Physics-Informed Machine Learning For Sound Field Estimation

Invited paper for the IEEE SPM Special Issue on Model-based Data-Driven Audio Signal Processing

Shoichi Koyama, *Senior Member, IEEE,* Juliano G. C. Ribeiro, *Member, IEEE,* Tomohiko Nakamura, *Member, IEEE,* Natsuki Ueno, *Member, IEEE,* and Mirco Pezzoli, *Member, IEEE*

Abstract

The area of study concerning the estimation of spatial sound, i.e., the distribution of a physical quantity of sound such as acoustic pressure, is called *sound field estimation*, which is the basis for various applied technologies related to spatial audio processing. The sound field estimation problem is formulated as a function interpolation problem in machine learning in a simplified scenario. However, high estimation performance cannot be expected by simply applying general interpolation techniques that rely only on data. The physical properties of sound fields are useful a priori information, and it is considered extremely important to incorporate them into the estimation. In this article, we introduce the fundamentals of *physics-informed machine learning (PIML)* for sound field estimation and overview current PIML-based sound field estimation methods.

Index Terms

Sound field estimation, kernel methods, physics-informed machine learning, physics-informed neural networks

This work was supported by JST FOREST Program, Grant Number JPMJFR216M, JSPS KAKENHI, Grant Number 23K24864, and the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on “Telecommunications of the Future” (PE00000001—program “RESTART”).

Manuscript received April xx, 20xx; revised August xx, 20xx.

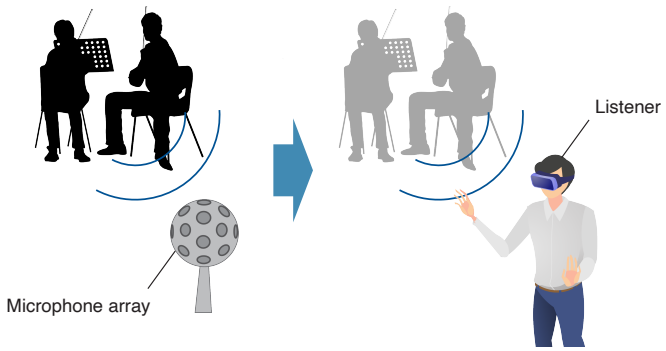


Fig. 1. Sound field recording for VR audio. The sound field captured by a microphone array is reproduced by loudspeakers or headphones. The estimated sound field should account for listener movement and rotation within large regions.

 DOI 10.1109/MSP.2024.3465896

I. Introduction

Sound field estimation, which is also referred to as sound field reconstruction, capturing, and interpolation, is a fundamental problem in acoustic signal processing, which is aimed at reconstructing a spatial acoustic field from a discrete set of microphone measurements. This essential technology has a wide variety of applications, such as room acoustic analysis, the visualization/auralization of an acoustic field, spatial audio reproduction using a loudspeaker array or headphones, and active noise cancellation in a spatial region. In particular, virtual/augmented reality (VR/AR) audio would be one of the most remarkable recent applications of this technology, as it requires capturing a sound field in a large region by multiple microphones (see Fig. 1). The spatial reconstruction of room impulse responses (RIRs) or acoustic transfer functions (ATFs), which is a special case of sound field estimation, can be applied to the estimation of steering vectors for beamforming and also head-related transfer functions (HRTFs) for binaural reproduction.

Sound field estimation has been studied for a number of years. Estimating a sound field in the inverse direction of wave propagation based on wave domain (or spatial frequency domain) processing, i.e., the basis expansion of a sound field into plane wave or spherical wave functions, has been particularly applied to acoustic imaging tasks [1]. This technique has been introduced in the signal-processing field in recent decades. In particular, spherical harmonic domain processing using a spherical microphone array has been intensively investigated because its isotropic nature is suitable for spatial sound processing [2].

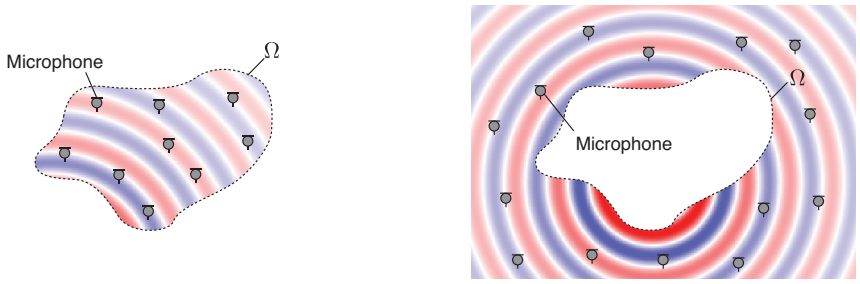
Wave domain processing has been applied to, for instance, spatial audio recording, source localization, source separation, and spatial active noise control.

The theoretical foundation of wave domain processing is derived from expansion representations by the solutions of the governing partial differential equations (PDEs) of acoustic fields, i.e., the wave equation in the time domain and the Helmholtz equation in the frequency domain. Since the basis functions used in wave domain processing, such as plane wave and spherical wave functions, are solutions of these PDEs in a region not including sound sources or scattering objects, the estimate is guaranteed to satisfy the wave and Helmholtz equations.

Although the classical methods are solely based on physics, advanced signal processing and machine learning techniques have also been applied to the sound field estimation problem. In particular, sparse optimization techniques have been investigated to improve the estimation accuracy while preserving the physical constraints for the estimate [3]–[5]. The physical constraints are usually imposed by the expansion representation of the sound field as used in the classical methods.

