# Sound field reconstruction using a compact acoustics-informed neural network

Fei Ma ⓘ ; Sipei Zhao ⓘ ; Ian S. Burnett

🔴 Check for updates

🌐
View
Online

↪
Export
Citation

09 January 2025 04:10:11

# Sound field reconstruction using a compact acoustics-informed neural network

Fei Ma,[1] (iD) Sipei Zhao,[1,a)] (iD) and Ian S. Burnett[2]

[1]*Center for Audio, Acoustics and Vibration, Faculty of Engineering and IT, University of Sydney Technology, Ultimo, New South Wales 2007, Australia*

[2]*DVC and VP STEM, RMIT University, Melbourne, Victoria 3000, Australia*

**ABSTRACT:**

Sound field reconstruction (SFR) augments the information of a sound field captured by a microphone array. Using basis function decomposition, conventional SFR methods are straightforward and computationally efficient but may require more microphones than needed to measure the sound field. Recent studies show that pure data-driven and learning-based methods are promising in some SFR tasks, but they are usually computationally heavy and may fail to reconstruct a physically valid sound field. This paper proposes a compact acoustics-informed neural network (AINN) method for SFR, whereby the Helmholtz equation is exploited to regularize the neural network. As opposed to pure data-driven approaches that solely rely on measured sound pressures, the integration of the Helmholtz equation improves robustness of the neural network against variations during the measurement processes and prompts the generation of physically valid reconstructions. The AINN is designed to be compact and able to predict not only the sound pressures but also sound pressure gradients within a spatial region of interest based on measured sound pressures along the boundary. Experiments with acoustic transfer functions measured in different environments demonstrate the superiority of the AINN method over the traditional cylindrical harmonics and singular value decomposition methods. © *2024 Acoustical Society of America.* https://doi.org/10.1121/10.0029022

## I. INTRODUCTION

Microphone arrays are commonly used to measure a sound field and maximize the information about the source (Benesty *et al.*, 2008). A large-aperture array with densely spaced microphones is preferred for sound field measurements; however, this is not always possible because of practical considerations such as cost and microphone arrangement (Rafaely, 2015). This necessitates sound field reconstruction (SFR; Fernandez-Grande, 2016; Williams, 1999; Zhang *et al.*, 2008), a task that aims to reconstruct a sound field apart from the limited (sparse) measurement locations.

Existing SFR methods can be broadly classified into two categories: conventional methods based on basis function decomposition and recent learning-based methods. The conventional methods decompose measured sound fields into some basis functions, such as cylindrical harmonics (CHs) (Williams, 1999), spherical harmonics (Chen *et al.*, 2015; Tang *et al.*, 2022; Verburg and Fernandez-Grande, 2018; Wabnitz *et al.*, 2011; Williams, 1999), prolate spheroidal wave functions (Zhang *et al.*, 2023), and plane waves (Antonello *et al.*, 2017; Fernandez-Grande, 2016; Schmid *et al.*, 2021; Williams, 1999). The basis functions are solutions of the Helmholtz equation (Skudrzyk, 2012; Williams, 1999), the governing partial differential equation (PDE) of time-harmonic wave propagation, and are continuous spatial

functions that can be evaluated at arbitrary positions. These two factors make the conventional methods easy to compute and generate a physically valid reconstruction of the sound field away from the measurement positions. However, the basis functions are designed with respect to some coordinate systems (Williams, 1999) without considering the statistical characteristics of sound fields. Thus, conventional methods may require more than the necessary number of measurements (spatial sampling points) to determine the basis function weights to reconstruct a sound field. The reality is that the statistical characteristics of a sound field can be used to reduce the number of sampling points based on singular value decomposition (SVD; Zhu *et al.*, 2020, 2021), compressive sensing (Verburg and Fernandez-Grande, 2018; Wabnitz *et al.*, 2011), statistical learning (Hahmann *et al.*, 2021), or Bayesian inference (Schmid *et al.*, 2021).

In contrast to conventional methods, recent learning-based techniques do not rely on predesigned basis functions. Instead, they exploit the learned statistical characteristics of sound fields for SFR. Lluis *et al.* (2020) and Kristoffersen *et al.* (2021) developed U-net-like neural networks, which were trained with simulated or measured room impulse responses (RIRs). The U-net-like neural networks achieved superior SFR performance over some of the conventional methods in the low frequency range (<300 Hz). Hahmann *et al.* (2021) proposed to learn basis functions in local subdomains that are subsequently generalized across different rooms and frequencies, which showed potential for modeling

a)Electronic mail: Sipei.Zhao@uts.edu.au

complex sound fields according to their local (spatial) or statistical characteristics. By further enforcing self-similarity between adjacent local subdomains (Hahmann and Fernandez-Grande, 2022), the method attained better SFR performance when few measurements were available. Most recently, Fernandez-Grande et al. (2023) proposed to use the generative adversarial networks for SFR to recover some of the sound field energy at high frequencies that would otherwise be lost due to under-sampling, demonstrating the promise of using statistical learning methods to overcome the sampling limitations. Although the learning-based methods outperformed the conventional methods in some SFR tasks, their computations are time-consuming. Furthermore, they are purely data-driven and, thus, do not necessarily reconstruct physically valid sound fields (Fernandez-Grande et al., 2023).

Recently, physical laws have been integrated into neural networks for various acoustic studies, such as the Kirchhoff–Helmholtz-based convolutional neural network (CNN) for nearfield acoustic holography (Olivieri et al., 2021), the physics-informed convolutional neural network (PI-CNN) for sound field estimation (Shigemi et al., 2022), the physics-informed neural network (PINN) for RIR reconstruction (Karakonstantis et al., 2024; Pezzoli et al., 2023), the PINN for acoustic boundary admittance estimation (Schmid et al., 2024), the DeepONet-based method for sound propagation simulations with moving sources (Borrel-Jensen et al., 2024), and the Fourier neural network-based method for estimating the sound field due to a source whose position is unknown (Middleton et al., 2023). Most of these studies (Karakonstantis et al., 2024; Olivieri et al., 2021; Pezzoli et al., 2023; Shigemi et al., 2022) attempted to reconstruct the sound field within a region of interest (ROI) with a few measurements inside the ROI.

In an alternative approach, this paper proposes an acoustics-informed neural network (AINN) to reconstruct the sound field within the ROI based on the sound pressures measured on its boundary (Cuomo et al., 2022; Raissi et al., 2019). The AINN is designed to approximate the sound field at the measurement positions and guided by the Helmholtz equation to generate physically valid reconstructions away from the measurement positions. The AINN is designed in the frequency domain and its size, i.e., the number of neurons in the hidden layers, is determined based on the physical principle. The AINN is compact and lightweight, making it easier to train than large neural networks. In addition, owing to the automatic-differentiation of deep-learning libraries (Pezzoli et al., 2023), the AINN is able to reconstruct not only sound pressures but also their gradients within the ROI. Experiments with transfer functions measured with two microphone arrays in three different rooms (Zhao et al., 2022) are conducted to compare the proposed method with the conventional CH (Williams, 1999) and SVD (Zhu et al., 2021) methods. Experimental results demonstrate the superiority of the proposed AINN method over the CH and SVD methods.

The remainder of this paper is organized as follows. The problem is formulated in Sec. II. The CH and SVD methods are reviewed in Sec. III, followed by the proposed AINN method. Numerical experiments are presented in Sec. IV to validate the performance of the proposed AINN method in comparison to the CH and the SVD methods. Section V concludes this work.

## II. PROBLEM FORMULATION

The problem of interest is illustrated in Fig. 1, where $(x,y)$ and $(r, \phi)$ denote the Cartesian and polar coordinates with respect to the coordinate origin $O$, respectively. The stars denote sound sources that generate a sound field in the ROI, which is depicted as the gray area. An array of microphones on the boundary of the ROI, displayed as the dots in Fig. 1, measure the sound pressures at $\{x_q, y_q\}_{q=1}^{Q}$ (or $\{r_q, \phi_q\}_{q=1}^{Q}$) as $P(\omega, x_q, y_q)_{q=1}^{Q}$ [or equivalently $P(\omega, r_q, \phi_q)_{q=1}^{Q}$], where $Q$ is the number of measurement points, and $\omega = 2\pi f$ is the angular frequency, where $f$ is the frequency. The objective is to estimate the sound pressures and their gradients inside the ROI based on the measured sound pressures on the boundary. Hereafter, the symbol $\omega$ is omitted in some quantities for notational simplicity.

It is assumed that there is no sound source inside the ROI, hence, the sound field within the ROI is governed by the homogeneous Helmholtz equation (Williams, 1999), i.e.,

$$P + \frac{1}{(\omega/c)^2} \nabla^2 P = 0, \tag{1}$$

where $c = 340 \, \text{m/s}$ is the speed of sound in air at room temperature, and $\nabla^2$ denotes the Laplacian operator. In Cartesian coordinates, the Laplacian operator is given by (Williams, 1999)

$$\nabla^2 = \frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2}, \tag{2}$$

and in polar coordinates, it is given by
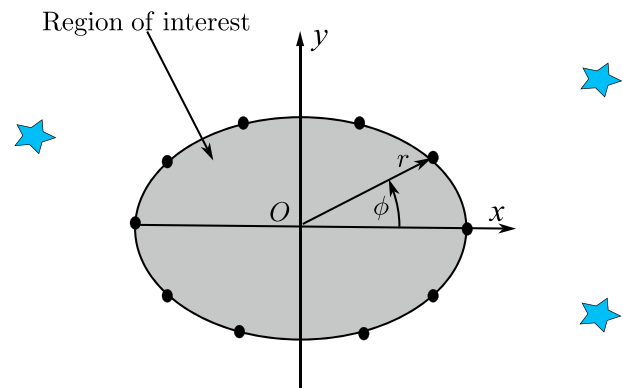


FIG. 1. (Color online) Illustration of problem formulation. A number of sound sources (⋆) generate a sound field within a ROI, and the objective is to estimate the sound pressures and their gradients within the ROI based on the sound pressures measured by microphones (•) placed along the ROI boundary. (a) cAINN, single coupled network and (b) dAINN, two decoupled networks, are shown.

