to avoid overfitting.

We update the network parameters iteratively through back propagation. With n indicating the iteration index, the frequency-dependent weights of the v^{th} point neuron can be updated as

$$w_v(k; n+1) = w_v(k; n) - \xi_w \frac{\partial \mathcal{L}(n)}{\partial (w_v(k;n))^*}, \tag{8}$$

where ξ_w is the learning rate of the weights and $(\cdot)^*$ de-

notes complex conjugation. The gradient term is given by

$$\frac{\partial \mathcal{L}(n)}{\partial (w_v(k;n))^*} = \sum_{q=1}^{Q} \bar{P}^{(n)}(\boldsymbol{x}_q, k) + \frac{1}{2}\lambda e^{i\theta_v(k;n)}, \quad (9)$$

where

$$\bar{P}^{(n)}(\boldsymbol{x}_q, k) = \left(\hat{P}^{(n)}(\boldsymbol{x}_q, k) - P_m(\boldsymbol{x}_q, k)\right) \times \frac{D_v(n)}{D_q^v(n)} e^{-ik\left(D_q^v(n) - D_v(n)\right)}, \quad (10)$$

$\hat{P}^{(n)}(\boldsymbol{x}_q, k)$ and $\theta_v(k;n)$ denote the model output and phase of the complex weight $w_v(k;n)$, respectively, after the n^{th} iteration.

Similarly, we can update the bias parameters B_v^α, where $\alpha \in \{x, y, z\}$, by

$$B_v^\alpha(n+1) = B_v^\alpha(n) - \xi_B \frac{\partial \mathcal{L}(n)}{\partial B_v^\alpha(n)}, \quad (11)$$

where ξ_B is the learning rate of the bias,

$$\frac{\partial \mathcal{L}(n)}{\partial B_v^x(n)} = \sum_{k=k_1}^{k_K} \sum_{q=1}^{Q} 2\,\Re\left\{ \left(\bar{P}^{(n)}(\boldsymbol{x}_q, k)\right)^* w_v(k;n) \left[\frac{(-ikD_v(n) - 1)}{D_v(n)^2} B_v^x(n) + \frac{(ikD_q^v(n) - 1)}{D_q^v(n)^2} \left(B_v^x(n) - x_q\right) \right] \right\}, \quad (12)$$

$$\frac{\partial \mathcal{L}(n)}{\partial B_v^y(n)} = \sum_{k=k_1}^{k_K} \sum_{q=1}^{Q} 2\,\Re\left\{ \left(\bar{P}^{(n)}(\boldsymbol{x}_q, k)\right)^* w_v(k;n) \left[\frac{(-ikD_v(n) - 1)}{D_v(n)^2} B_v^y(n) + \frac{(ikD_q^v(n) - 1)}{D_q^v(n)^2} \left(B_v^y(n) - y_q\right) \right] \right\}, \quad (13)$$

$$\frac{\partial \mathcal{L}(n)}{\partial B_v^z(n)} = \sum_{k=k_1}^{k_K} \sum_{q=1}^{Q} 2\,\Re\left\{ \left(\bar{P}^{(n)}(\boldsymbol{x}_q, k)\right)^* w_v(k;n) \left[\frac{(-ikD_v(n) - 1)}{D_v(n)^2} B_v^z(n) + \frac{(ikD_q^v(n) - 1)}{D_q^v(n)^2} \left(B_v^z(n) - z_q\right) \right] \right\}, \quad (14)$$

and $\Re$ indicates the real part of the argument.

During training, care must be taken to avoid placing virtual sources at or too close to observation points, as the gradients in Eqns. (12)–(14) become unbounded when $D_q^v(n)$ approaches zero. To mitigate this, the distances should be continuously monitored to satisfy a minimum threshold, and the corresponding spatial biases should be reinitialized to maintain numerical stability.

It is worth noting that the entire training process described above relies exclusively on microphone observations, without the need for additional datasets, making the proposed method particularly suitable for data-scarce acoustic environments.

3.3 System Initialization

The point neuron weights $w_v(k)$ are typically initialized with random complex values constrained in magnitude to $[-1, 1]$. In room acoustic scenarios, where frequency responses often exhibit consistent trends across space, normalized microphone observations can be used to guide the initialization of $w_v(k)$. Similarly, spatial biases may be initialized based on prior knowledge of the acoustic environment. For example, in height-invariant sound fields, initial biases may be distributed in the xy-plane, while in reflective spaces, their placement can be aligned with the geometry of dominant boundaries. Such informed initialization improves training efficiency and reduces the risk of convergence to suboptimal local minima.

