Audio Engineering Society

Conference Paper 11

Presented at the AES International Conference on Artificial Intelligence and Machine Learning in Audio
2025 September 8–10, London, UK

This paper was peer-reviewed as a complete manuscript for presentation at this conference. 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.

Complex-valued physics-informed neural networks for sound field estimation

Vlad S. Paul, Nara Hahn, and Philip A. Nelson

Institute of Sound and Vibration Research, University of Southampton, United Kingdom

Correspondence should be addressed to Vlad S. Paul (v.s.paul@soton.ac.uk)

ABSTRACT

Sound field reconstruction from sparse measurements is a challenging problem in acoustics, with applications in room acoustics, spatial audio capture, and active noise control. In this work, a complex-valued physics-informed neural network (PINN) is proposed for reconstructing sound fields from data collected by an open-sphere microphone array at control points distributed on a 2D grid. Unlike conventional PINNs that use only spatial coordinates as input, our model additionally incorporates measured pressure values, enabling it to learn the mapping between observed fields and their spatial distributions. In order to investigate the impact of the physical constraints on the model's performance, the network is trained on sound fields generated from superposed single-frequency plane waves and its performance is compared to that of a complex-valued multilayer perceptron (MLP). Results show that the complex-valued PINN yields smoother training curves and extended regions of low reconstruction error, highlighting its enhanced spatial consistency. Future work will focus on investigating the model's generalization to unseen sound fields and varying control point grids, as well as optimizing the placement and number of training points.

1 Introduction

Sound field estimation involves reconstructing the spatial distribution of acoustical variables from a finite number of discrete measurements. This fundamental approach is used in applications such as room acoustics [1, 2], spatial sound capture [3] and active noise control [4]. Conventional methods often rely on physical models, such as finite element methods (FEM), which discretise the domain to numerically solve the wave equation. While effective, these methods become computationally expensive, especially at higher frequencies, due to the need for dense spatial sampling. Alternatively, analytical approaches like spherical harmonic (SH) expansions provide efficient representations by leveraging known wave solutions but might struggle in complex environments with irregular boundaries or unknown scattering effects.

Virtual sensing extends this concept by estimating the sound field at unmeasured locations, known as "virtual sensors". This approach is particularly useful when direct measurements are impractical or invasive. By leveraging data from available sensors and incorporating physical principles, virtual sensing aims to provide accurate estimations of the sound field in regions of interest.

In recent years, machine learning techniques have been used for sound field estimation applications to enhance accuracy and efficiency [5, 6]. In these data-driven approaches, the network's parameters are adjusted using a training dataset to estimate the sound fields, without any prior knowledge of physical properties. Physics-informed neural networks (PINNs) are a notable development in this area, since these models incorporate the governing physical laws, such as those leading to the wave equation, directly into the loss function of the neural network. This integration ensures that the network's predictions adhere to known physical behaviors, even when trained on limited data [7]. PINNs have recently gained attention in the field of sound field reconstruction. In [8], the authors proposed the use of PINNs to address the issue of spherical Bessel function nulls in open sphere microphone arrays (OSMAs). The model is trained from measurements on an OSMA to estimate the sound field on a sphere of a different radius. This approach leverages the fact that nulls in spherical Bessel functions vary with radius. Using the PINN enables the retrieval of sound field coefficients that would otherwise be difficult to obtain directly from the OSMA measurements. In [9], a PINN-based approach is proposed for recovering acoustic fields, particularly focusing on speech signals. Comparisons with various frequency-domain and time-domain reconstruction benchmark techniques revealed that the proposed PINN method achieved superior accuracy across various measurement setups.

Traditional neural networks are typically real-valued and may struggle with the phase and magnitude relationships present in complex data. Combining PINNs with complex-valued neural networks (CVNNs) offers a promising approach for sound field estimation applications and has been used very recently in [10] for acoustic holography. CVNNs process complex numbers directly, allowing them to inherently represent the phase relationships crucial for accurate sound field representation. The benefits of CVNNs have been investigated in various works such as [11, 12, 13] and it has been shown that in some cases, using complex numbers during training increases the performance of the model compared to using just real data.

This work will investigate the use of complex-valued PINNs for a sound field reconstruction task. Individual sound fields, each generated from superposed single-frequency plane waves, are measured using a microphone array placed on the surface of an open sphere. The task is to estimate the pressure at some control points on the horizontal plane inside the open sphere. Unlike conventional PINNs, which typically take only spatial (and possibly temporal) coordinates as inputs, the proposed model incorporates the measured sound fields into the training data as an additional condition. By including both microphone pressure measurements and spatial coordinates as inputs, the network can infer the relationship between multiple sound field measurements and their corresponding spatial distributions. This design is inspired by the work of [14], who introduced Conditional PINNs. These models incorporate "tag vectors" as conditional inputs to represent a family of physical scenarios.

Such an approach has promising implications for a range of audio applications. For instance, in Active Noise Control (ANC) systems, accurately reconstructing various sound fields at virtual sensor locations is essential for generating the anti-noise signals needed to cancel unwanted sounds. Moreover, this method can enhance immersive audio capturing and reproduction, where high-fidelity spatial reconstruction enables 3D sound experiences for virtual reality (VR) and augmented reality (AR) environments.