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}. \tag{3}$$

In this paper, we build a compact neural network informed by the Helmholtz equation [Eq. (1)] to reconstruct the sound field within the ROI based on the microphone measurements.

## III. METHODOLOGY

This section first reviews the CH method (Williams, 1999) and SVD method (Zhu et al., 2021) for SFR, and subsequently proposes the AINN method.

### A. The CHs method

Acoustic quantities are expressed in polar coordinates for the ease of computing the CHs. Sound pressures can be decomposed into CHs as (Williams, 1999)

$$\mathbf{P}_M = \mathbf{J}\mathbf{A}, \tag{4}$$

where $\mathbf{P}_M = [P(r_1,\phi_1), P(r_2,\phi_2),\ldots,P(r_Q,\phi_Q)]^\top$ denote the measured sound pressures at $\{r_q,\phi_q\}_{q=1}^Q$, $(\cdot)^\top$ is the transpose operation, $\mathbf{A} = [A_{-N}, A_{-N+1},\ldots,A_N]^\top$ denote the CH weights, and

$$\mathbf{J} = \begin{bmatrix} J_{-N}\left(\frac{\omega}{c}r_1\right)e^{iN\phi_1} & \cdots & J_N\left(\frac{\omega}{c}r_1\right)e^{-iN\phi_1} \\ J_{-N}\left(\frac{\omega}{c}r_2\right)e^{iN\phi_2} & \cdots & J_N\left(\frac{\omega}{c}r_2\right)e^{-iN\phi_2} \\ \vdots & \vdots & \vdots \\ J_{-N}\left(\frac{\omega}{c}r_Q\right)e^{iN\phi_Q} & \cdots & J_N\left(\frac{\omega}{c}r_Q\right)e^{-iN\phi_Q} \end{bmatrix} \tag{5}$$

is a $Q \times (2N+1)$ matrix whose entry $J_n((\omega/c)r_q)e^{-in\phi_q}$ is the $n$th-order CH (Williams, 1999) evaluated at $(r_q,\phi_q)$, $i$ is the imaginary unit, and $J_n(\cdot)$ is the Bessel function of order $n$ (Williams, 1999). In Eqs. (4) and (5), $N$ is the dimensionality of the sound field under CH decomposition and normally chosen as (Kennedy et al., 2007)

$$N = \lceil 2\pi fr/c \rceil, \tag{6}$$

where $\lceil \cdot \rceil$ is the ceiling operation.

The CH method estimates the weights $\hat{\mathbf{A}} = [\hat{A}_{-N}, \hat{A}_{-N+1},\ldots,\hat{A}_N]^\top$ through

$$\hat{\mathbf{A}} = \mathbf{J}^\dagger \mathbf{P}_M, \tag{7}$$

where $(\cdot)^\dagger$ denotes the pseudo-inverse operation. The sound pressure and radial gradient for an arbitrary position $(r_e,\phi_e)$ can be reconstructed as (Williams, 1999)

$$\hat{P}_{CH}(r_e,\phi_e) \approx \sum_{n=-N}^N \hat{A}_n J_n\left(\frac{\omega}{c}r_e\right)e^{-in\phi_e}, \tag{8}$$

and

$$\frac{\partial \hat{P}_{CH}(r_e,\phi_e)}{\partial r_e} \approx \frac{\omega}{c}\sum_{n=-N}^N \hat{A}_n J'_n\left(\frac{\omega}{c}r_e\right)e^{-in\phi_e}, \tag{9}$$

where $J'_n(\cdot)$ denotes the derivative of the Bessel function with respect to the argument. The pressure gradients along the $x$ axis and $y$ axis at $(x_e,y_e)$ can be reconstructed as

$$\frac{\partial \hat{P}_{CH}(x_e,y_e)}{\partial x_e} = \frac{\partial \hat{P}_{CH}(r_e,\phi_e)}{\partial r_e}\cos(\phi_e), \tag{10}$$

$$\frac{\partial \hat{P}_{CH}(x_e,y_e)}{\partial y_e} = \frac{\partial \hat{P}_{CH}(r_e,\phi_e)}{\partial r_e}\sin(\phi_e). \tag{11}$$

### B. The SVD method

The SVD method (Zhu et al., 2020, 2021) regards a source at $(x_s,y_s)$ as a cluster of virtual point sources whose positions are $\{x_{s,j},y_{s,j}\}_{j=1}^J$ and constructs two matrices with respect to the virtual point sources. The first is a matrix of transfer functions between the virtual point sources and microphones such that

$$\mathbf{H}_{SM} = \begin{bmatrix} H(x_1,y_1,x_{s,1},y_{s,1}) & \cdots & H(x_1,y_1,x_{s,J},y_{s,J}) \\ \vdots & \vdots & \vdots \\ H(x_Q,y_Q,x_{s,1},y_{s,1}) & \cdots & H(x_Q,y_Q,x_{s,J},y_{s,J}) \end{bmatrix}, \tag{12}$$

where the $q$th row and $j$th column entry, $H(x_q,y_q,x_{s,j},y_{s,j})$, is the free-field transfer function (Williams, 1999) between the virtual point source located at $(x_{s,j},y_{s,j})$ and the microphone located at $(x_q,y_q)$, i.e. (Williams, 1999),

$$H(x_v,y_v,x_{s,j},y_{s,j}) = \frac{1}{4\pi}\frac{\exp(-i\omega r_{v,j}/c)}{r_{v,j}}, \tag{13}$$

where $r_{v,j} = \sqrt{(x_v - x_{s,j})^2 + (y_v - y_{s,j})^2}$. The second is a matrix of transfer functions between the virtual point sources and pressure estimation points, namely,

$$\mathbf{H}_{SV} = \begin{bmatrix} H(x_1,y_1,x_{s,1},y_{s,1}) & \cdots & H(x_1,y_1,x_{s,J},y_{s,J}) \\ \vdots & \vdots & \vdots \\ H(x_V,y_V,x_{s,1},y_{s,1}) & \cdots & H(x_V,y_V,x_{s,J},y_{s,J}) \end{bmatrix}, \tag{14}$$

where the $v$th row and $j$th column entry, $H(x_v,y_v,x_{s,j},y_{s,j})$, is the free-space transfer function between the virtual point source located at $(x_{s,j},y_{s,j})$ and the pressure estimation point located at $(x_v,y_v)$, which is defined similarly as in Eq. (13). The two matrices are decomposed as (Zhu et al., 2021)

$$\mathbf{H}_{SM} = \mathbf{U}_{SM}\boldsymbol{\Sigma}_{SM}\mathbf{V}_{SM}^*, \tag{15}$$

and

$$\mathbf{H}_{SV} = \mathbf{U}_{SV}\boldsymbol{\Sigma}_{SV}\mathbf{V}_{SV}^*, \tag{16}$$

where $(\cdot)^*$ denotes the complex conjugate operation, $\mathbf{U}_{SM}$ and $\mathbf{U}_{SV}$ are unitary matrices whose columns represent the

basis functions of the receiver space, $\mathbf{V}_{\text{SM}}$ and $\mathbf{V}_{\text{SV}}$ are unitary matrices whose columns represent the basis functions of the source space, and $\boldsymbol{\Sigma}_{\text{SM}}$ and $\boldsymbol{\Sigma}_{\text{SV}}$ are diagonal matrices whose elements represent the capability of a source space basis function to excite a receiver space basis function. The dimensions of $U_{\text{SM}}$, $\Sigma_{\text{SM}}$, $V_{\text{SM}}$, $U_{\text{SV}}$, $\Sigma_{\text{SV}}$, and $V_{\text{SV}}$ are $Q \times Q$, $Q \times J$, $J \times J$, $V \times V$, $V \times J$, and $J \times J$, respectively.

The SVD method (Zhu *et al.*, 2021) reconstructs the sound pressure $\hat{\mathbf{P}}_{\text{V}} = [\hat{P}(x_1, y_1), ..., \hat{P}(x_v, y_v), ..., \hat{P}(x_V, y_V)]^{\top}$ as

$$\hat{\mathbf{P}}_{\text{V}} = \mathbf{U}_{\text{SV}}\boldsymbol{\Sigma}_{\text{SV}}\mathbf{V}_{\text{SV}}^{*}\mathbf{V}_{\text{SM}}\boldsymbol{\Sigma}_{\text{SM}}^{-1}\mathbf{U}_{\text{SV}}^{*}\mathbf{P}_{\text{M}}, \tag{17}$$

where $(\cdot)^{-1}$ denotes the matrix inversion operation. To reconstruct the pressure gradient, the SVD method first reconstructs the pressure at two closely spaced positions $(x_v \pm \delta_x, y_v \pm \delta_y)$, and then approximates the pressure gradient by

$$\frac{\partial \hat{P}_{\text{SVD}}(x_v, y_v)}{\partial x_v} \approx \frac{\hat{P}(x_v + \delta_x, y_v) - \hat{P}(x_v - \delta_x, y_v)}{2\delta_x}, \tag{18}$$

$$\frac{\partial \hat{P}_{\text{SVD}}(x_v, y_v)}{\partial y_v} \approx \frac{\hat{P}(x_v, y_v + \delta_y) - \hat{P}(x_v, y_v - \delta_y)}{2\delta_y}. \tag{19}$$

The radial gradient can be reconstructed as

$$\frac{\partial \hat{P}_{\text{SVD}(r_e, \phi_e)}}{\partial r_e} = \frac{\partial \hat{P}_{\text{SVD}}(x_e, y_e)}{\partial x_e}\cos(\phi_e)$$
$$+ \frac{\partial \hat{P}_{\text{SVD}}(x_e, y_e)}{\partial y_e}\sin(\phi_e). \tag{20}$$

## C. The AINN

This section proposes an AINN method for SFR. Two designs of the AINN, depending on whether the real and imaginary parts of the sound pressure are modeled separately or collaboratively, are investigated in this paper. In the first design,

the real and imaginary parts of the sound pressure are modeled with a single network and a single loss function, as shown in Fig. 2(a). In the second design, by contrast, the real and imaginary parts of the sound pressure are modeled separately with two small networks with individual loss functions, as illustrated in Fig. 2(b). Hereinafter, the two designs are referred to as the coupled acoustics-informed neural network (cAINN) and decoupled acoustics-informed neural network (dAINN).