On the basis of great successes in various fields, (deep) neural networks [6] have also been incorporated into sound field estimation techniques in recent years to exploit their high representational power. The network weights are optimized to predict the sound field by using a set of training data. Basically, this approach is purely data-driven; that is, prior information on physical properties is not taken into account, as opposed to the methods mentioned above. To induce the estimate to satisfy the governing PDEs, the loss function for evaluating the deviation from the governing PDEs has been incorporated, and such a technique is referred to as the *physics-informed neural networks (PINNs)* [7], [8], which have been applied to various forward and inverse problems for physical fields, such as flow field [9], seismic field [10], and acoustic field [11]. PINNs help prevent unnatural distortions in the estimate, which are unlikely to occur in practical acoustic fields, particularly when the number of observations is small. Currently, there exist various methods using PINNs or other neural networks for sound field estimation as described in detail in the later sections [12]–[15].

Physical properties are useful prior information in sound field estimation and methods based on physical properties have many advantages over purely data-driven approaches. Although sound field estimation methods have historically used physical constraints explicitly, with the recent development of machine learning techniques, *physics-informed machine learning (PIML)* in sound field estimation continues to grow significantly, especially after the advent of PINN [16]–[19]. We describe how to embed physical properties for various machine learning techniques, such as linear regression, kernel regression, and neural networks, in the sound field estimation problem and summarize the current PIML-based methods for sound field estimation. Furthermore, current limitations and future outlooks are also discussed.



(a) Interior problem (b) Exterior problem

Fig. 2. Sound field estimation of interior and exterior regions. In the interior problem, the sound field in the source-free interior region Ω is estimated. In contrast, the sound field in the source-free exterior region $\mathbb{R}^3\backslash\Omega$ is estimated in the exterior problem.

II. Sound field estimation problem

First, we formulate the sound field estimation problem. In this article, we mainly focus on an interior problem, that is, estimating a sound field inside a source-free region of the target with multiple microphones (see Fig. 2). Several methods introduced in this article are also applicable to the exterior or radiation problem.

The target region in the three-dimensional space is denoted as $\Omega \subset \mathbb{R}^3$. The pressure distribution in Ω is represented by $U : \Omega \times \mathbb{R} \to \mathbb{R}$ in the time domain and $u : \Omega \times \mathbb{R} \to \mathbb{C}$ in the frequency domain, which is generated by sound waves propagating from one or more sound sources outside Ω and also by reverberation. Here, U and u are continuous functions of the position $\boldsymbol{r} \in \Omega$ and the time $t \in \mathbb{R}$ or the angular frequency $\omega \in \mathbb{R}$. M microphones are placed at arbitrary positions inside Ω. The objective of the sound field estimation problem is to reconstruct U or u from microphone observations.

In the frequency domain, the estimation can be performed at each frequency assuming signal independence between frequencies. If the estimator of u is linear in the frequency domain, a finite-impulse-response (FIR) filter for estimating a sound field generated by usual sounds, such as speech and music, can be designed. Otherwise, the estimation is performed in the time–frequency domain, for example, by the short-time Fourier transform (STFT), or in the time domain. Particularly when the estimator is adapted to the environment, the direct estimation in the time domain is generally more challenging because the estimator for U may have a large dimension of spatiotemporal functions. When the source signal is an impulse signal, which means that U is a spatiotemporal function of an RIR, the problem can be relatively tractable because the time samples to be estimated are reduced to the length of RIRs at most and the estimation is generally performed offline.

For simplicity, the microphones are assumed to be omnidirectional, i.e., pressure microphones. In this scenario, the sound field estimation problem is equivalent to the general interpolation problem

because the observations are spatial (or spatiotemporal) samples of the multivariate scalar function to be estimated. Specifically, when the microphone observations in the frequency domain are denoted by $\boldsymbol{s}(\omega) = [s_1(\omega), \ldots, s_M(\omega)]^{\mathsf{T}} \in \mathbb{C}^M$ ($(\cdot)^{\mathsf{T}}$ is the transpose) obtained by the pressure microphones at $\{\boldsymbol{r}_m\}_{m=1}^M$, the observed values of $\{s_m(\omega)\}_{m=1}^M$ can be regarded as $\{u(\boldsymbol{r}_m, \omega)\}_{m=1}^M$ contaminated by some noise. In the time domain, the sound field should be estimated from the observations of M microphones and T time samples. Most methods introduced in this article can be generalized to the case of directional microphones because the measurement process can be assumed to be a linear operation. It is also possible for most methods to be generalized to estimate expansion coefficients of spherical wave functions of the sound field at a predetermined expansion center, which is particularly useful for spatial audio applications such as *ambisonics* [20].

We start with techniques of embedding physical properties in general interpolation techniques. Then, current studies of sound field estimation based on PIML are introduced.

III. Embedding physical properties in interpolation techniques

In general interpolation techniques, the estimation of the continuous function $f : \mathbb{R}^P \to \mathbb{K}$ ($\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$, i.e., f is U or u) from a discrete set of observations $\boldsymbol{y} \in \mathbb{K}^I$ at the sampling points $\{\boldsymbol{x}_i\}_{i=1}^I$ is achieved by representing f with some model parameters $\boldsymbol{\theta}$ and solving the following optimization problem:

$$\underset{\boldsymbol{\theta}}{\text{minimize}}\, \mathcal{L}\left(\boldsymbol{y}, \boldsymbol{f}(\{\boldsymbol{x}_i\}_{i=1}^I; \boldsymbol{\theta})\right) + \mathcal{R}(\boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{f}(\{\boldsymbol{x}_i\}_{i=1}^I; \boldsymbol{\theta}) = [f(\boldsymbol{x}_1; \boldsymbol{\theta}), \ldots, f(\boldsymbol{x}_I; \boldsymbol{\theta})]^{\mathsf{T}} \in \mathbb{K}^I$ is the vector of the discretized function f represented by $\boldsymbol{\theta}$, $\mathcal{L}$ is the loss function for evaluating the distance between $\boldsymbol{y}$ and f at $\{\boldsymbol{x}_i\}_{i=1}^I$, and $\mathcal{R}$ is the regularization term for $\boldsymbol{\theta}$ to prevent overfitting. A typical loss function is the squared error between $\boldsymbol{y}$ and $\boldsymbol{f}$, which is written as

$$\mathcal{L}\left(\boldsymbol{y}, \boldsymbol{f}(\{\boldsymbol{x}_i\}_{i=1}^I; \boldsymbol{\theta})\right) = \|\boldsymbol{y} - \boldsymbol{f}(\{\boldsymbol{x}_i\}_{i=1}^I; \boldsymbol{\theta})\|^2. \tag{2}$$