2 Problem Formulation

2.1 Preliminaries

Assume a three-dimensional space $\mathbb{R}^3$, where a spatial coordinate vector that represents a point in space is defined as $\mathbf{r} = (x, y, z)$. The acoustic pressure field $p(\mathbf{r}, k)$ is governed by the Helmholtz equation, which describes the propagation of time-harmonic waves. For a given frequency f and corresponding wavenumber $k = 2\pi f/c$, where c is the speed of sound, the Helmholtz equation is expressed as

$$\nabla^2 p(\mathbf{r}, k) + k^2 p(\mathbf{r}, k) = 0, \quad \mathbf{r} \in \Omega \tag{1}$$

where ∇^2 denotes the Laplacian with respect to $\mathbf{r}$ and where Ω is the spatial domain of interest. Consider a microphone located at position $\mathbf{r}_\mathrm{m}$ and a plane wave propagating in the direction specified by the unit vector $\hat{\mathbf{r}}_\mathrm{s}$. The pressure at $\mathbf{r}_\mathrm{m}$ for a single-frequency plane wave is given by

$$p(\mathbf{r}_\mathrm{m}, k) = A e^{-j(k\hat{\mathbf{r}}_\mathrm{s} \cdot \mathbf{r}_\mathrm{m} + \alpha)}, \tag{2}$$

where A is the amplitude of the plane wave. Here, $\hat{\mathbf{r}}_\mathrm{s} \cdot \mathbf{r}_\mathrm{m}$ denotes the dot product between the propagation direction and the microphone position, effectively projecting the position onto the direction of the incident wave. Note that $\hat{\mathbf{r}}_\mathrm{s}$ is normalised to unit length, ensuring it represents the direction of wave propagation. The phase term α can be used to add a random offset to simulate slight differences in the arrival times of the plane wave at different microphones in the array.

2.2 Data model

Consider a simulated microphone array positioned on an open sphere of radius $R_\mathrm{mic} = 0.4$ m in free-field conditions. The microphone array contains 64 sensors positioned at the same angular positions (azimuth and colatitude) as those of an Eigenmike em64 [15], so that an even coverage of the open sphere is ensured. The sensors capture the sound field of superposed plane waves, constituting a spatial sampling of the complex spectrum. Assuming the positions of the 64 sensors are given by $\{\mathbf{r}_{\mathrm{m}_1}, \mathbf{r}_{\mathrm{m}_2}, \ldots, \mathbf{r}_{\mathrm{m}_{64}}\}$, the corresponding sensor measurements form a complex vector $\mathbf{p}_\mathrm{m} = \{p(\mathbf{r}_{\mathrm{m}_1}, k), p(\mathbf{r}_{\mathrm{m}_2}, k), \ldots p(\mathbf{r}_{\mathrm{m}_{64}}, k)\} \in \mathbb{C}$ based on Eq. (2). A set of control points are defined inside the open sphere

$$\Omega = \{\mathbf{r}_{\mathrm{c}_1}, \mathbf{r}_{\mathrm{c}_2}, \ldots, \mathbf{r}_{\mathrm{c}_N}\} \subset \mathbb{R}^3, \tag{3}$$

where the task is to estimate the sound field based on the measurements $\mathbf{p}_\mathrm{m}$.

A complex-valued PINN is implemented to learn the unknown pressure field $p(\mathbf{r}_\mathrm{c}, k)$ at any control point from the sensor measurements $\mathbf{p}_\mathrm{m}$. Unlike a standard PINN, which might take the spatial coordinates $\mathbf{r}_\mathrm{c}$ as input, the proposed network also incorporates the microphone measurements associated with a given sound field. Specifically, for each generated sound field s, a vector of sensor data is defined as

$$\mathbf{p}_\mathrm{m}^{(s)} = \left[p^{(s)}(\mathbf{r}_\mathrm{m1}, k),\, p^{(s)}(\mathbf{r}_\mathrm{m2}, k), \ldots,\, p^{(s)}(\mathbf{r}_\mathrm{m64}, k)\right] \tag{4}$$

which captures the pressures recorded at the 64 microphone locations. Each entry reflects how the sound field computed from superpositions of plane waves interacts with the array geometry. The network is therefore designed to learn a mapping

$$\mathscr{N} : \left(\mathbf{r}_\mathrm{c}, \mathbf{p}_\mathrm{m}^{(s)}\right) \mapsto p^{(s)}(\mathbf{r}_\mathrm{c}, k), \tag{5}$$

where $\mathbf{r}_\mathrm{c} = (x_\mathrm{c}, y_\mathrm{c}, z_\mathrm{c}) \in \mathbb{R}^3$ is the location of the control point and $\mathscr{N}$ denotes the neural network model, viewed here as a nonlinear function that takes both the spatial coordinates and the microphone measurements as inputs and outputs the predicted pressure.