To model the real and imaginary parts of the sound pressure, one network design, cAINN, should have more expressive power and, hence, more hidden layers or more neurons in each hidden layer (Goodfellow *et al.*, 2016). As shown in Figs. 2(a) and 2(b), more neuron connections between the two hidden layers make the cAINN more complicated than the dAINN. If, in Fig. 2(a), the $N$ neurons in the upper half network are disconnected from the other $N$ neurons in the lower half network, the cAINN will be the same as the simpler dAINN when two losses are used to train the network separately. Although in theory, cAINN is capable of exploiting the real and imaginary parts of the sound pressure for the training process and, therefore, may achieve better SFR performance, its training is complicated but may not attain the desired performance in practice.

As depicted in Fig. 2, for cAINN and dAINN, the inputs are the Cartesian coordinates $(x,y)$ and the outputs are the real and imaginary parts of the reconstructed sound pressure, denoted as $\mathcal{N}^{\Re}(x,y)$ and $\mathcal{N}^{\Im}(x,y)$, respectively. Similar to the conventional data-driven methods, a data loss is used to minimize the differences between the reconstructed and measured sound pressures at the measurement locations. The data loss for the real parts of the sound pressures $\mathfrak{L}_{\text{d}}^{\Re}$ is given by

$$\mathcal{L}_{\text{d}}^{\Re} = \frac{1}{Q}\sum_{q=1}^{Q}(P^{\Re}(x_q, y_q) - \mathcal{N}^{\Re}(x_q, y_q))^2, \tag{21}$$

where $P^{\Re}(x_q, y_q)$ and $\mathcal{N}^{\Re}(x_q, y_q)$ denote the real parts of the measured and reconstructed sound pressure at $\{x_q, y_q\}_{q=1}^{Q}$, respectively.
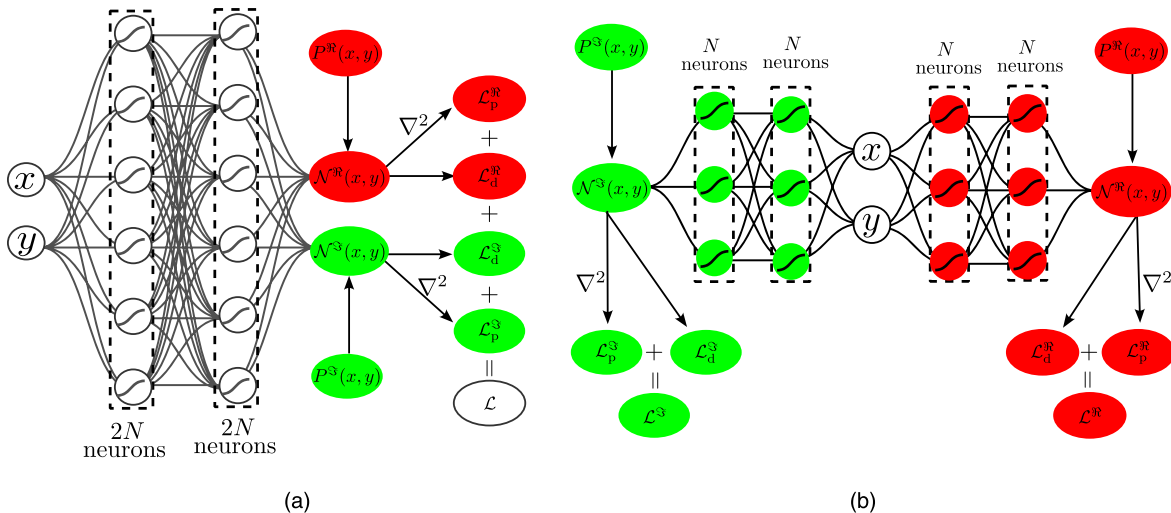


FIG. 2. (Color online) (a) The single coupled network (cAINN) and (b) the two decoupled network (dAINN) are displayed.

To incorporate acoustic information into the design of the neural network, the Helmholtz Eq. (1) is used to calculate an extra PDE loss. It is noted that, unlike from the data loss that is calculated only for the measured locations, the PDE loss is calculated for both of the measurement locations on the ROI boundary and multiple spatially sampled points within the ROI. By uniformly sampling $D$ positions within the ROI at $\{x_d, y_d\}_{d=1}^D$ and referring to Eqs. (1) and (2), the PDE loss for the real part of the sound pressure $\mathfrak{L}_p^\Re$ is given by

$$
\mathcal{L}_p^\Re = \frac{1}{D} \sum_{d=1}^D \left( \mathcal{N}^\Re(x_d, y_d) + \frac{1}{(w/c)^2} \right.
$$
$$
\left. \times \left[ \frac{\partial^2 \mathcal{N}^\Re(x_d, y_d)}{\partial x_d^2} + \frac{\partial^2 \mathcal{N}^\Re(x_d, y_d)}{\partial y_d^2} \right] \right)^2, \quad (22)
$$

where $\mathcal{N}^\Re(x_d, y_d)$ denotes the real part of the reconstructed sound pressure at $\{x_d, y_d\}$. The number $D$ should be chosen to be sufficiently large such that the distance between two PDE loss calculation points is no more than one-tenth of the wavelength of interest. The data loss and PDE loss are combined to calculate the total loss. The definitions of the imaginary-part data loss $\mathcal{L}_d^\Im$ and PDE loss $\mathcal{L}_p^\Im$ are similar to Eqs. (21) and (22), respectively, and are not shown here for brevity.

For the cAINN in Fig. 2(a), a single network is used to model the real and imaginary parts of the sound pressure, and there are $2N$ neurons in each hidden layer. The trainable parameters of the network are updated to minimize a single total loss function, i.e.,

$$
\mathcal{L} = \mathcal{L}_d^\Re + \mathcal{L}_p^\Re + \mathcal{L}_d^\Im + \mathcal{L}_p^\Im. \quad (23)
$$

For the dAINN in Fig. 2(b), two independent networks are used to model the real and imaginary parts of the sound pressure independently with $N$ neurons in each hidden layer. The trainable parameters of two networks are updated to minimize the real-part total loss,

$$
\mathcal{L}^\Re = \mathcal{L}_d^\Re + \mathcal{L}_p^\Re, \quad (24)
$$

and the imaginary-part total loss,

$$
\mathcal{L}^\Im = \mathcal{L}_d^\Im + \mathcal{L}_p^\Im. \quad (25)
$$

Once trained, the AINN method can reconstruct the sound pressure at an arbitrary position $(x_e, y_e)$ as $\mathcal{N}_e(x_e, y) = \mathcal{N}^\Re(x_e, y_e) + i\mathcal{N}^\Im(x_e, y_e)$. The sound pressure gradient at that position can be reconstructed as $\partial \mathcal{N}(x_e, y_e)/\partial x_e$ along the $x$ direction and as $\partial \mathcal{N}(x_e, y_e)/\partial y_e$ along the $y$ direction through differentiation on the network output. The pressure gradient along the radial direction can be reconstructed as

$$
\frac{\partial \mathcal{N}(r_e, \phi_e)}{\partial r_e} = \frac{\partial \mathcal{N}(x_e, y_e)}{\partial x_e} \cos(\phi_e) + \frac{\partial \mathcal{N}(x_e, y_e)}{\partial y_e} \sin(\phi_e). \quad (26)
$$

Here, are some comments on the proposed AINN method and recommended configurations.

### 1. Using tanh as the activation function

We use tanh as the activation function for two reasons. First, the tanh function is a smooth function, whose second-order gradient can be computed. This is necessary for the Laplacian operators in Eqs. (2) and (3). Second, the tanh function outputs positive or negative values according to the input. This makes it easier to model sound pressure, whose value can be either positive or negative. Sinusoidal functions also meet these two criteria and, thus, may also be used as the activation function. However, our trial results indicated that using tanh as the activation function shows better results than using the sin function. Therefore, we choose tanh over sin in this paper.

### 2. Cartesian coordinates vs polar coordinates

For the AINN method, we express sound pressures in Cartesian coordinates instead of polar coordinates for two reasons. First, the presence of the $1/r$ term can make the polar-coordinate Laplacian operator in Eq. (3) numerically unstable. Second, when a circular microphone array is used, no information about the sound field variation along the radial direction can be measured. In this case, the AINN method is unable to accurately estimate the first- and second-order radial gradients needed for calculating the Laplacian operator in polar coordinates [Eq. (3)].

### 3. Loss function

The loss functions in Eqs. (23)–(25) consist of the data loss and PDE loss. The data loss prompts the network output to approximate the measured sound pressure at positions $\{x_q, y_q\}_{q=1}^Q$, which are on the boundary of the ROI as shown in Fig. 1. The PDE loss, on the other hand, regularizes the network output to conform with the Helmholtz equation at positions $\{x_d, y_d\}_{d=1}^D$ on the boundary of and within the ROI. $D$ should be chosen to be sufficiently large and, thus, the distance between adjacent positions is at most half (ideally, one-tenth) of the wavelength for the frequency of interest. It is noted that when calculating the PDF loss in Eq. (22), the Helmholtz equation was divided by the term $(\omega/c)^2$ such that the output error has a consistent physical dimension with sound pressure. Therefore, theoretically, there is no need to further balance the data loss and PDE loss with an extra weighting factor.

### 4. Neuron number

As displayed in Eq. (8), the sound pressure can be expressed as a linear combination of $2N + 1$ CHs $\{J_n(kr)e^{-in\phi}\}_{n=-N}^N$, which are solutions of the Helmholtz equation [Eq. (1)]. As depicted in Figs. 2(a) and 2(b), the very same sound pressure can also be expressed as a linear combination of the output of a number of neurons. This fact inspires us to set the number of neurons in hidden layers according to the CH decomposition of the sound pressure.