One of the simplest regularization terms is the squared ℓ_2-norm penalty (Tikhonov regularization) for $\boldsymbol{\theta}$ defined as

$$\mathcal{R}(\boldsymbol{\theta}) = \lambda \|\boldsymbol{\theta}\|^2, \tag{3}$$

where λ is the positive constant parameter. By using the solution of (1), which is denoted as $\hat{\boldsymbol{\theta}}$, we can estimate the function f as $f(\boldsymbol{x}; \hat{\boldsymbol{\theta}})$.

The general interpolation techniques described above highly depend on data. Prior information on the function to be estimated is basically useful for preventing overfitting, whose simplest example is the squared ℓ_2-norm penalty (3). In the context of sound field estimation, one of the most beneficial lines of

information will be the physical properties of the sound field. In particular, a governing PDE of the sound field, i.e., the wave equation in the time domain or the Helmholtz equation in the frequency domain, is an informative property because the function to be estimated should satisfy the governing PDE. The homogeneous wave and Helmholtz equations are respectively written as

$$\left(\nabla_{\boldsymbol{r}}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) U(\boldsymbol{r}, t) = 0 \tag{4}$$

and

$$\left(\nabla_{\boldsymbol{r}}^2 + k^2\right) u(\boldsymbol{r}, \omega) = 0, \tag{5}$$

where c is the sound speed and $k = \omega/c$ is the wave number. We introduce several techniques to embed these physical properties in the interpolation.

A. Regression with basis expansion into element solutions of governing PDEs

A widely used model for f is a linear combination of a finite number of basis functions. We define the basis functions and their weights as $\{\varphi_l(\boldsymbol{x})\}_{l=1}^L$ and $\{\gamma_l\}_{l=1}^L$, respectively, with $\varphi_l : \mathbb{R}^P \to \mathbb{K}$ and $\gamma_l \in \mathbb{K}$. Then, f is expressed as

$$\begin{aligned} f(\boldsymbol{x}; \boldsymbol{\gamma}) &= \sum_{l=1}^{L} \gamma_l \varphi_l(\boldsymbol{x}) \\ &= \boldsymbol{\varphi}(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{\gamma}, \end{aligned} \tag{6}$$

where $\boldsymbol{\varphi}(\boldsymbol{x}) = [\varphi_1, \ldots, \varphi_L]^{\mathsf{T}} \in \mathbb{K}^L$ and $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_L]^{\mathsf{T}} \in \mathbb{K}^L$. Thus, f is interpolated by estimating $\boldsymbol{\gamma}$ from $\boldsymbol{y}$.

A well-established technique used to constrain the estimate f to the governing PDEs is the expansion into the element solutions of PDEs, which is particularly investigated in the frequency domain to constrain the estimate to satisfy the Helmholtz equation. Several expansion representations for the Helmholtz equation have been applied to the sound field estimation problem. When considering the interior problem, the following representations are frequently used [21]:

- Plane wave expansion (or Herglotz wave function)

$$u(\boldsymbol{r}, \omega) = \int_{\mathbb{S}_2} \tilde{u}(\boldsymbol{\eta}, \omega) \mathrm{e}^{-\mathrm{j}k\langle \boldsymbol{\eta}, \boldsymbol{r} \rangle} \mathrm{d}\boldsymbol{\eta}, \tag{7}$$

  where $\tilde{u}$ is the complex amplitude of plane waves arriving from the direction $\boldsymbol{\eta} \in \mathbb{S}_2$ ($\mathbb{S}_2$ is the unit sphere). Although the plane wave function is usually defined by using the propagation direction $\tilde{\boldsymbol{\eta}} = -\boldsymbol{\eta}$, (7) is defined by using the arrival direction $\boldsymbol{\eta}$ for simplicity in the later formulations.

- Spherical wave function expansion

$$u(\boldsymbol{r},\omega)=\sum_{\nu=0}^{\infty}\sum_{\mu=-\nu}^{\nu}\mathring{u}_{\nu,\mu}(\boldsymbol{r}_{\mathrm{o}},\omega)j_{\nu}(k\|\boldsymbol{r}-\boldsymbol{r}_{\mathrm{o}}\|)Y_{\nu,\mu}\left(\frac{\boldsymbol{r}-\boldsymbol{r}_{\mathrm{o}}}{\|\boldsymbol{r}-\boldsymbol{r}_{\mathrm{o}}\|}\right), \tag{8}$$

  where $\mathring{u}_{\nu,\mu}$ is the expansion coefficient of the order ν and degree μ, $\boldsymbol{r}_{\mathrm{o}}$ is the expansion center, j_{ν} is the νth-order spherical Bessel function, and $Y_{\nu,\mu}$ is the spherical harmonic function of the order ν and degree μ, accepting as arguments the azimuth and zenith angles of the unit vector $(\boldsymbol{r}-\boldsymbol{r}_{\mathrm{o}})/\|\boldsymbol{r}-\boldsymbol{r}_{\mathrm{o}}\|$.