Including the microphone pressure vectors $\mathbf{p}_\mathrm{m}^{(s)}$ as part of the network input offers several important benefits. First, each sound field s generates a unique set of pressure measurements across the microphone array, capturing the spatial variations in phase and amplitude that depend on the direction of arrival of the individual plane waves and the constructive/destructive interferences. By conditioning the network on $\mathbf{p}_\mathrm{m}^{(s)}$, the model learns which sound field is being observed and can adapt its prediction of the pressure field accordingly. In contrast, if only the coordinate (x, y, z) were used, the network would be forced to learn a single field solution, failing to account for the differences in measurement patterns that arise from different superpositions of sound fields. Secondly, the pressure measurements naturally encapsulate valuable physical information relating to the interaction of the incident wave with the array, such as phase shifts and amplitude variations, which is essential for an accurate reconstruction of the sound field. Finally, this approach allows us to train a single network that can generalize across multiple sound fields. Thus, the inclusion of microphone pressures as part of the input not only preserves crucial physical context but also promotes efficiency and flexibility in handling sound fields consisting of superpositions of plane waves.

2.3 MSE and PDE control points

In the approach presented here, the interior control points Ω are split into two complementary sets (Ω_MSE, Ω_PDE) to make use of both direct supervision from known pressure values and physical constraints. The subset Ω_MSE contains only a small number Q of control points, where ground-truth pressure measurements are assumed to be available. At these locations, the mean squared error (MSE) is computed between the complex PINN's predicted pressure $\hat{p}^{(s)}(\mathbf{r}_\mathrm{c}, k)$ and the true pressure $p^{(s)}(\mathbf{r}_\mathrm{c}, k)$. This yields the loss term

$$\mathscr{L}_\mathrm{MSE} = \frac{1}{Q} \sum_{\mathbf{r}_{\mathrm{c}_q} \in Q} \left\| \hat{p}^{(s)}(\mathbf{r}_{\mathrm{c}_q}, k) - p^{(s)}\left(\mathbf{r}_{\mathrm{c}_q}, k\right) \right\|_2^2, \tag{6}$$

which compares the network solution to actual data at a number Q of critical points. Meanwhile, the larger set

Ω_{PDE} enforces the Helmholtz equation by minimizing the PDE residual adapted from Eq. (1) at L control points leading to the loss

$$\mathscr{L}_{\text{PDE}} = \frac{1}{L} \sum_{\mathbf{r}_{c_l} \in L} \left\| \frac{\nabla^2 \hat{p}(\mathbf{r}_{c_l}, k) + k^2 \hat{p}(\mathbf{r}_{c_l}, k)}{k^2} \right\|_2^2. \quad (7)$$

Note that the regular Helmholtz equation in Eq. (1) is normalized by k^2 and used as PDE loss. This operation results in keeping the Helmholtz residual on a comparable scale to the data loss, but it also diminishes the direct influence of the frequency on the training process. Especially at higher frequencies, the term k^2 grows quickly, resulting in an inflation of $\mathscr{L}_{\text{PDE}}$ relative to $\mathscr{L}_{\text{MSE}}$. Splitting the control points this way ensures that the network learns to match known pressures where they are available (through Ω_{MSE}) while also satisfying the fundamental wave equation over the entire domain (through Ω_{PDE}). The network is then trained using a weighted combination of the two losses, given by

$$\mathscr{L}_{\text{tot}} = \mathscr{L}_{\text{MSE}} + \alpha_{\text{PDE}} \mathscr{L}_{\text{PDE}}, \quad (8)$$

where α_{PDE} is a factor that weights the importance of the PDE loss.

2.4 Data generation

To form the training set, single frequency plane waves are first simulated arriving only in the horizontal plane, each having a random phase. All sound fields are simulated under free-field conditions. Fifty directions of arrival (DoAs) are defined around the open spherical microphone array of radius $R_{\text{mic}} = 0.4$ m, uniformly distributed in azimuth. From these 50 candidate plane waves, 5 distinct DoAs are randomly picked and their pressure fields are superposed, creating one combined sound field. By repeating this random selection process 10 times, 10 distinct superpositions are generated, each comprising 5 plane waves from different azimuthal directions. This superposition approach reflects more realistic acoustical scenarios, where multiple waves arrive simultaneously from various directions. This application is based on previous work [16], where a simple complex-valued MLP has been trained to estimate the sound field of single-frequency plane waves on a sphere of control points. Figure 1 shows an example of the plane waves grouped to create the superposed sound fields.

Due to the size of the open sphere and the number of microphones, the array has a known spatial aliasing frequency that can be calculated from the maximum spherical harmonic order of the array $N = 7$, the radius of the sphere $R_{\text{mic}} = 0.4$ m and the speed of sound $c = 343$ m/s as

$$f_{\text{alias}} = \frac{cN}{2\pi R_{\text{mic}}} \approx 955 \text{ Hz}. \quad (9)$$

In order to investigate the performance of the network at different frequencies, three different datasets are generated with plane wave frequencies $f_{\text{low}} = f_{\text{alias}}/2$, f_{alias} and $f_{\text{high}} = 2f_{\text{alias}}$. By generating these three different datasets, a variety of aliasing behaviors are captured and the proposed method can be investigated across different wave numbers. For each frequency, an analogous set of plane wave superpositions are generated, preserving the same DoA configurations and plane wave phases, while varying the wave number k.