J. Acoust. Soc. Am. **156** (3), September 2024

Ma *et al.* 2013

TABLE I. Trainable parameter number of the dAINN method as a function of frequency for reconstructing a 0.1 m radius sound field.

| Frequency | Number of layers, $L$ | Number of neurons, $N$ | Number of trainable parameters |
|---|---|---|---|
| $f = 1$ kHz | 2 | 2 | 11 |
| $f = 2$ kHz | 2 | 4 | 29 |
| $f = 3$ kHz | 2 | 6 | 57 |
| $f = 4$ kHz | 2 | 8 | 99 |

Specifically, for the cAINN, the single network design, the neuron number is set to be $2N$, and for the dAINN, the two network design, the neuron number is set to be $N$, where $N$ can be calculated from Eq. (6). Based on Eq. (6), for a sound field within a circular ROI with a $r = 0.1$ m radius, the number of trainable parameters $N$ of the dAINN method are listed in Table I when the number of hidden layers is $L = 2$. As shown in Table I, a dAINN with no more than 200 trainable parameters is sufficient for modeling a $r = 0.1$ m radius sound field up to 4 kHz, considering the real and imaginary parts of the complex sound pressure. To model a sound field of a different size and at other frequencies, we can design the cAINN and dAINN based on Eq. (6) accordingly. The AINN is compact in comparison to other learning-based methods, thus, is easier and faster to train. For example, Karakonstantis *et al.* (2024) employed a network with 5 hidden layers, each with 512 neurons. It is worth noting that unlike the time domain approach in Karakonstantis *et al.* (2024), our proposed AINN works in the frequency domain. To reconstruct the time-domain sound field, multiple compact AINNs, each for a frequency bin, can be trained in parallel. It is also noted that this paper designed the AINN architecture and parameters based on the CH expansion order, aiming to make it more explainable. Hyperparameter optimization of the AINN may lead to better performance but is not covered in this paper and will be investigated in the future.

## IV. EXPERIMENTS

Experiments were conducted to validate the performance of the proposed AINN method and compare it with the CH and SVD methods.

## A. Data processing

The SFR methods were evaluated using the University of Technology Sydney (UTS) multi-zone sound field reproduction dataset (Zhao *et al.*, 2022). The measurement setup is shown in Figs. 3 and 4. The RIRs between a loudspeaker array [Fig. 3(a)] and two microphone arrays [Figs. 3(b) and 3(c)] were measured in an anechoic chamber [Fig. 4(a)], a medium meeting room [Fig. 4(b)], and a small meeting room [Fig. 4(c)]. The microphone arrays were placed at the center of the circular array of 60 loudspeakers. The radius of the loudspeaker array is 1.5 m.

The loudspeaker array consisted of 60 Genelec 8010A Studio Monitors (Iisalmi, Finland; Zhao *et al.*, 2022), as shown in Fig. 3(a), and was placed approximately at the center of each room as shown in Fig. 4. The loudspeaker positions are

$$x_l = -R \sin((2l - 1)\pi/60) \, \text{m}, \qquad (27)$$

$$y_l = R \cos((2l - 1)\pi/60) \, \text{m}, \qquad (28)$$

where $R = 1.5$ m and $l = 1, 2, \ldots, 60$.

Two microphone arrays were constructed using the DPA 4060 Series Miniature Omnidirectional microphones (Kokkedal, Denmark; Zhao *et al.*, 2022), which were calibrated at 1 kHz. The first array is a planar 64-microphone square array with a side length of 0.28 m, as shown in Fig. 3(b). The microphone positions are

$$x_m = -0.14 + 0.04 \times \text{mod}(m - 1, 8) \, \text{m}, \qquad (29)$$

$$y_m = 0.14 - 0.04 \times \lfloor (m - 1)/8 \rfloor \, \text{m}, \qquad (30)$$

where $\lfloor \cdot \rfloor$ is the floor operation, and $m = 1, 2, \ldots, 64$. The second array is a dual-circular array with 30 microphones uniformly placed on each circle, as displayed in Fig. 3(c). The microphone positions are

$$\theta_m = 2\pi(m - 1)/30, \qquad (31)$$

$$x_m = -r \times \sin(\theta_m) \, \text{m}, \qquad (32)$$

$$y_m = r \times \cos(\theta_m) \, \text{m}, \qquad (33)$$

where $m = 1, 2, \ldots, 30$, and $r_1 = 0.12$ m for the exterior circle and $r_2 = 0.1$ m for the interior circle.
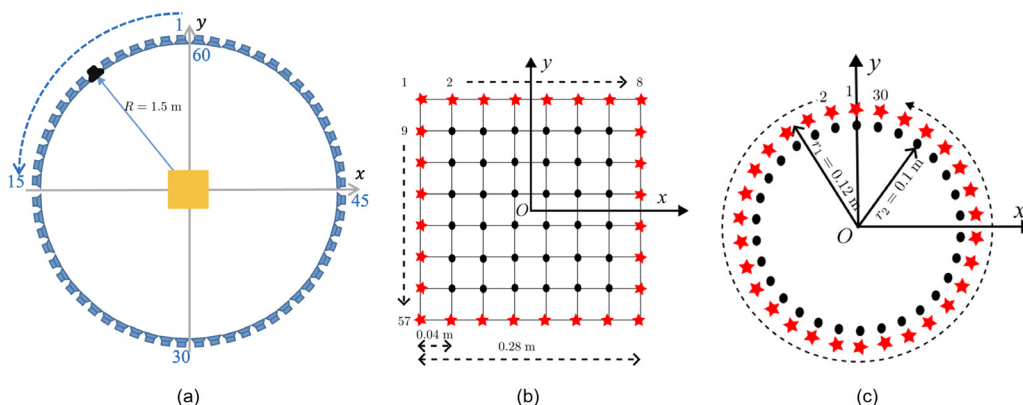


FIG. 3. (Color online) Experimental setup shows (a) the 60-loudspeaker array with a microphone array in the center, (b) the 64-microphone planar array with 28 boundary microphones ($\star$) and 36 interior microphones, and (c) the dual-circular microphone array with 30 exterior microphones ($\star$) and 30 interior microphones.
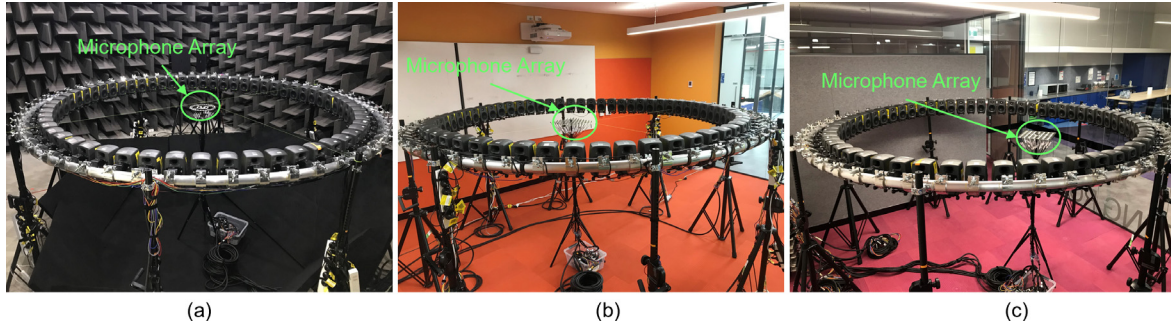
FIG. 4. (Color online) Photos for the experimental setups show (a) the Anechoic chamber, (b) the medium meeting room, and (c) the small meeting room.

The loudspeaker and microphone arrays were connected to a 64-input-64-output audio interface, which consists of four Yamaha RIO1608-D2 and four Yamaha RIO8-D (Shizuoka, Japan). The audio interface was commanded by a MATLAB program (MathWorks, Natick, MA) through a Dante virtual sound card (Sydney, Australia). A logarithmic sine sweep signal of 3 s duration with a frequency range from 20 Hz to 20 kHz was produced through each loudspeaker. The microphone recorded signals were processed to obtain the RIRs, which are 43 680 taps long at a sampling rate of 48 kHz. The RIRs were transformed into frequency domain transfer functions through the discrete Fourier transform (Oppenheim *et al.*, 1997), resulting in the complex sound pressures used in the experiments here. For more details about the measurement, please refer to Zhao *et al.* (2022).

### B. Implementation

The CH method was implemented based on Eqs. (4)–(11). According to Eq. (6), the dimensionalities of the sound field within the planar array are $N = 3, 6, 8$ for $f = 1$, 2, 3 kHz, respectively, and the dimensionalities of the sound field within the dual-circular array are $N = 3, 5, 7$ for $f = 1$, 2, 3 kHz, respectively.

The SVD method was implemented based on Eqs. (12)–(20). For a source (loudspeaker) located at $(x_s, y_s)$, the virtual point sources are uniformly arranged around the source in a 0.1 m × 0.1 m square, and the distance between two virtual point sources is 0.01 m. This amounts to 120 virtual point sources in total. The arrangement of the sources for the SVD method is shown in Fig. 5. It is noted that in the implementation of the CH and SVD methods according to Eqs. (7) and (17), respectively, no regularization has been employed. For the CH method, the condition numbers of the matrix $\mathbf{J}$ in Eq. (7) for the current experimental setup are less than ten, therefore, the regularization is not necessary. For the SVD method, incorporating regularization does not provide sensible improvement over current implementation.

To implement the proposed AINN method, we used the TensorFlow library and initialized the trainable parameters according to the Xavier initialization (Glorot and Bengio, 2010). The ADAM algorithm (Kingma and Ba, 2014), with a learning rate of 0.001, was used as the optimizer. The neuron number was set based on the dimensionality of the sound field under CH decomposition. The hidden layer number was set as one for $f = 1$ kHz and as two for $f = 2$, 3 kHz based on a trial-and-error process.