- Equivalent source distribution (or single-layer potential)

$$u(\boldsymbol{r},\omega)=\int_{\partial\Omega}\breve{u}(\boldsymbol{r}',\omega)\frac{\mathrm{e}^{\mathrm{j}k(\boldsymbol{r}-\boldsymbol{r}')}}{4\pi\|\boldsymbol{r}-\boldsymbol{r}'\|}\mathrm{d}\boldsymbol{r}', \tag{9}$$

  where $\breve{u}$ is the weight of the equivalent source, i.e., the point source, and $\partial\Omega$ is the boundary of Ω. The boundary surface of the equivalent source is not necessarily identical to $\partial\Omega$ if it encloses the region Ω.

Examples of the element solutions are plotted in Fig. 3. By discretizing the integration operation in (7) and (9) or truncating the infinite-dimensional series expansion in (8), and setting the element solutions as $\{\varphi_l(\boldsymbol{x})\}_{l=1}^{L}$ and their weights as $\{\gamma_l\}_{l=1}^{L}$, we can approximate the above representations as the finite-dimensional basis expansions in (6).

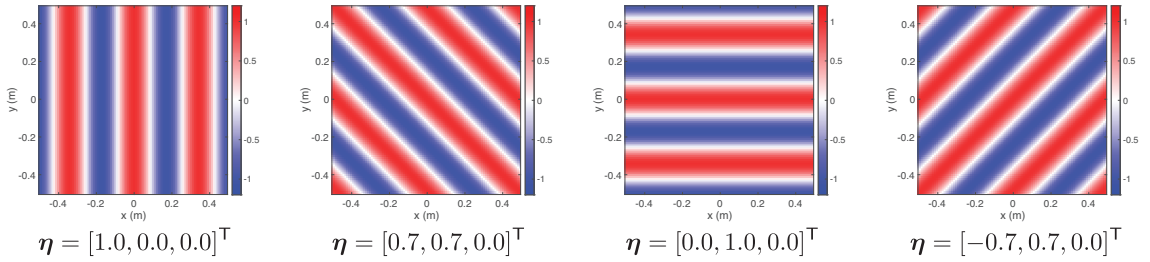
When the squared error loss function (2) and ℓ_2-norm penalty (3) are used, (6) becomes a simple least squares problem (or linear ridge regression). The closed-form solution of $\boldsymbol{\gamma}$ is obtained as

$$\begin{aligned}\hat{\boldsymbol{\gamma}}&=\underset{\boldsymbol{\gamma}\in\mathbb{C}^{L}}{\arg\min}\,\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{\gamma}\|^2+\lambda\|\boldsymbol{\gamma}\|^2\\&=\left(\boldsymbol{\Phi}^{\mathsf{H}}\boldsymbol{\Phi}+\lambda\boldsymbol{I}\right)^{-1}\boldsymbol{\Phi}^{\mathsf{H}}\boldsymbol{y},\end{aligned} \tag{10}$$

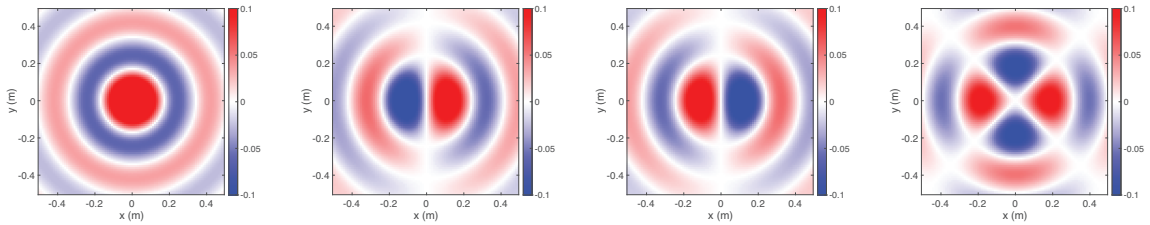
where $\boldsymbol{\Phi}=[\boldsymbol{\varphi}(\boldsymbol{x}_1),\ldots,\boldsymbol{\varphi}(\boldsymbol{x}_I)]^{\mathsf{T}}\in\mathbb{K}^{I\times L}$, $\boldsymbol{I}$ is the identity matrix, and $(\cdot)^{\mathsf{H}}$ is the conjugate transpose (equivalent to the transpose $(\cdot)^{\mathsf{T}}$ for a real-valued matrix). Then, f is estimated using $\hat{\boldsymbol{\gamma}}$ as in (6) under the constraint that f satisfies the governing PDE.

Various regularization terms for $\mathcal{R}$ other than the ℓ_2-norm penalty have been investigated to incorporate prior knowledge on the structure of $\boldsymbol{\gamma}$. For example, the ℓ_1-norm penalty is widely used to promote sparsity on $\boldsymbol{\gamma}$ when $L\gg I$ in the context of sparse optimization and compressed sensing [22]. However, iterative algorithms, such as the fast iterative shrinkage thresholding algorithm and iteratively reweighted least squares, must be used to obtain the solution [23].

When applying the linear ridge regression, how to decide the number of basis functions L, i.e., the number of plane waves (7) or point sources (9) and truncation order in (8), is not a trivial task. For the spherical wave function expansion, using the Ω radius R, $\lceil kR\rceil$ and $\lceil \mathrm{e}kR/2\rceil$ are empirically known to be



(a) Plane wave functions



(b) Spherical wave functions

Fig. 3. Basis functions used to constrain the solution to satisfy the Helmholtz equation on the x–y plane at the frequency of 1000 Hz. (a) Plane wave functions. (b) Spherical wave functions. Real parts of the functions are plotted.

an appropriate truncation order when the target region Ω is a sphere. The sparsity-promoting regularization can help find an admissible number of basis functions using a dictionary matrix ($\boldsymbol{\Phi}$ in (10)) constructed using a large number of basis functions [3]–[5].