The control points, at which the pressure field is to be reconstructed, are placed on the horizontal plane within a smaller sphere of radius 0.1 m. By limiting the control points to a two-dimensional cross-section, the spatial sampling is simplified while still capturing the critical features of the pressure field, making it sufficient for validating the reconstruction accuracy. To avoid spatial aliasing at these control points the spacing between neighboring points is set to $\frac{\lambda}{8}$, where λ is the wavelength corresponding to the highest simulated frequency (f_{high}). This ensures that the separation of control points is sufficiently fine to capture the pressure variations accurately. The resulting 61 control points are split into 5 MSE points (Ω_{MSE}), where target pressures are assumed to be known. These points correspond to the four diagonal (extreme) positions plus the center of the grid. The remaining 56 points (Ω_{PDE}) enforce the Helmholtz equation through the PDE loss. Thus, a small number of directly supervised points are combined with a larger set of physics-driven points to ensure the final pressure field solution is both data-consistent and physically accurate. Figure 2 illustrates the split between MSE and PDE control points.

For validation, a different set of control points is generated on the same horizontal plane but with a finer spacing of $\frac{\lambda}{16}$, resulting in 241 points. The

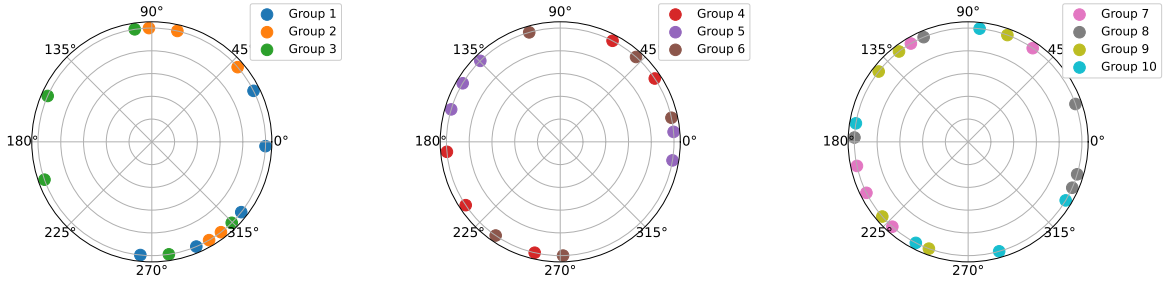


Fig. 1: Plane wave directions and their grouping into superposed sound fields

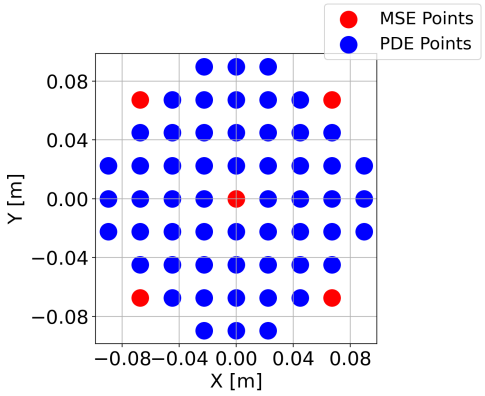


Fig. 2: Control points split into MSE and PDE points.

same superposed sound fields are kept during the validation process, but the networks performance is evaluated on this denser grid. This approach tests whether the model's predictions generalize well to new, more finely sampled positions inside the sphere.

Figure 3 shows an example of the sound field computed from the superposition of 5 plane waves and recorded by the microphone array on the open sphere, together with the sound field at the control points for training and validation.

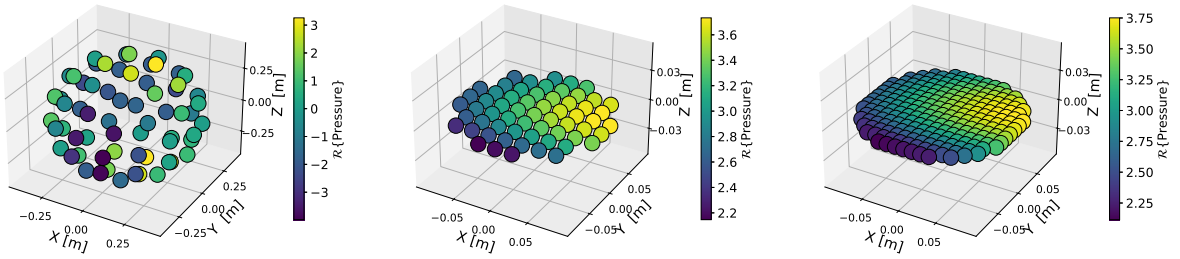
2.5 Network parameters

For each superposed sound field s, the complex pressure $\mathbf{p}_{\mathrm{m}}^{(s)}$ is computed at the 64 microphones. During training, these 64 complex-valued measurements are concatenated with the (x, y, z) coordinates of a selected control point $\mathbf{r}_{\mathrm{c}}$, also treated as a complex input by adding a zero imaginary part. Hence, each input vector is of length 67 (64 for the microphones plus 3 for the coordinates). The network then outputs a single complex value, corresponding to the predicted pressure at that control point $\mathbf{r}_{\mathrm{c}}$.