The data loss was calculated with respect to the measured sound pressures $\mathbf{P}_M$ at the corresponding coordinates $\{x_q, y_q\}_{q=1}^{Q}$. The PDE loss was calculated with respect to coordinates $\{x_d, y_d\}_{d=1}^{D}$ on the boundary of and within the ROI, which were consistently sampled with a uniform interval distance of 0.01 m between adjacent points, as depicted in Fig. 6. The AINN method was trained for $10^5$ epochs, and no early stop strategy was implemented. At each epoch, $\{\mathbf{P}_M, \{x_q, y_q\}_{q=1}^{Q}, \{x_d, y_d\}_{d=1}^{D}\}$ were fed to the AINN all at once (no batching). The specific values of $Q$ and $D$ depend on the experimental setup. For the sound pressure reconstruction experiments in Secs. IV D and IV E, $Q = 28$ because 28 measurement microphones are used and $D = 841$ for the uniformly sampled spatial points with an interval distance of 0.01 within the 0.28 m square (Fig. 6). For the pressure gradient reconstruction experiments in Secs. IV F and IV G, $Q = 30$ because 30 measurement microphones are used and $D = 625$ for the uniformly sampled spatial points with an interval distance of 0.01 within the 0.24 m square around the circular microphone array. The AINN was trained on a NVIDIA RTX 6000 graphics processing unit (GPU) (NVIDIA, Santa Clara, CA) with a 24 G random
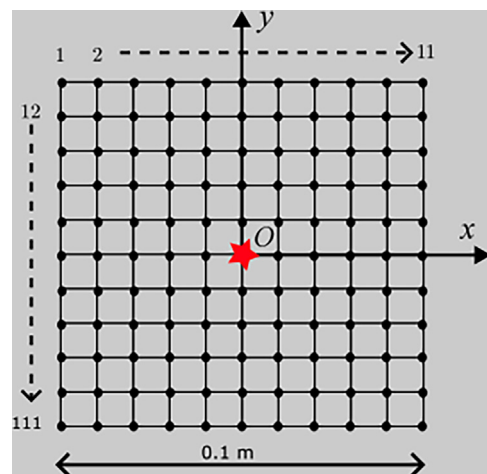


FIG. 5. (Color online) Arrangement of the sources for the SVD method, where $\star$ denotes the speaker position and $\bullet$ denotes the virtual point source positions, are shown.

J. Acoust. Soc. Am. **156** (3), September 2024

Ma *et al.* 2015

access memory (RAM), and the training of each AINN for a frequency took about 3 min.

In Secs. IV C–IV G, the performance of the proposed AINN method is compared to that of the CH and SVD methods when the same amount of data are available for all three methods, which is a fair comparison. It would also be interesting to compare the amount of training data required by these methods to reconstruct a sound field with a certain accuracy, which is supposed to depend on frequency and the accuracy threshold. This will be investigated in future work.

## C. Performance metrics

The performance of all methods was evaluated by the relative reconstruction error,

$$\xi = 10 \log 10 \frac{\sum_{e=1}^{E} \|P(x_e, y_e) - \hat{P}(x_e, y_e)\|_2^2}{\sum_{e=1}^{E} \|P(x_e, y_e)\|_2^2}, \quad (34)$$

where $\| \cdot \|_2$ denotes the $\ell_2$ norm, $P(x_e, y_e)$ and $\hat{P}(x_e, y_e)$ are the ground truth sound pressure and its reconstruction, respectively, and $\{x_e, y_e\}_{e=1}^{E}$ are coordinates of the sound pressure reconstruction positions, where $E$ is the total number of reconstruction positions. The pressure gradient reconstruction error was defined similarly to Eq. (34). The logarithmic error measure was chosen because it can represent a wide range of values compactly.

## D. Sound pressure reconstruction: Loudspeaker 7

Based on the sound pressures measured by the 28 exterior microphones of the planar array [Fig. 3(b)], we reconstructed the sound pressures at the 36 interior microphones within the array. Figures 7–9 show the real parts of the sound pressure due to loudspeaker 7 [the black loudspeaker in Fig. 3(a)] at 1, 2, and 3 kHz in the anechoic chamber, the medium room, and the small room, respectively. Figures 7–9
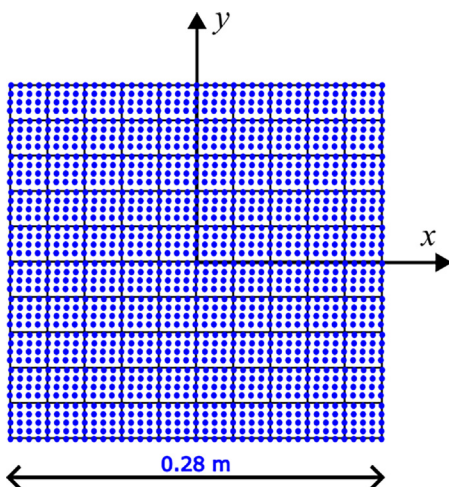


FIG. 6. (Color online) Arrangement of points (•) used for PDE loss calculation for the AINN method are displayed.

also show the reconstructions by the CH method, SVD method, AINN method, and corresponding (real and imaginary parts) reconstruction errors $\xi$. The results for the imaginary parts are similar and are not shown here for brevity.

As shown in Figs. 7–9, the dAINN, two network design, outperforms the cAINN, single network design, at all frequencies and in all room environments. Experiments for the cAINN with more or less neurons on hidden layers, i.e., $\lceil 3N/2 \rceil$ or $3N$, were also conducted, and the results were also inferior to those for the dAINN, thus, they are not revealed for brevity. Hereinafter, we focus only on the dAINN.

As shown in the second and third rows of Figs. 7–9, the SVD method performs better than the CH method and in most of the cases for $f = 2, 3$ kHz, with a reduction in the overall reconstruction error by 1.0–2.6 dB at 2 kHz and



(a0) 1 kHz Ground truth   (b0) 2 kHz Ground truth   (c0) 3 kHz Ground truth

(a1) 1 kHz, CH −13.5 dB   (b1) 2 kHz, CH −10.6 dB   (c1) 3 kHz, CH −6.5 dB

(a2) 1 kHz, SVD −11.8 dB   (b2) 2 kHz, SVD −11.6 dB   (c2) 3 kHz, SVD −10.6 dB

(a3) 1 kHz, cAINN −10.5 dB   (b3) 2 kHz, cAINN −5.9 dB   (c3) 3 kHz, cAINN −7.2 dB