The finite-dimensional expansion into the element solutions of the governing PDE is a powerful and flexible technique to embed the physical properties, which has also been applied to the estimation of the exterior field and region including sources or scattering objects. This technique is also generalized to the cases of using directional microphones and estimating expansion coefficients of spherical wave functions [24], [25].

B. Kernel regression using reproducing kernel Hilbert space for governing PDE

Kernel regression is classified as a nonparametric regression technique. The kernel function $\kappa : \mathbb{R}^P \times \mathbb{R}^P \to \mathbb{K}$ should be defined as a similarity function expressed as an inner product defined on some functional space $\mathscr{H}$, $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{\varphi}(\boldsymbol{x}), \boldsymbol{\varphi}(\boldsymbol{x}') \rangle_{\mathscr{H}}$. Thus, $\boldsymbol{\varphi}$ can be infinite-dimensional, which is one of the major differences from the basis expansion described in Section III-A. On the basis of the representer

theorem [26], f is represented by the weighted sum of κ at the sampling points as

$$\begin{aligned} f(\boldsymbol{x};\boldsymbol{\alpha}) &= \sum_{i=1}^{I} \alpha_i \kappa(\boldsymbol{x}, \boldsymbol{x}_i) \\ &= \boldsymbol{\kappa}(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{\alpha}, \end{aligned} \tag{11}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_I]^{\mathsf{T}} \in \mathbb{K}^I$ are the weight coefficients and $\boldsymbol{\kappa}(\boldsymbol{x}) = [\kappa(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, \kappa(\boldsymbol{x}, \boldsymbol{x}_I)]^{\mathsf{T}}$ is the vector of kernel functions. In the kernel ridge regression [26], $\boldsymbol{\alpha}$ is obtained as

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{y}, \tag{12}$$

with the Gram matrix $\boldsymbol{K} \in \mathbb{K}^{I \times I}$ defined as

$$\boldsymbol{K} = \begin{bmatrix} \kappa(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & \kappa(\boldsymbol{x}_1, \boldsymbol{x}_I) \\ \vdots & \ddots & \vdots \\ \kappa(\boldsymbol{x}_I, \boldsymbol{x}_1) & \cdots & \kappa(\boldsymbol{x}_I, \boldsymbol{x}_I) \end{bmatrix}. \tag{13}$$

Then, f is interpolated by substituting $\hat{\boldsymbol{\alpha}}$ into (11).

In the kernel regression, the reproducing kernel Hilbert space to seek a solution must be properly defined, which also defines the reproducing kernel function κ. In general machine learning, some assumptions for the data structure are imposed to define κ. For example, the Gaussian kernel function [26] is frequently used to induce the smoothness of the function to be interpolated. The reproducing kernel Hilbert space $\mathscr{H}$, as well as the kernel function κ, can be defined to constrain the solution to satisfy the Helmholtz equation [27]. The solution space of the homogeneous Helmholtz equation is defined by the plane wave expansion (7) or spherical wave function expansion (8) with square-integrable or square-summable expansion coefficients. We here define the inner product over $\mathscr{H}$ using the plane wave expansion with positive directional weighting $w : \mathbb{S}_2 \to \mathbb{R}_{\geq 0}$ as [28]

$$\langle u_1, u_2 \rangle_{\mathscr{H}} = \frac{1}{4\pi} \int_{\mathbb{S}_2} \frac{1}{w(\boldsymbol{\eta})} \tilde{u}_1(\boldsymbol{\eta})^* \tilde{u}_2(\boldsymbol{\eta}) \mathrm{d}\boldsymbol{\eta}. \tag{14}$$

The directional weighting function w is designed to incorporate prior knowledge of the directivity pattern of the target sound field by setting $\boldsymbol{\xi}$ in the approximate directions of sources or early reflections. We represent w as a unimodal function using the von Mises–Fisher distribution [29] defined by

$$w(\boldsymbol{\eta}) = \frac{1}{C(\beta)} \mathrm{e}^{\beta \langle \boldsymbol{\eta}, \boldsymbol{\xi} \rangle}, \tag{15}$$

where $\boldsymbol{\xi} \in \mathbb{S}_2$ is the peak direction, β is the parameter for controlling the sharpness of w, and $C(\beta)$ is the normalization factor

$$C(\beta) = \begin{cases} 1, & \beta = 0 \\ \frac{\sinh(\beta)}{\beta}, & \text{otherwise} \end{cases}. \tag{16}$$

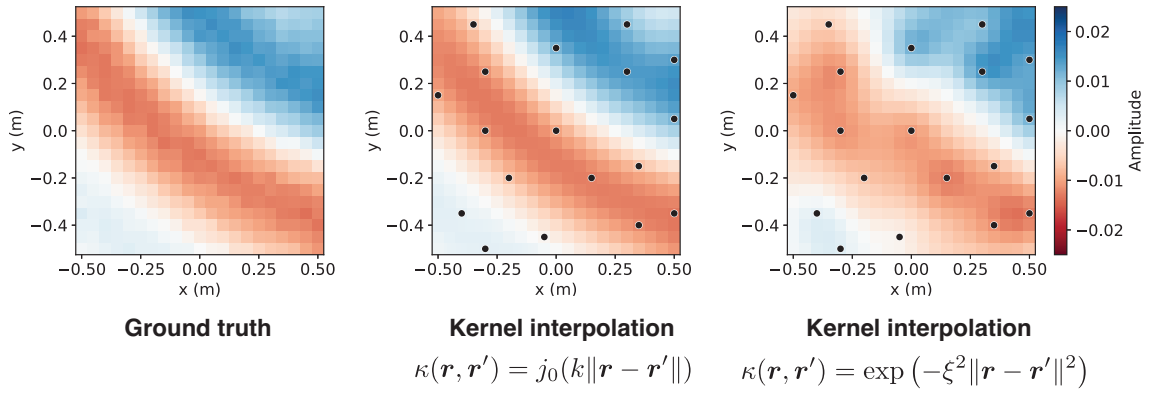


Fig. 4. Pressure distribution on the two-dimensional plane estimated by the kernel method using MeshRIR dataset [30]. Black dots indicate the microphone positions ($M = 18$). The kernel function of the spherical Bessel function (18) and Gaussian kernel ($\xi = 1.0k$) are compared. The source is an ordinary loudspeaker, and the source signal is the pulse signal lowpass-filtered up to 500 Hz. The use of the physics-constrained kernel function improved the reconstruction accuracy. The figure is taken from [30], with an additional comparison with Gaussian kernel.