The proposed network architecture is a fully complex-valued multilayer-perceptron implemented using the complexPyTorch library [17]. Two hidden layers, each with 200 complex neurons were used and the activation function in each of them was the complex cardioid function [18] to effectively capture the phase-amplitude relationships in the data. The output layer contained a single complex neuron without an activation function to estimate the complex pressure at the control point.

For the MSE points (Ω_{MSE}), the $\mathscr{L}_{\mathrm{MSE}}$ was used during training, while for the PDE points (Ω_{PDE}), the normalised $\mathscr{L}_{\mathrm{PDE}}$ was used. Keeping a comparable scale between the two losses reduces the need for manual tuning of the loss weight λ_{PDE}, which is a well-known challenge when training PINNs [19]. An imbalance can lead to training difficulties, where one loss term may dominate, causing the network to prioritize either the data fit or the physical constraints inadequately.

During training, an initial learning rate of 0.01 was used and the model trained for 5000 epochs using a batch size of 64. A learning rate scheduler halved the rate every 1000 epochs, improving the convergence stability. The network weights were updated using the Adam optimizer. It is important to note that the same network architecture was trained individually from scratch for the three different datasets. All network models were trained on a MacBook Pro equipped with a 16-core CPU, 16-core Neural Engine and 36 GB of unified memory. Training each model took



(a) Open sphere **(b)** Training control points **(c)** Validation control points

Fig. 3: Example of sound field computed from a superposition of 5 plane waves at $f_{\text{low}} = 478$ Hz, corresponding to Group 2 (orange) in Figure 1.

Table 1: Comparison between the two network models in terms of averaged error expressed in dB.

	$f_{\text{low}} = 478$ Hz		$f_{\text{alias}} = 955$ Hz		$f_{\text{high}} = 1911$ Hz	
Control points	Train	Val	Train	Val	Train	Val
complexPINN	-14.6	-14.9	-5.1	-5.4	5.2	5.0
complexMLP	-9.9	-10.1	-2.9	-3.2	6.4	6.2

approximately 1 hour and 30 minutes. Data simulation and model training were performed using Python 3.11 and PyTorch 2.2.

3 Results

The objective is to evaluate the impact of incorporating physical laws into neural network training by comparing complex PINNs with standard complex neural networks that do not embed such physics-based constraints. To achieve this, two types of models are trained. The complexPINN models are trained based on the loss function presented in Eq. (8), with a weight $\alpha_{\text{PDE}} = 1$. The second type of models have $\alpha_{\text{PDE}} = 0$, effectively removing the physics constraint from the training process. Thus, the network learns only from the data-driven loss, which is based on the 5 control points Ω_{MSE} previously defined.

Table 1 shows the averaged error between the network predictions and the targets at both training and validation control points for the three different datasets corresponding to the different plane wave frequencies. The error is computed using Eq. (6). The error is first averaged over the control points for one sound field s and then averaged again over all 10 sound fields. The averaged errors from Table 1 show a couple of interesting observations for the models trained in this work. First, as expected, the sound field reconstruction accuracy at the control points is higher if the frequency is lower than the aliasing frequency. This is not surprising, since below the aliasing frequency, the spatial structure of the sound field can be extracted accurately from the microphone signals, while around and above f_{alias} the distance between the microphones is larger than the wavelength, which leads to an increase in sound field estimation error.

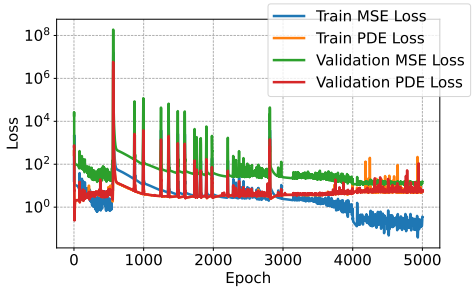
Incorporating the physical laws into the network model shows an improved sound field reconstruction. The complexPINN outperforms the complexMLP models at all investigated frequencies, for both training and validation control points. Interestingly, the reconstruction error at the validation control points is slightly lower than that at the training points. This may be attributed to several factors. Although both sets are drawn from the same grid, the training set is a sparser sampling (61 points) compared to the denser validation sampling (241 points). While this is a rather unusual way of splitting the training and validation control points, this was done intentionally to test the model's ability to generalize to a denser, unseen grid. This sparser selection may disproportionately include points in regions where the model's prediction error is higher, making the average error appear larger. In contrast, the denser validation set provides a more comprehensive sampling of the pressure field, so the averaging process smooths out localized high-error regions and results in a lower overall error.