(a4) 1 kHz, dAINN −14.5 dB   (b4) 2 kHz, dAINN −18.5 dB   (c4) 3 kHz, dAINN −11.4 dB
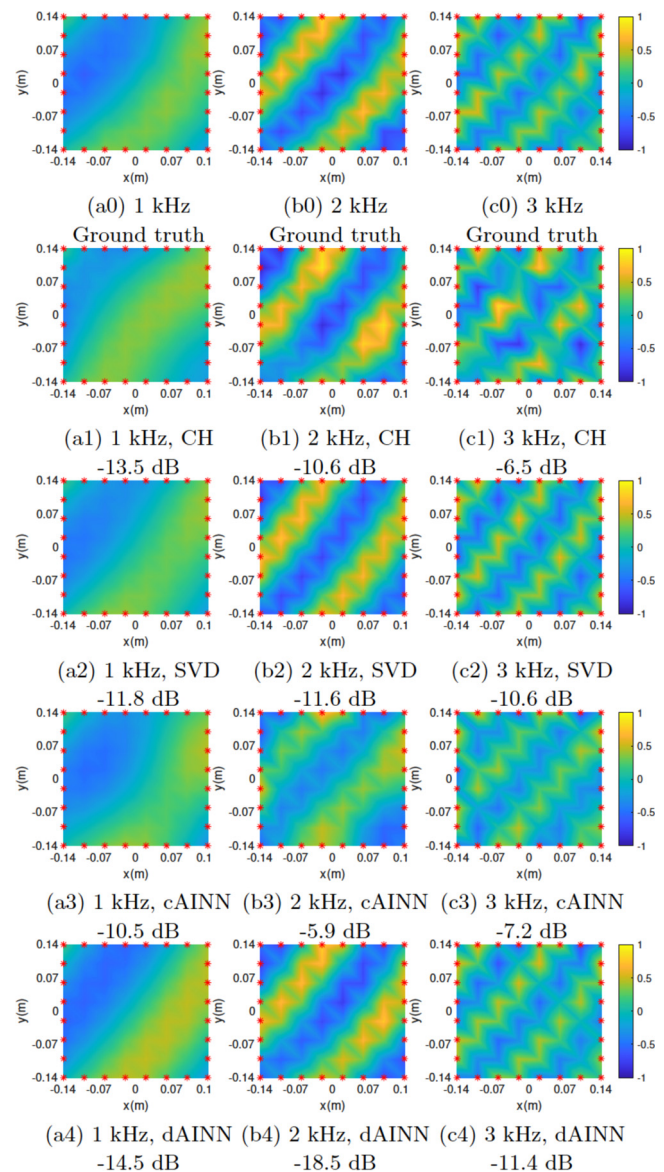
FIG. 7. (Color online) Comparison of measured and reconstructed sound pressures within the planar array based on the measured sound pressures by the boundary microphones (★) in the anechoic chamber. The pressure reconstruction error $\xi$ is calculated based on Eq. (34).
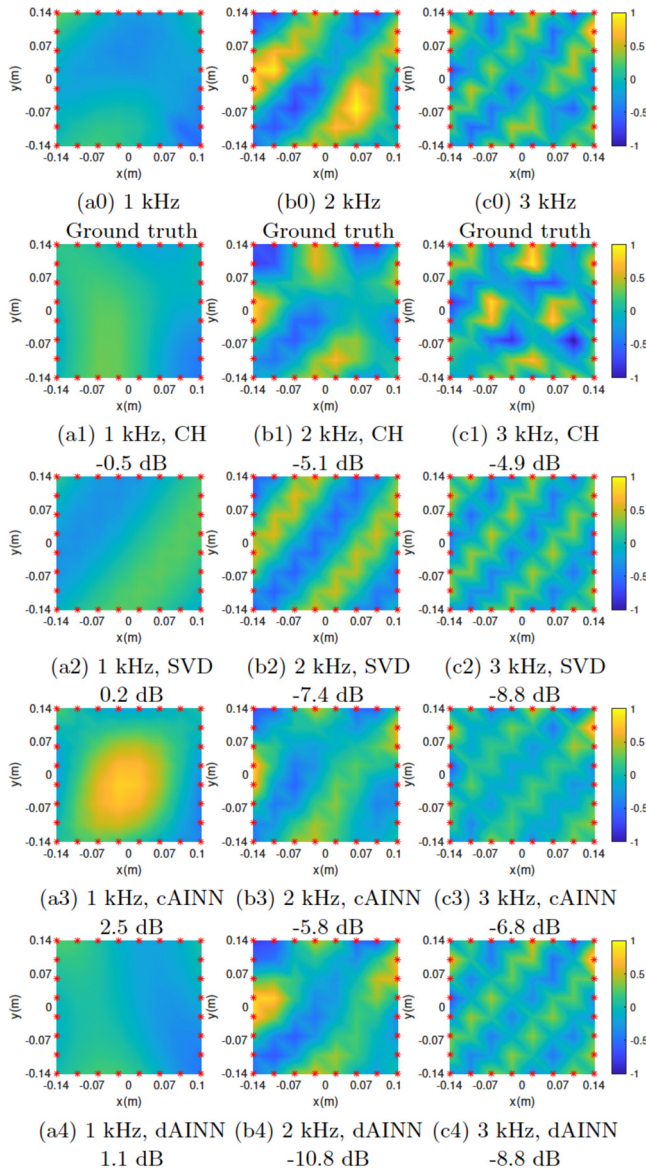
FIG. 8. (Color online) Comparison of measured and reconstructed sound pressures within the planar array based on the measured sound pressures by the boundary microphones (⋆) in the medium meeting room. The pressure reconstruction error $\xi$ is calculated based on Eq. (34).
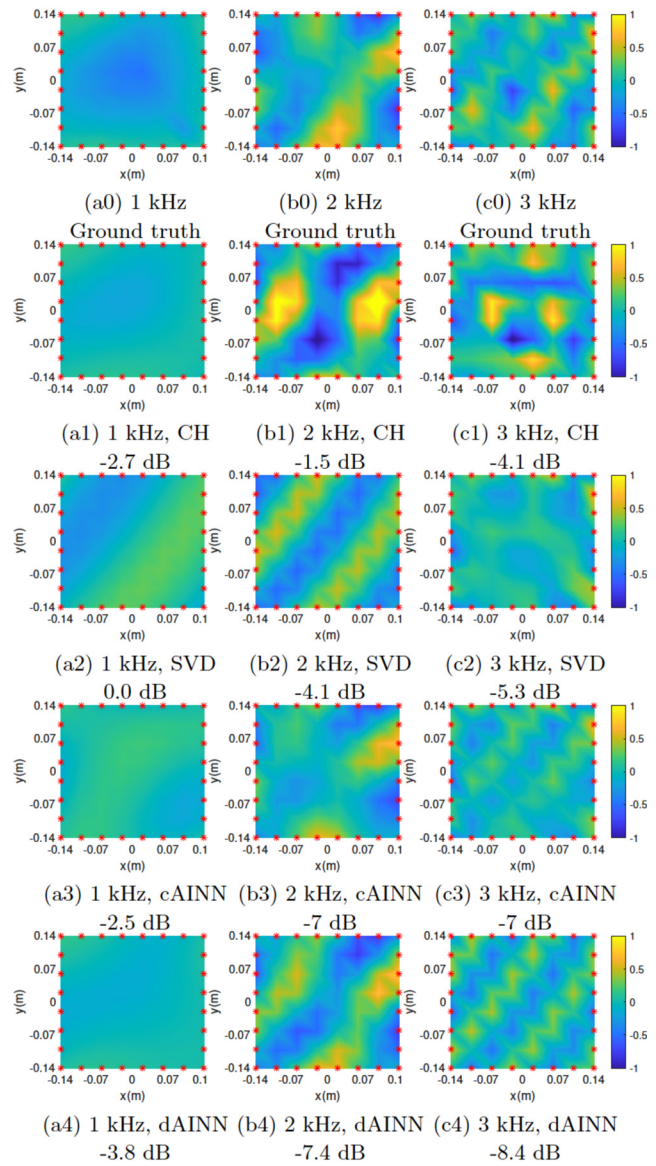


FIG. 9. (Color online) Comparison of measured and reconstructed sound pressures within the planar array based on the measured sound pressures by the boundary microphones (⋆) in the small meeting room. The pressure reconstruction error $\xi$ is calculated based on Eq. (34).

1.2–4.1 dB at 3 kHz across the three rooms. This could be attributed to the facts that prior information of the sound source location was used in the SVD method (Zhu *et al.*, 2020, 2021) or the noncircular/cylindrical geometry of the array does not allow the CH method to work effectively. In contrast, the proposed dAINN outperforms the SVD method in all of the tested rooms at $f = 2, 3$ kHz, although no prior information about the sound source location was required by the dAINN method. This could be attributed to the fact that the CH and SVD methods relied on only the sound pressures measured on the edge of the planar array for SFR. The measured pressures did not necessarily contain sufficient information to fully determine the sound field within the planar array. The dAINN, on the other hand, exploited the Helmholtz equation for regularizing the SFR within the

array through the PDE loss and reconstructed the sound field within the planar array more accurately.

As shown in the first column of Fig. 7, at 1 kHz, all methods achieve lower than $-10$ dB reconstruction errors in the anechoic chamber. However, Figs. 8(a1)–8(a4), 9(a11), 9(a12)–9(a41), and 9(a42) show that in the medium and small rooms, the pressure reconstruction errors of all of the methods are poor. Specifically, the SVD method, cAINN method, and dAINN method achieve larger than 0 dB pressure reconstruction errors in the medium room. This may be attributed to the fact that in the medium [Fig. 8(a0)] and small rooms [Fig. 9(a0)], under the particular loudspeaker-microphone array setup, the interference between the direct sound and room reflected sound makes the sound field within the planar array show no viable wave front as in the
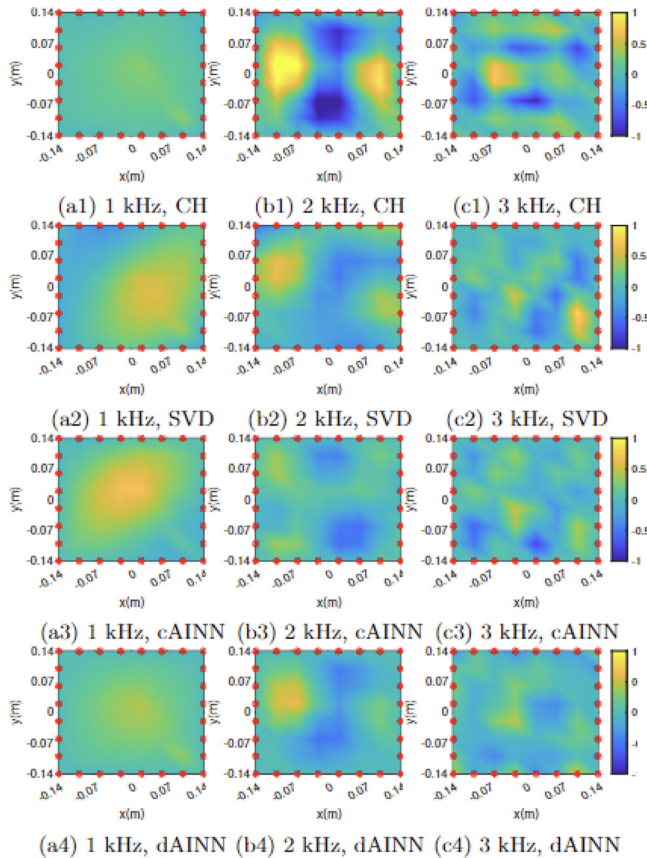
J. Acoust. Soc. Am. **156** (3), September 2024

Ma *et al.*      2017

(a1) 1 kHz, CH  (b1) 2 kHz, CH  (c1) 3 kHz, CH

(a2) 1 kHz, SVD  (b2) 2 kHz, SVD  (c2) 3 kHz, SVD

(a3) 1 kHz, cAINN  (b3) 2 kHz, cAINN  (c3) 3 kHz, cAINN

(a4) 1 kHz, dAINN  (b4) 2 kHz, dAINN  (c4) 3 kHz, dAINN

FIG. 10. (Color online) Sound pressure reconstruction errors $[P^{\Re}(x_e, y_e) - \hat{P}^{\Re}(x_e, y_e)]$ within the planar array based on the measured pressures by the boundary microphones ($\star$) in the small meeting room.
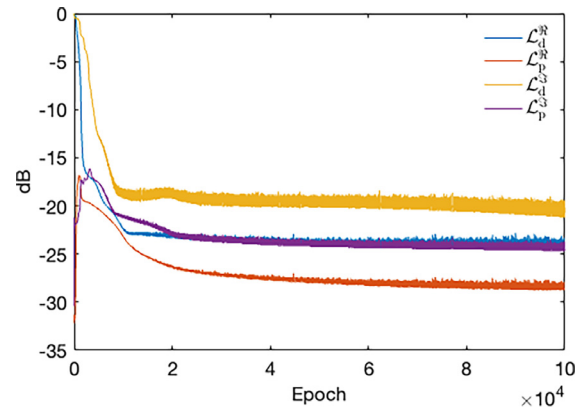


FIG. 11. (Color online) The learning curves for the real/imaginary part data and PDE losses of the dAINN method in the medium room at 2 kHz.

anechoic chamber [Fig. 7(a0)]. This causes the methods to be unable to accurately reconstruct the sound field. An extension of this work is to exploit knowledge of the sound source(s) position to further improve the performance of the AINN method, which will be investigated in the future.