Then, the kernel function is analytically expressed as

$$\kappa(\boldsymbol{r}, \boldsymbol{r}') = \frac{1}{C(\beta)} j_0\left(\sqrt{(k(\boldsymbol{r} - \boldsymbol{r}') - \mathrm{j}\beta\boldsymbol{\xi})^{\mathsf{T}}\,(k(\boldsymbol{r} - \boldsymbol{r}') - \mathrm{j}\beta\boldsymbol{\xi})}\right). \tag{17}$$

When the directional weighting w is uniform, i.e., no prior information on the directivity, w is set as 1, and the kernel function is simplified as

$$\kappa(\boldsymbol{r}, \boldsymbol{r}') = j_0\left(k\|\boldsymbol{r} - \boldsymbol{r}'\|\right). \tag{18}$$

This function is equivalent to the spatial covariance function of diffused sound fields. Thus, the kernel ridge regression with the constraint of the Helmholtz equation is achieved by using the kernel function defined in (17) or (18). Figure 4 is an example of the experimental comparison of the physics-constrained kernel function (18) and the Gaussian kernel function.

The kernel regression described above is applied to interpolate the pressure distribution from discrete pressure measurements. This technique can be generalized to the cases of using directional microphones and estimating expansion coefficients of spherical wave functions [28], which is regarded as an infinite-dimensional generalization of the finite-dimensional basis expansion method [24].

C. Regression by neural networks incorporating governing PDEs

When applying neural networks to regression problems, the target output is typically a discretized value, which is denoted as $\boldsymbol{t} \in \mathbb{R}^J$, considering the real-valued function f. In a simple manner, $\boldsymbol{t}$ is

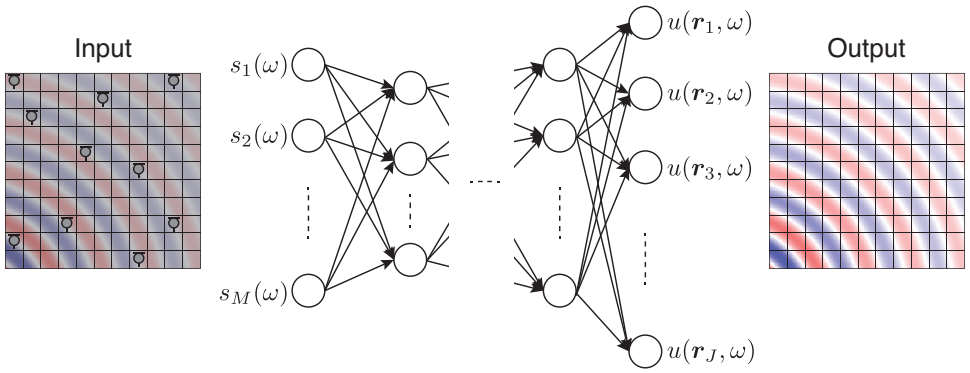


Fig. 5. Feedforward neural network for sound field estimation in the frequency domain. The input is the observation $\boldsymbol{s}$ and the target output is the discretized value of u. The network weights are optimized by using a pair of datasets.

defined as a discretization of the domain into $\{\boldsymbol{x}_j\}_{j=1}^J$ with J ($> I$) sampling points, i.e., $t_j = f(\boldsymbol{x}_j)$. The neural network with input $\boldsymbol{y} \in \mathbb{R}^I$ and output $\boldsymbol{g}(\boldsymbol{y}) \in \mathbb{R}^J$, as shown in Fig. 5, is trained using a pair of datasets for various sampling points and function values $\{(\boldsymbol{y}_d, \boldsymbol{t}_d)\}_{d=1}^D$, where the subscript d is the index of the data and D is the number of training data. There are various possibilities of the network architecture, but its parameters are collectively denoted as $\boldsymbol{\theta}_{\mathrm{NN}}$. The loss function used to train the network, i.e., to obtain $\boldsymbol{\theta}_{\mathrm{NN}}$, is, for example, the squared error written as

$$\mathcal{J}(\boldsymbol{\theta}_{\mathrm{NN}}) = \sum_{d=1}^{D} \|\boldsymbol{t}_d - \boldsymbol{g}(\boldsymbol{y}_d; \boldsymbol{\theta}_{\mathrm{NN}})\|^2. \tag{19}$$

The parameters $\boldsymbol{\theta}_{\mathrm{NN}}$ are optimized basically by minimizing $\mathcal{J}$ using steepest-descent-based algorithms [6]. In the inference, the output for the unknown observation $\boldsymbol{y}$, i.e., $\boldsymbol{g}(\boldsymbol{y})$, is computed by forward propagation, which is regarded as a discretized estimate of f. When estimating complex-valued functions, a simple strategy is to treat real and imaginary values separately. Other techniques are, for example, separating the complex values into magnitudes and phases and using complex-valued neural networks [31].

It is not straightforward to incorporate physical properties into the neural-network-based regression because the output is a discretized value. One of the techniques to constrain the estimate to satisfy the Helmholtz equation is to set the expansion coefficients of element solutions introduced in Section III-A to the target output [18], [32]. By using a pair of datasets $\{(\boldsymbol{y}_d, \boldsymbol{\gamma}_d)\}_{d=1}^D$, we can train the network for predicting expansion coefficients. By using the estimated $\boldsymbol{\gamma}$, we can obtain the continuous function f satisfying the Helmholtz equation. This technique is referred to as the *physics-constrained neural networks (PCNNs)*.