Figure 4 shows the training and validation losses as

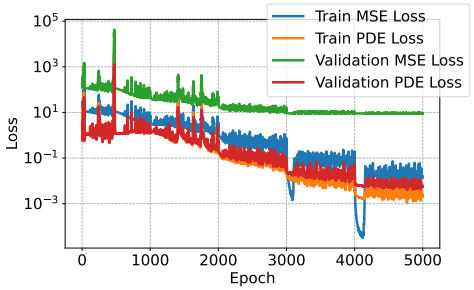
training progresses for both complexPINN and complexMLP models. The case shown here involves sound fields computed from plane waves at $f_{\text{alias}} = 955$ Hz. The plots display the following: the difference between predictions and targets at Ω_{MSE} (Train MSE Loss), the Helmholtz residual for the training control points Ω_{PDE} (Train PDE Loss), the difference between predictions and targets at validation control points (Validation MSE Loss) and Helmholtz residual at the validation points (Validation PDE Loss). First, it can be seen that adding the Helmholtz residual as regularization parameter in the model results in smoother training curves. By forcing the model to satisfy the Helmholtz equation during training, the loss landscape becomes less erratic, leading to more stable gradients and a more gradual reduction in the loss over epochs. This additional constraint effectively reduces the degrees of freedom of the solution.

Not surprisingly, if $\alpha_{\text{PDE}} = 0$, the model does not learn to satisfy the Helmholtz equation and therefore the PDE losses at training and validation control points (orange and red curves) remain constant during training. When $\alpha_{\text{PDE}} = 1$, the complexPINN model is learning to satisfy the Helmholtz equation as evidenced by the decreasing PDE losses over time. Interestingly, the Helmholtz residual at the training points behaves similarly to that at the validation points, even though the model does not see these control points during training. This can be explained by the fact that the validation control points are sampled from the same grid as the training points, just using a finer grid resolution. Because the solutions to the Helmholtz equation change gradually over small distances, once the model learns to satisfy the equation at one point, it is likely to achieve similar accuracy in neighboring regions. Toward the end of training, the PDE loss on the training points becomes slightly lower than on the validation points, which is expected because the model directly optimizes the physics constraint at the training locations. This result emphasises the model's ability to generalize the learned physics to unseen regions, even though it cannot perfectly satisfy the PDE at every point.

An interesting observation in Figure 4 is that, even though the complexPINN model significantly reduces the Helmholtz residual compared to the complexMLP, the overall error between predictions and targets at validation control points shows only a small improvement. This is not necessarily because the model prioritizes fitting the Ω_{PDE} data, but rather, it arises from the underdetermined nature of the Helmholtz equation in the absence of boundary conditions. With a 1:1 weighting between the data loss and the PDE loss, both are considered equally. However, since the Helmholtz equation admits an infinite family of solutions that locally satisfy the PDE, the network's single-neuron output for pressure may converge to a physically valid solution that minimizes the PDE residual without necessarily matching the desired target pressure. A possible solution to improve the performance would be to reduce α_{PDE} so that the model is encouraged to match the target pressure values. However, this comes with a trade-off, because the model might find a solution that is less physically consistent. Another approach would be to incorporate additional constraints during training.



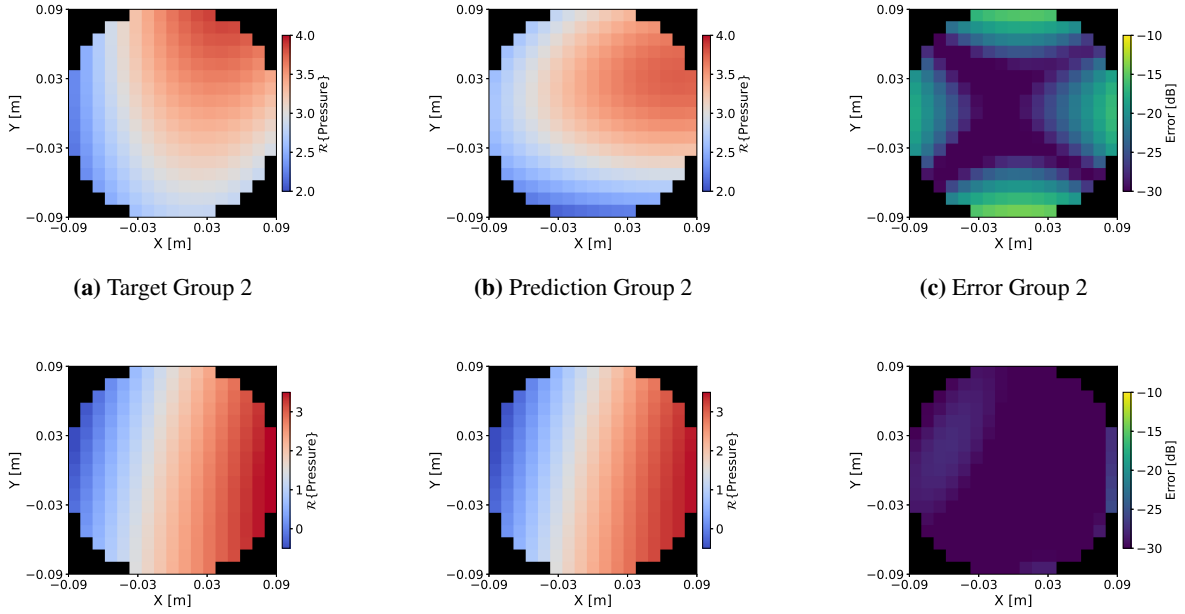
(a) complexMLP



(b) complexPINN

Fig. 4: Illustration of losses as training progresses for the complexMLP and complexPINN models for the dataset with plane waves at $f_{\text{alias}} = 955$ Hz.