Figure 10 further shows the difference, i.e., $P^{\Re}(x_e, y_e) - \hat{P}^{\Re}(x_e, y_e)$, between the ground truth and the reconstructions by different methods in the small meeting room. It can be observed that the difference is close to zero on the boundary of the microphone array but is larger in the interior of the microphone array. The results are expected as the pressures on the boundary are known and used in the calculations or training. Larger errors are observed at 1 kHz than at 2 kHz and 3 kHz, which is consistent with Fig. 9.

Figure 11 depicts the learning curves for real/imaginary-part data and PDE losses of the dAINN method in the medium room at 2 kHz, which corresponds to Fig. 6(b4). The learning curves of the dAINN method at other frequencies and other room environments are similar to those in Fig. 11 and, thus, are not shown here for brevity. As displayed in Fig. 11, the learning curves converge after 50 000 epochs of training. After convergence, the difference between the data and PDE losses for the real part is $|\mathcal{L}_d^{\Re} - \mathcal{L}_p^{\Re}| \approx 4$ dB, and the corresponding difference for the imaginary part is $|\mathcal{L}_d^{\Im} - \mathcal{L}_p^{\Im}| \approx 3$ dB.

### E. Sound pressure reconstruction: All loudspeakers

The SFR experiment was further conducted for other loudspeakers with one loudspeaker working each time. For each loudspeaker, we set the transfer functions between it and the planar array as the sound pressure. We used the sound pressure measured by the exterior 28 microphones to reconstruct the sound pressure at the interior 36 microphones within the planar array. Figure 12 compares the sound pressure reconstruction errors of the CH, SVD, and dAINN methods in the three rooms for all loudspeakers at 3 kHz.

The dAINN achieves the smallest average reconstruction errors of −12.5, −8, and −7.2 dB, in the anechoic chamber, medium room, and small room, respectively. Similar experiments were also conducted at 1 and 2 kHz. Except that the reconstruction errors were larger for the medium room and small room at 1 kHz, the results were similar to those in Fig. 12 and, thus, are not shown here for brevity. It is noted that Fig. 12 displays that the reconstruction errors in the anechoic chamber are smaller than those in other two rooms. That is because the boundaries and scatterers in the other two rooms make the sound field complex or of high frequency in the spatial domain, and the AINN has a hard time to learn the high-frequency spatial feature (Xu et al., 2022).

### F. Pressure gradient reconstruction: Loudspeaker 7

The sound pressures measured by the 30 microphones on the exterior circle of the dual-circular array [Fig. 3(c)] were used to reconstruct the radial pressure gradients between the two circles. The ground truth radial pressure gradients could not be directly measured by the setup shown in Figs. 3 and 4, thus, were approximated as

$$\frac{\partial P(r, \theta_m)}{\partial r} \approx \frac{P(r_1, \theta_m) - P(r_2, \theta_m)}{r_1 - r_2}, \tag{35}$$

where $m = 1, 2, \ldots, 30$. The radial pressure gradients were reconstructed by the CH method, SVD method, and dAINN method through Eq. (9), Eq. (20), and Eq. (26), respectively.
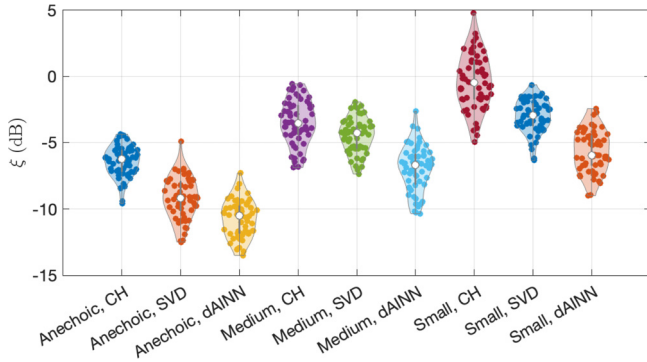
FIG. 12. (Color online) Sound pressure reconstruction errors for all loudspeakers at 3 kHz in three rooms. The label "small, CH" denotes the sound pressure reconstruction error of the CH method in the small room, and other labels are similarly defined.

To examine the effectiveness of the pressure gradient approximation, we performed a simulation based on the ideal free-field environment before experimenting with the real-measured transfer functions in different rooms. The

transfer functions between loudspeaker 7 and the dual-circular array were simulated using the free-field Green's function (Williams, 1999) with added white Gaussian noise at a signal-to-noise ratio (SNR) of 20 dB to model disturbances. The reconstruction errors are depicted in Fig. 13. As shown in Figs. 13(a1), 13(b1), and 13(c1), the reconstruction errors of three methods are relatively small, i.e., $< -10$ dB, when using the simulated sound pressures for reconstruction, which demonstrate the effectiveness of the approach.

The experimental results for the real-measured transfer functions in different rooms are shown in Figs. 13(a2)–13(c4). It can be observed that the reconstruction errors of all three methods show signs of degradation compared to the results with simulated transfer functions. The CH and SVD methods exhibit significant deviations from the ground truth in the medium and small rooms. In contrast, by exploiting the data-independent Helmholtz equation for regularization, the dAINN method is less susceptible to the inherent disturbances of the measurement processes. Thus, the dAINN method achieves the least pressure gradient reconstruction errors in most cases.



(a1) Free-space, 1 kHz
(b1) Free-space, 2 kHz
(c1) Free-space, 3 kHz
(a2) Anechoic chamber, 1 kHz
(b2) Anechoic chamber, 2 kHz
(c2) Anechoic chamber, 3 kHz
(a3) Medium room, 1k
(b3) Medium room, 2 kHz
(c3) Medium room, 3 kHz
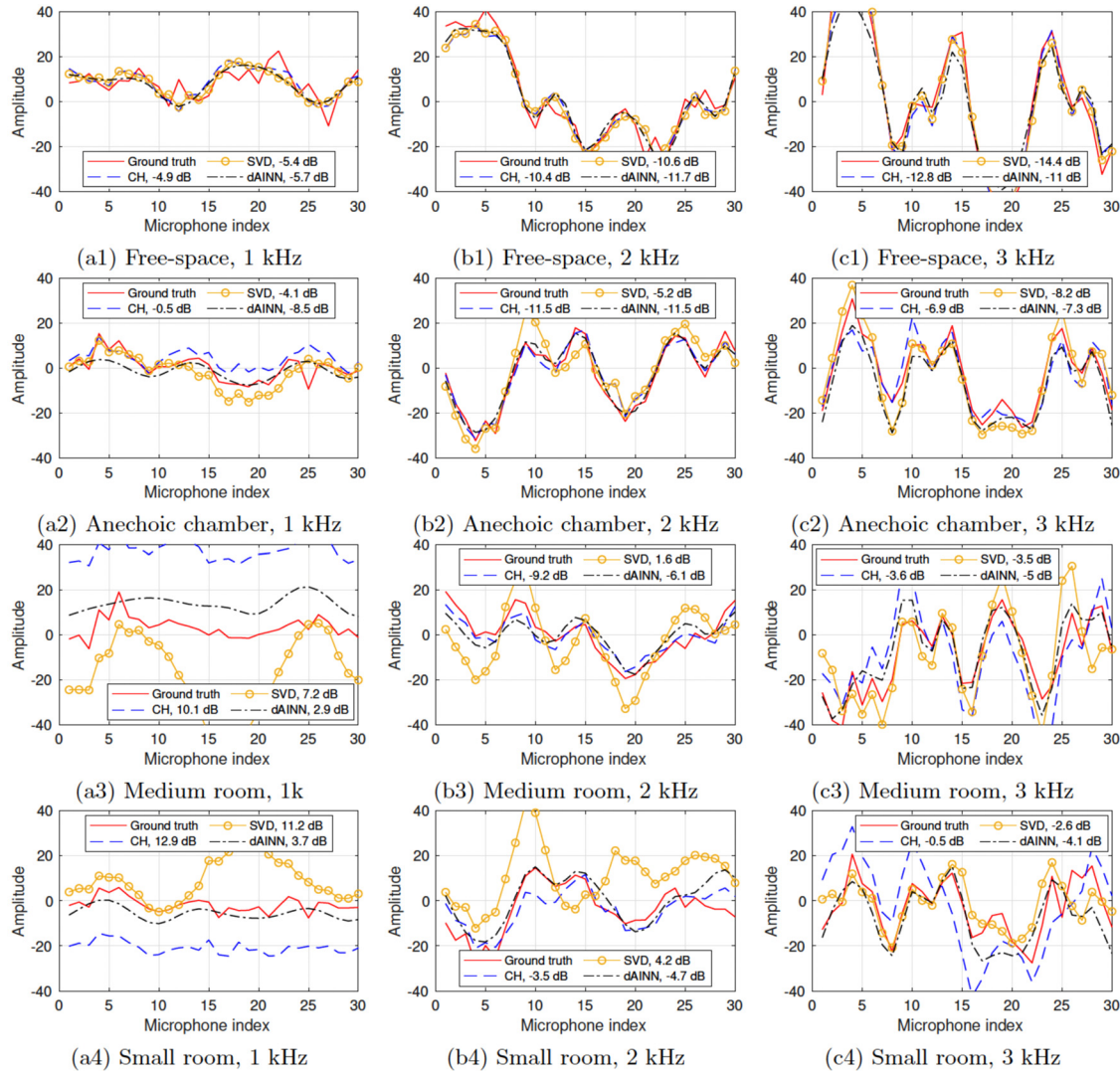(a4) Small room, 1 kHz
(b4) Small room, 2 kHz
(c4) Small room, 3 kHz

FIG. 13. (Color online) Comparison of measured and reconstructed radial pressure gradients based on the measurement of 30 microphones along the exterior circle of the dual-circular array for loudspeaker 7. The numbers in the legends denote the reconstruction errors by the SH, SVD, and dAINN methods.
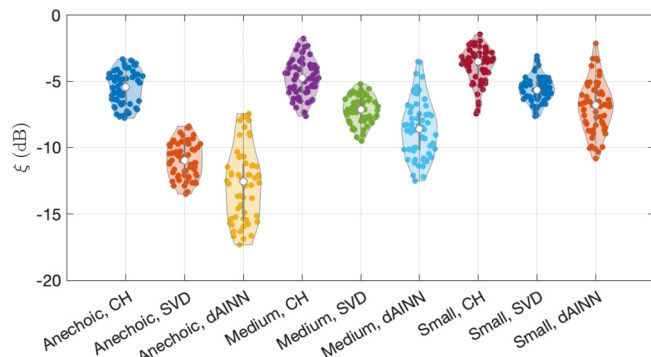
FIG. 14. (Color online) Pressure gradient reconstruction errors for all loud-speakers at 3 kHz in three rooms. The label "small, CH" denotes the sound pressure reconstruction error of the CH method in the small room, and other labels are similarly defined.