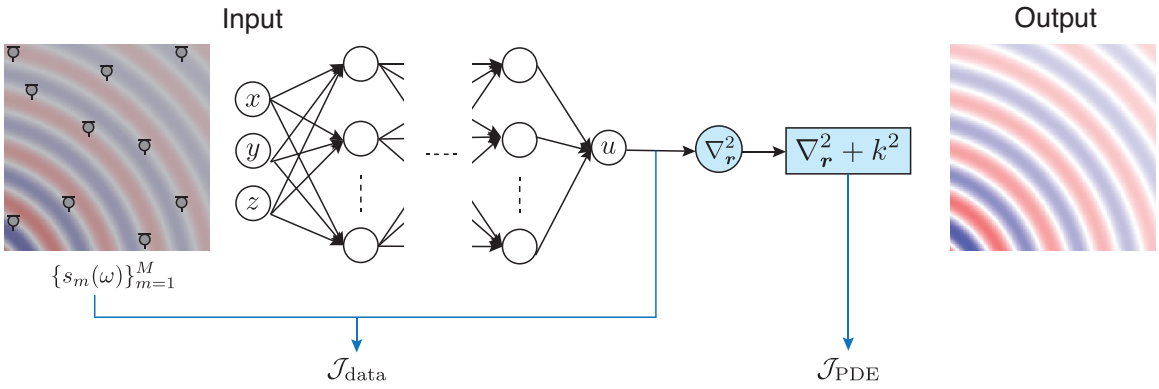


Fig. 6. PINN using implicit neural representation for sound field estimation in the frequency domain. The input is the positional coordinates $\boldsymbol{r} = (x, y, z)$ and the target output is $u(\boldsymbol{r})$. The network weights are optimized by using the observation $\boldsymbol{y}$. In PINNs, the loss function is composed of the data loss $\mathcal{J}_{\mathrm{data}}$ and PDE loss $\mathcal{J}_{\mathrm{PDE}}$ to penalize the deviation from the governing PDE.

D. PINNs based on implicit neural representation

PINNs were first proposed to solve forward and inverse problems containing PDEs by using the implicit neural representation [7], [8]. In the implicit neural representation, neural networks (typically, multilayer perceptions) are used to implicitly represent a continuous function f. The input and output of a neural network are the argument of f, i.e., $\boldsymbol{x} \in \mathbb{R}^P$, and the scalar function value $g(\boldsymbol{x}; \boldsymbol{\theta}_{\mathrm{NN}}) \in \mathbb{R}$ approximating $f(\boldsymbol{x})$, respectively (see Fig. 6). The training data in this scenario are the pair of sampling points and function values $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^I$. The cost function is, for example, the squared error written as

$$\mathcal{J}_{\mathrm{INR}}(\boldsymbol{\theta}_{\mathrm{NN}}) = \sum_{i=1}^{I} |y_i - g(\boldsymbol{x}_i; \boldsymbol{\theta}_{\mathrm{NN}})|^2. \tag{20}$$

Therefore, it is unnecessary to discretize the target output or prepare a pair of datasets because the observation data are regarded as the training data in this context.

A benefit of the implicit neural representation is that some constraints on g including its (partial) derivatives can be included in the loss function:

$$\mathcal{J}_{\mathrm{INR}}(\boldsymbol{\theta}_{\mathrm{NN}}) = \sum_{i=1}^{I} |y_i - g(\boldsymbol{x}_i; \boldsymbol{\theta}_{\mathrm{NN}})|^2 + \epsilon \sum_{n=1}^{N} \left| H\left(g(\boldsymbol{x}_n), \nabla_{\boldsymbol{x}} g(\boldsymbol{x}_n), \nabla_{\boldsymbol{x}}^2 g(\boldsymbol{x}_n), \ldots\right)\right|^2 \tag{21}$$

with the positive constant ϵ and given constraints at $\{\boldsymbol{x}_n\}_{n=1}^N$

$$H\left(g(\boldsymbol{x}_n), \nabla_{\boldsymbol{x}} g(\boldsymbol{x}_n), \nabla_{\boldsymbol{x}}^2 g(\boldsymbol{x}_n), \ldots\right) = 0. \tag{22}$$

Here, $\nabla_{\boldsymbol{x}}$ represents the gradient with respect to $\boldsymbol{x}$. The evaluation points of the constraints $\{\boldsymbol{x}_n\}_{n=1}^N$ can be determined independently from the sampling points $\{\boldsymbol{x}_i\}_{i=1}^I$. The loss function including derivatives is usually calculated using automatic differentiation [33].

Therefore, the physical properties represented by a PDE can be embedded into the loss function to infer the function approximately satisfying the governing PDE. When considering the sound field estimation in the frequency domain, as shown in Fig. 6, the loss function to train the neural network g with the input $\boldsymbol{r} \in \Omega$ is expressed as

$$\mathcal{J}_{\mathrm{PINN}}(\boldsymbol{\theta}_{\mathrm{NN}}) = \mathcal{J}_{\mathrm{data}} + \epsilon \mathcal{J}_{\mathrm{PDE}}, \tag{23}$$

where

$$\mathcal{J}_{\mathrm{data}} = \sum_{m=1}^{M} |s_m(\omega) - g(\boldsymbol{r}_m; \boldsymbol{\theta}_{\mathrm{NN}})|^2, \tag{24}$$

$$\mathcal{J}_{\mathrm{PDE}} = \sum_{n=1}^{N} |(\nabla_{\boldsymbol{r}}^2 + k^2) g(\boldsymbol{r}_n; \boldsymbol{\theta}_{\mathrm{NN}})|^2. \tag{25}$$

Thus, g is trained to minimize the deviation from the Helmholtz equation as well as the observations at the sampling points $\{\boldsymbol{r}_m\}_{m=1}^M$. Since this technique is based on an implicit neural representation, the observed signal $\{s_m(\omega)\}_{m=1}^M$ corresponds to the training data for the neural network. In this sense, PINNs can be classified as an unsupervised approach. Figure 7 shows an experimental example of RIR reconstruction using a neural network and the same architecture trained using the PINN loss (23). It can be seen in Fig. 7 that the neural network can estimate the RIRs, although they contain noise and incoherent wavefronts in the missing channels. In contrast, the result of the PINN is more accurate with reduced noise and coherent wavefronts owing to the adoption of the PDE loss.