It is also interesting to investigate how the network models perform for each individual sound field. Figure 7 shows the averaged error over training or

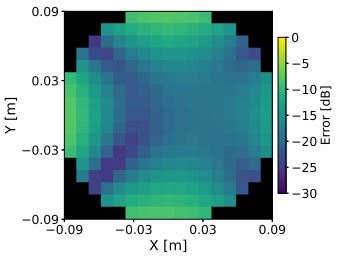


(a) Target Group 2 **(b)** Prediction Group 2 **(c)** Error Group 2

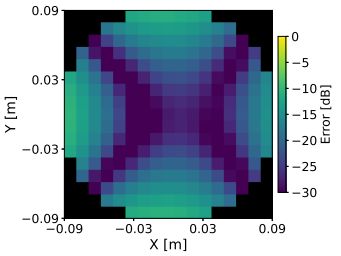
(d) Target Group 5 **(e)** Prediction Group 5 **(f)** Error Group 5

Fig. 5: Illustration of target, prediction and sound field error for two different groups of superposed single-frequency plane waves at $f_{\text{low}} = 478$ Hz. The plots are shown for the complexPINN model with $\alpha_{\text{PDE}} = 1$.

validation points for each of the 10 sound fields computed from the superposition of plane waves at frequency f_{low}. It can be observed that the performance of the network models varies across different superposed sound fields. This variability is not necessarily surprising and can be explained by the complexity of each sound field. When multiple plane waves are superposed, their relative directions and phase relationships produce interference patterns that can vary significantly in their complexity. As a result, the difficulty of accurately reconstructing the pressure field depends on the specific configuration of the underlying plane waves. In fields where the interference patterns are more complex, the network faces a greater challenge, leading to higher reconstruction errors. For example, the complex PINN achieves the lowest reconstruction error for the 5-th sound field, which corresponds to Group 5 in Figure 1. The superposed sound field in Group 5 is computed from adding the plane waves corresponding to the purple circles. Two of them are close to each other at 0° while the other three are also very close to each other between 135° and 180°. The superposition of these plane waves together with their assigned random phases creates a sound field that is less complex, compared for example to the sound field generated by the plane waves from Group 2. This can be seen in Figure 5, where the target and the estimated sound fields are shown for the validation control points as 2D plots together with the error distributed across the control points. The performance is shown for the complexPINN model with $\alpha_{\text{PDE}} = 1$. It can be observed that if the generated sound field is more complex due to the different plane wave directions and different phases (Group 2), the network achieves a lower reconstruction accuracy. Interestingly, in the error plot for Group 2, a pattern emerges: the reconstruction error is smaller along the two diagonals of the horizontal circle. This pattern likely corresponds to the locations of the Ω_{MSE} shown in Figure 2, where the network is directly supervised. In the complexPINN model, the Helmholtz equation helps propagate this low-error information to neighboring points, effectively broadening the region of accurate predictions. In contrast, the standard MLP, which relies solely on point-wise MSE, only achieves a narrow band of low error around these training points. This is shown in Figure 6 as a comparison of error plots between complexMLP and complexPINN for the



(a) complexMLP



(b) complexPINN

Fig. 6: Comparison of error plots between targets and estimations at the validation control points for the complexMLP and complexPINN models for the sound field generated in Group 10 with plane waves at $f_{\text{alias}} = 955$ Hz.

soundfield generated in Group 10 with plane waves at f_{alias}.

4 Conclusion

The work presented here demonstrates that incorporating physics-based constraints via complex-valued PINNs enhances the physical consistency of sound field reconstructions and improves the model's ability to estimate the sound field at control points, where it is assumed that no training data is available. By integrating measured pressure data as an additional input, the network can learn to estimate different sound fields and the reconstruction performance depends on the complexity of the sound field. Future work will focus on evaluating the network's ability to generalize to entirely new sound fields that were not included in the training set, as well as its performance when predicting pressure at control points sampled on grids that differ from those used during training. Additionally, the influence of the number and spatial distribution of the MSE training points on the model's accuracy could be investigated

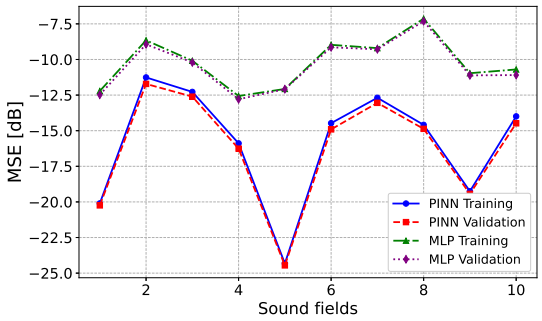


Fig. 7: Averaged reconstruction error across the different sound fields used during training for plane waves at frequency $f_{\text{low}} = 478$ Hz.

further. These studies will help refine the approach and further enhance its robustness for practical sound field reconstruction applications.