### G. Pressure gradient reconstruction: All loudspeakers

The pressure gradient reconstruction experiment was repeated over all 60 loudspeakers. We set each loudspeaker to the dual-circular array's transfer function as the sound pressure, approximated the radial pressure gradient according to Eq. (35), and reconstructed the pressure gradient in the same way as in Sec. IV F.

Figure 14 shows reconstruction errors for all loudspeakers in the three rooms by the CH, SVD, and dAINN methods at 3 kHz. The experimental results for 1 and 2 kHz showed similar results as for that in Fig. 14 and, thus, are not shown for brevity. Comparing Fig. 14 with Fig. 12, we can see that the pressure gradient is more challenging to reconstruct than the sound pressure. The dAINN achieves the smallest average reconstruction errors of –11.5, –7.0, and –5.5 dB in the anechoic chamber, medium room, and small room, respectively. The average reconstruction errors of the SVD method are about 2 dB higher than those of the dAINN method. The reconstruction errors of the CH method are the worst and exceed 0 dB in the small room.

It is noted that in this paper, the pressure gradient was approximated by the finite difference [Eq. (35)]. This may not be accurate in some circumstances and may contribute to the relatively high reconstruction errors depicted in Figs. 13 and 14. The velocity sensor (De Bree, 2003) may be used for measuring the pressure gradient and testing the performance of the pressure gradient reconstruction in the future.

### V. CONCLUSION

This paper proposed a compact AINN method for SFR. A neural network was designed to approximate the measured sound pressure and obey the Helmholtz equation, which regularized the network to generate physically valid output at and beyond the measurement positions. The performance of the AINN method was validated by sound pressure and pressure gradient reconstruction experiments and outperformed the CH and SVD methods. The design of the AINN method, specifically its width and depth, is still empirical. Further theoretical investigation is needed to

provide better guidance on the AINN design, which will be one of our future research topics.

### AUTHOR DECLARATIONS
#### Conflict of Interest

The authors have no conflicts to disclose.

### DATA AVAILABILITY

The data that support the findings of this study are openly available in UTS Stash at https://doi.org/10.26195/0wx8-v473, Zhao et al. (2022).

Antonello, N., De Sena, E., Moonen, M., Naylor, P. A., and Van Waterschoot, T. (2017). "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," IEEE/ACM Trans. Audio. Speech. Lang. Process. 25(10), 1929–1941.

Benesty, J., Chen, J., and Huang, Y. (2008). Microphone Array Signal Processing (Springer Science and Business Media, Heidelberg, Germany), Vol. 1.

Borrel-Jensen, N., Goswami, S., Engsig-Karup, A. P., Karniadakis, G. E., and Jeong, C.-H. (2024). "Sound propagation in realistic interactive 3D scenes with parameterized sources using deep neural operators," Proc. Natl. Acad. Sci. U.S.A. 121(2), e2312159120.

Chen, H., Abhayapala, T. D., and Zhang, W. (2015). "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," J. Acoust. Soc. Am. 138(5), 3081–3092.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F. (2022). "Scientific machine learning through physics–informed neural networks: Where we are and what's next," J. Sci. Comput. 92(3), 88.

De Bree, H.-E. (2003). "The microflown: An acoustic particle velocity sensor," Acoust. Aust. 31(3), 91–94.

Fernandez-Grande, E. (2016). "Sound field reconstruction using a spherical microphone array," J. Acoust. Soc. Am. 139(3), 1168–1178.

Fernandez-Grande, E., Karakonstantis, X., Caviedes-Nozal, D., and Gerstoft, P. (2023). "Generative models for sound field reconstruction," J. Acoust. Soc. Am. 153(2), 1179–1190.

Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, May 13–15, 2010, Sardinia, Italy, pp. 249–256.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning (MIT Press, Cambridge, MA).

Hahmann, M., and Fernandez-Grande, E. (2022). "A convolutional plane wave model for sound field reconstruction," J. Acoust. Soc. Am. 152(5), 3059–3068.

Hahmann, M., Verburg, S. A., and Fernandez-Grande, E. (2021). "Spatial reconstruction of sound fields using local and data-driven functions," J. Acoust. Soc. Am. 150(6), 4417–4428.

Karakonstantis, X., Caviedes-Nozal, D., Richard, A., and Fernandez-Grande, E. (2024). "Room impulse response reconstruction with physics-informed deep learning," arXiv:2401.01206.

Kennedy, R. A., Sadeghi, P., Abhayapala, T. D., and Jones, H. M. (2007). "Intrinsic limits of dimensionality and richness in random multipath fields," IEEE Trans. Signal Process. 55(6), 2542–2556.

2020    J. Acoust. Soc. Am. 156 (3), September 2024

Ma et al.

Kingma, D. P., and Ba, J. (**2014**). "Adam: A method for stochastic optimization," arXiv:1412.6980.

Kristoffersen, M. S., Møller, M. B., Martínez-Nuevo, P., and Østergaard, J. (**2021**). "Deep sound field reconstruction in real rooms: Introducing the ISOBEL sound field dataset," arXiv:2102.06455.

Lluis, F., Martinez-Nuevo, P., Bo Møller, M., and Ewan Shepstone, S. (**2020**). "Sound field reconstruction in rooms: Inpainting meets super-resolution," J. Acoust. Soc. Am. **148**(2), 649–659.

Middleton, M., Murphy, D. T., and Savioja, L. (**2023**). "The application of Fourier neural operator networks for solving the 2D linear acoustic wave equation," in *Forum Acusticum*, September 11–15, 2023, Torino, Italy (European Acoustics Association).

Olivieri, M., Pezzoli, M., Antonacci, F., and Sarti, A. (**2021**). "A physics-informed neural network approach for nearfield acoustic holography," Sensors **21**(23), 7834.

Oppenheim, A. V., Willsky, A. S., Nawab, S. H., and Ding, J.-J. (**1997**). *Signals and Systems* (Prentice Hall, Upper Saddle River, NJ), Vol. 2.

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

Rafaely, B. (**2015**). *Fundamentals of Spherical Array Processing* (Springer, New York), Vol. 8.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (**2019**). "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. **378**, 686–707.

Schmid, J. D., Bauerschmidt, P., Gurbuz, C., Eser, M., and Marburg, S. (**2024**). "Physics-informed neural networks for acoustic boundary admittance estimation," Mech. Syst. Signal Process. **215**, 111405.

Schmid, J. M., Fernandez-Grande, E., Hahmann, M., Gurbuz, C., Eser, M., and Marburg, S. (**2021**). "Spatial reconstruction of the sound field in a room in the modal frequency range using Bayesian inference," J. Acoust. Soc. Am. **150**(6), 4385–4394.

Shigemi, K., Koyama, S., Nakamura, T., and Saruwatari, H. (**2022**). "Physics-informed convolutional neural network with bicubic spline interpolation for sound field estimation," in *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 5–8, 2022, Bamberg, Germany (IEEE, New York), pp. 1–5.

Skudrzyk, E. (**2012**). *The Foundations of Acoustics: Basic Mathematics and Basic Acoustics* (Springer Science &and Business Media, Vienna, Austria).

Tang, X., Zhang, J., Alon, D. L., Ben-Hur, Z., Samarasinghe, P., and Abhayapala, T. (**2022**). "Wave domain sound field interpolation using two spherical microphone arrays," in *2022 30th European Signal Processing Conference (EUSIPCO)*, August 29–September 2, 2022, Belgrade, Serbia (IEEE, New York), pp. 319–323.

Verburg, S. A., and Fernandez-Grande, E. (**2018**). "Reconstruction of the sound field in a room using compressive sensing," J. Acoust. Soc. Am. **143**(6), 3770–3779.

Wabnitz, A., Epain, N., van Schaik, A., and Jin, C. (**2011**). "Time domain reconstruction of spatial sound fields using compressed sensing," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 22–27, 2011, Prague, Czech Republic (IEEE, New York), pp. 465–468.

Williams, E. G. (**1999**). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York).

Xu, Z., Zhang, Y., and Luo, T. (**2022**). "Overview frequency principle/spectral bias in deep learning," arXiv:2201.07395.

Zhang, W., Kennedy, R. A., and Abhayapala, T. D. (**2008**). "Iterative extrapolation algorithm for data reconstruction over sphere," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 31–April 4, 2008, Las Vegas, NV (IEEE, New York), pp. 3733–3736.

Zhang, X., Lou, J., Zhu, S., Lu, J., and Li, R. (**2023**). "Sound field reconstruction using prolate spheroidal wave functions and sparse regularization," Sensors **23**(19), 8312.

Zhao, S., Zhu, Q., Cheng, E., and Burnett, I. S. (**2022**). "A room impulse response database for multizone sound field reproduction (L)," J. Acoust. Soc. Am. **152**(4), 2505–2512.

Zhu, Q., Qiu, X., Coleman, P., and Burnett, I. (**2020**). "A comparison between two modal domain methods for personal audio reproduction," J. Acoust. Soc. Am. **147**(1), 161–173.

Zhu, Q., Qiu, X., Coleman, P., and Burnett, I. (**2021**). "An experimental study on transfer function estimation using acoustic modelling and singular value decomposition," J. Acoust. Soc. Am. **150**(5), 3557–3568.

J. Acoust. Soc. Am. **156** (3), September 2024

Ma *et al.* 2021

09 January 2025 04:10:11