4. SIMULATION ANALYSIS

This section presents a preliminary evaluation of the proposed broadband point-neuron network for reconstructing the sound pressure field over a target region using limited microphone observations. We perform this simulation study for a rectangular room measuring $5.2 \times 6.4 \times 4.2$ m, with the coordinate origin defined at the bottom-left corner of the room. A spherical region of 1 m radius, centered at $[2.6,\ 3.2,\ 2.1]$ m, serves as the target region for reconstruction. A total of 75 microphones are randomly distributed within the target region, with their z-coordinates confined to lie within 15 cm of the xy-plane. A single acoustic source is placed at $[4.4,\ 4.8,\ 2.1]$ m. The resulting sound pressure at the microphones is simulated using the image source method [11, 12] for different T_{60} values. The sampling rate is 16 kHz. To simulate real-world conditions, white Gaussian noise is added to the measurements achieving a signal-to-noise ratio of 40 dB.

The weights of the point neuron network are initialized using the normalized average response of the microphone measurements, with random weighting applied assigned across the V units. Spatial biases are initialized over a 3D mesh-grid spanning $10 \times 10 \times 8$ m, with a resolution of 1.5 m in the xy-plane and 4.2 m along the z-axis, totaling to 147 units for the whole broadband range. The regularization parameter is set to $\lambda = 0.01$, and the learning rates for the weights and spatial biases are set to $\xi_w = 0.001$ and $\xi_b = 0.02$, respectively.

To evaluate performance, pressure reconstruction is performed at 1500 locations within the target region, over a frequency range from 100 to 2000 Hz. We compare the proposed method with two established techniques: (i) a least-squares-based plane-wave decomposition (PWD) method [13], and (ii) a kernel interpolation method [14] that uses zeroth-order spherical Bessel function kernels and also constrains the solution to satisfy the Helmholtz equation. Note that, unlike the PWD and kernel methods, which are applied independently at each frequency, the point-neuron network reconstructs $P(\boldsymbol{x}, k)$ jointly across all k, exploiting shared spatial structure and spectral continuity inherent in the broadband sound field.

We measure the reconstruction accuracy using the error metric defined by

$$E(\boldsymbol{x}_t, k) = \frac{|\hat{P}(\boldsymbol{x}_t, k) - P(\boldsymbol{x}_t, k)|^2}{|P(\boldsymbol{x}_t, k)|^2}, \qquad (15)$$

where $\hat{P}(\boldsymbol{x}_t, k)$ and $P(\boldsymbol{x}_t, k)$ denote the estimated and true sound pressures, respectively, at the target location $\boldsymbol{x}_t$. We use the image source method [11, 12] to generate $P(\boldsymbol{x}_t, k)$ for all the target locations. In addition to $E(\boldsymbol{x}_t, k)$, we compute the normalized correlation between $\hat{P}(\boldsymbol{x}_t, k)$ and $P(\boldsymbol{x}_t, k)$ across k to asses spectral consistency.

Figure 2 presents a comparison of the reconstructed sound pressure magnitudes at a target location for $T_{60} = 0.2$ s. The PWD method shows significant deviation from the ground-truth response, yielding a correlation of 0.39. The kernel-based method achieves better accuracy, with a correlation of 0.84. In comparison, the proposed point-neuron network achieves the highest correlation of 0.90, closely matching the true response across the entire frequency range and effectively capturing both fine spectral details and overall magnitude trend.

Figure 3 shows the frequency-dependent reconstruction error $E(\boldsymbol{x}_t, k)$ averaged over 1500 target locations for $T_{60} = 0.2$ s. The PWD method exhibits substantial errors,

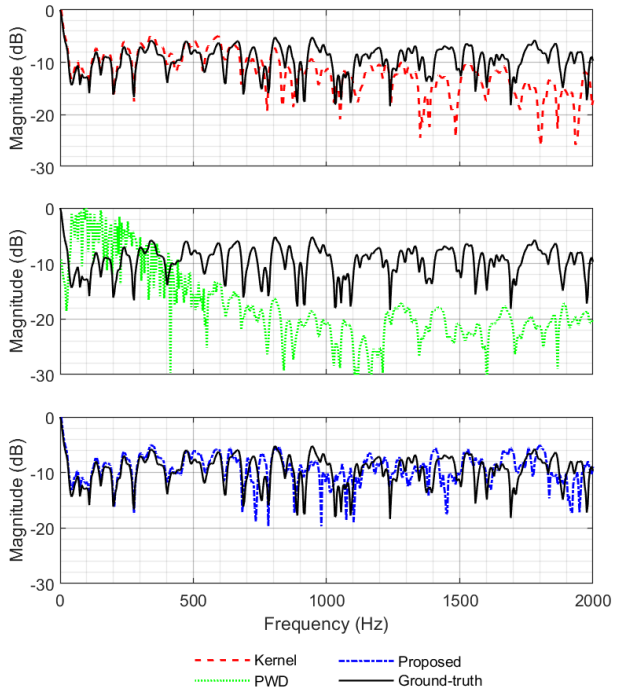

Figure 2: Sound pressure magnitude across frequency at the target location $[2.58,\ 2.21,\ 2.1]$ m for a reverberation time of $T_{60} = 0.2$ s, reconstructed using the proposed, kernel-based, and PWD-based methods against the ground-truth.