References

[1] Verburg, S. A. and Fernandez-Grande, E., "Reconstruction of the sound field in a room using compressive sensing," *The Journal of the Acoustical Society of America*, 143(6), pp. 3770–3779, 2018, ISSN 0001-4966, doi:10.1121/1.5042247.

[2] Fernandez-Grande, E., "Sound field reconstruction using a spherical microphone array)," *The Journal of the Acoustical Society of America*, 139(3), pp. 1168–1178, 2016, ISSN 0001-4966, doi:10.1121/1.4943545.

[3] Zotter, F. and Frank, M., *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer Topics in Signal Processing, Springer Cham, 1 edition, 2019, doi:10.1007/978-3-030-17207-7.

[4] Nelson, P. A., "Active control of acoustic fields and the reproduction of sound," *Journal of Sound and Vibration*, 177(4), pp. 447–477, 1994.

[5] Miotello, F., Comanducci, L., Pezzoli, M., Bernardini, A., Antonacci, F., and Sarti, A., "Reconstruction of Sound Field Through Diffusion Models," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1476–1480, 2024, doi:10.1109/ICASSP48485.2024.10446761.

[6] Fernandez-Grande, E., Karakonstantis, X., Caviedes-Nozal, D., and Gerstoft, P., "Generative models for sound field reconstruction," *The Journal of the Acoustical Society of America*, 153(2), pp. 1179–1190, 2023, ISSN 0001-4966, doi:10.1121/10.0016896.

[7] Koyama, S., Ribeiro, J. G. C., Nakamura, T., Ueno, N., and Pezzoli, M., "Physics-Informed Machine Learning for Sound Field Estimation: Fundamentals, state of the art, and challenges," *IEEE Signal Processing Magazine*, 41(6), pp. 60–71, 2024, doi:10.1109/MSP.2024.3465896.

[8] Ma, F., Abhayapala, T. D., and Samarasinghe, P. N., "Circumvent spherical Bessel function nulls for open sphere microphone arrays with physics informed neural network," in *10th Convention of the European Acoustics Association (Forum Acusticum 2023)*, Turin, Italy, 2023, doi:10.61782/fa.2023.0101.

[9] Olivieri, M., Karakonstantis, X., Pezzoli, M., Antonacci, F., Sarti, A., and Fernandez-Grande, E., "Physics-informed neural network for volumetric sound field reconstruction of speech signals," *EURASIP Journal on Audio, Speech, and Music Processing*, 2024(1), p. 42, 2024.

[10] Luan, X., Olivieri, M., Pezzoli, M., Antonacci, F., and Sarti, A., "Complex-Valued Physics-Informed Neural Network for Near-Field Acoustic Holography," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, pp. 126–130, IEEE, 2024.

[11] Paul, V. S. and Nelson, P. A., "Efficient design of complex-valued neural networks with application to the classification of transient acoustic signals," *The Journal of the Acoustical Society of America*, 156(2), pp. 1099–1110, 2024, ISSN 0001-4966, doi:10.1121/10.0028230.

[12] Hirose, A., "Complex-valued neural networks: The merits and their origins," in *Proceedins of the International Joint Conference on Neural Networks*, pp. 1237–1244, 2009, doi:10.1109/IJCNN.2009.5178754.

[13] Scardapane, S., Van Vaerenbergh, S., Hussain, A., and Uncini, A., "Complex-Valued Neural Networks With Nonparametric Activation Functions," *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(2), pp. 140–150, 2020, ISSN 2471-285X, doi:10.1109/TETCI.2018.2872600.

[14] Kovacs, A., Exl, L., Kornell, A., Fischbacher, J., Hovorka, M., Gusenbauer, M., Breth, L., Oezelt, H., Yano, M., Sakuma, N., Kinoshita, A., Shoji, T., Kato, A., and Schrefl, T., "Conditional physics informed neural networks," *Communications in Nonlinear Science and Numerical Simulation*, 104, p. 106041, 2022, ISSN 1007-5704, doi:https://doi.org/10.1016/j.cnsns.2021.106041.

[15] mh acoustics, *em64 Eigenmike release notes*, 25 Summit Ave Summit, NJ 07901, 2023.

[16] Paul, V. S., Hahn, N., and Nelson, P. A., "Sound Field Analysis Using a Complex-Valued Neural Network," in *Proceedings of the Institute of Acoustics, Vol. 46, Pt. 3*, 2024.

[17] Matthès, M. W., Bromberg, Y., de Rosny, J., and Popoff, S. M., "Learning and Avoiding Disorder in Multimode Fibers," *Phys. Rev. X*, 11, p. 021060, 2021, doi:10.1103/PhysRevX.11.021060.

[18] Virtue, P., Yu, S. X., and Lustig, M., "Better than real: Complex-valued neural nets for MRI fingerprinting," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 3953–3957, 2017, doi:10.1109/ICIP.2017.8297024.

[19] Hua, J., Li, Y., Liu, C., Wan, P., and Liu, X., "Physics-informed neural networks with weighted losses by uncertainty evaluation for accurate and stable prediction of manufacturing systems," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.