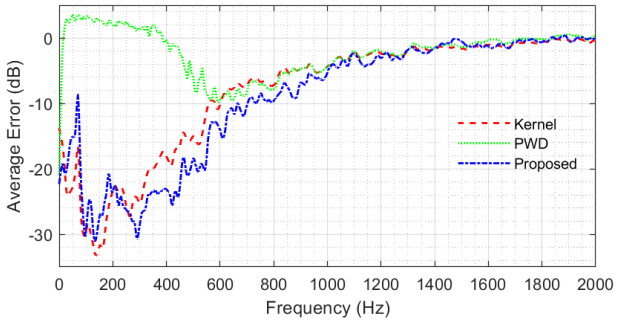

Figure 3: Reconstruction error averaged over 1500 target locations for $T_{60} = 0.2$ s

.

particularly at lower frequencies, attributed to its sensitivity to noise and inversion instability. While the kernel-based method offers improved accuracy over PWD, the proposed point-neuron network achieves the best recon-

struction performance, with notably lower error in the 300-1000 Hz range.

Table 1: Correlation between $P(\boldsymbol{x}_t, k)$ and $\hat{P}(\boldsymbol{x}_t, k)$ across k, reconstructed by different methods at $\boldsymbol{x}_t = [2.58,\ 2.21,\ 2.1]$ m under different T_{60}s.

T_{60} (s)	PWD	Kernel	Proposed
0.2	0.39	0.84	0.90
0.4	0.35	0.91	0.92
0.6	0.37	0.93	0.94

We performed the reconstruction under different reverberant conditions for the target location $\boldsymbol{x}_t = [2.58,\ 2.21,\ 2.1]$ m, and the corresponding correlation values are listed in Table 1. The PWD method consistently yields low correlation values, indicating its limitation in preserving spectral consistency in reverberant environments. The kernel method performs significantly better, with correlation improving as $T60$ increases. The proposed method outperforms both PWD and kernel methods across all T_{60} values, maintaining the highest correlation scores. Moreover, the correlation remains nearly consistent with a slight increase with T_{60}, suggesting the robustness of the proposed method under diverse reverberant conditions. It is worth noting that this performance across different T_{60}s is achieved without re-tuning the initial biases, learning rates, or regularization parameter of the network. With appropriate adjustments to these hyperparameters, the reconstruction performance can potentially be further improved.

5. CONCLUSION

In this paper, we introduced a physics-informed point neuron learning framework for broadband array signal processing, extending our prior narrowband formulation. The proposed architecture models the broadband sound field by learning frequency-dependent weights and shared spatial biases of all point neuron units jointly across the entire frequency range, while strictly adhering to the physical laws of wave propagation. Simulation results demonstrate that the proposed method significantly outperforms conventional PWD and kernel-based methods in reconstructing broadband sound fields, especially under varying reverberant conditions. The network achieves high spectral correlation with a compact architecture and minimal reliance on training data. Future work will focus on validating the approach with real-world measurements, benchmarking against deep neural network-based models, and exploring more compact representations of frequency responses to further improve efficiency.

6. REFERENCES

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

[2] 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," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

[3] K. Shigemi, S. Koyama, T. Nakamura, and H. Saruwatari, "Physics-informed convolutional neural network with bicubic spline interpolation for sound field estimation," in *Proc. Int. Workshop Acoust. Signal Enhancement.*, pp. 1–5, IEEE, 2022.

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

[5] 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. 2024, no. 1, p. 42, 2024.

[6] M. Olivieri, M. Pezzoli, F. Antonacci, and A. Sarti, "A physics-informed neural network approach for nearfield acoustic holography," *Sensors*, vol. 21, no. 23, p. 7834, 2021.

[7] M. Pezzoli, F. Antonacci, and A. Sarti, "Implicit neural representation with physics-informed neural networks for the reconstruction of the early part of room impulse responses," *Preprint at arXiv:2306.11509*, 2023.

[8] 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, 2024.

[9] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating physics-based modeling with machine learning: A survey," *Preprint at arXiv:2003.04919*, vol. 1, no. 1, pp. 1–34, 2020.

[10] H. Bi and T. D. Abhayapala, "Point neuron learning: a new physics-informed neural network architecture," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, p. 56, 2024.

[11] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Soc. America*, vol. 65, pp. 943–950, Apr. 1979.

[12] E. A. P. Habets, "Room Impulse Response Generator," tech. rep., Int. Audio Lab. Erlangen, Sep. 2010.

[13] D. Colton and R. Kress, "Direct Acoustic Obstacle Scattering," in *Inverse Acoustic and Electromagnetic Scattering Theory*, ch. 3, pp. 43–110, Cham, Switzerland: Springer Nature, 4 ed., 2019.

[14] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel Ridge Regression with Constraint of Helmholtz Equation for Sound Field Interpolation," in *Proc. 16th Int. Workshop Acoustic Signal Enhancement*, (Tokyo, Japan), pp. 436–440, Sep. 2018.