To extract the properties of an acoustic environment by using a set of training data, i.e., supervised learning for sound field estimation, the general neural networks in Section III-C can be applied [34]–[37]. Since the output of the neural network is a discretized value, the evaluation of the deviation from the governing PDE is not straightforward. A simple approach is difference approximation by finely discretizing the domain to compute the PDE loss [38]. Another strategy is the use of general interpolation techniques to obtain a continuous function from the output of the neural networks [17]; thus, the PDE loss is computed by using the interpolated function.

IV. Current studies of sound field estimation based on PIML

Current sound field estimation methods based on PIML are essentially based on the techniques introduced in Section III. The linear ridge regression based on finite-dimensional expansion into basis functions has been thoroughly investigated in the literature. In particular, the spherical wave function expansion has been used in spatial audio applications because of its compatibility with spherical microphone arrays [2], [25]. The kernel ridge regression (or infinite-dimensional expansion) with the constraint of the

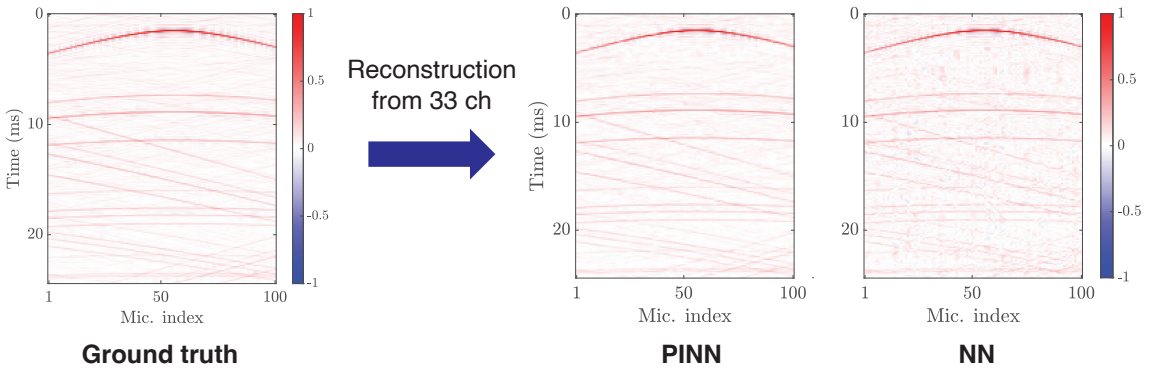


Fig. 7. RIRs (simulated shoebox-shaped room of $T_{60} = 500$ ms) measured by a uniform linear array of 100 microphones are reconstructed using only 33 randomly selected channels. PINN [14] and a plain neural network (NN) are compared. The addition of the physics loss improved the reconstruction accuracy of RIRs. Figures are taken from [39] in which the method proposed in [14] has initially appeared, and further details of the experimental setting are presented.

Helmholtz equation can be regarded as a generalization of linear-ridge-regression-based methods. Since it is applicable to the estimation in a region of an arbitrary shape with arbitrarily placed microphones and the number of parameters to be manually tuned can be reduced, kernel-ridge-regression-based methods are generally simpler and more useful than the linear-ridge-regression-based methods. One of the advantages of these methods is a linear estimator in the frequency domain, which means that the estimation is performed by the convolution of an FIR filter in the time domain. This property is highly important in real-time applications such as active noise control.

Sparsity-based methods have also been intensively investigated over the past ten or more years both in the frequency and time domains [3]–[5]. Sparsity is usually imposed on expansion coefficients of plane waves (7) and equivalent sources (9). Since the sound field is represented by a linear combination of sparse plane waves or equivalent sources, it is guaranteed that the interpolated function satisfies the Helmholtz equation. These techniques are effective when the sound field to be estimated fits the sparsity assumption, such as a less reverberant sound field. However, iterative algorithms are required to obtain a solution; thus, the estimator is nonlinear, and STFT-based processing is necessary in practice. The sparse plane wave model is also used in hierarchical-Bayesian-based Gaussian process regression, which can be regarded as a generalization of the kernel ridge regression [40]. Then, the inference process becomes a linear operation.

Neural-network-based sound field estimation methods have gained interest in recent years, which makes the inference more adaptive to the environment owing to the high representational power of neural networks [34]–[37], [41], [42]. Physical properties have recently been incorporated into neural networks

to mitigate physically unnatural distortions even with a small number of microphones. PINNs for the sound field estimation problem are investigated both in the frequency [12], [13], [43] and time domains [14], [15]. A major difference from the methods based on basis expansions and kernels is that the estimated function does not strictly satisfy the governing PDEs, i.e., the Helmholtz and wave equations, because its constraint is imposed by the loss function $\mathcal{J}_{\mathrm{PDE}}$ as in (23). The governing PDEs are strictly satisfied by using PCNN, i.e., setting the target output as expansion coefficients of the basis expansions [18], [32]; however, this approach can sacrifice the high representational power of neural networks by strongly restricting the solution space. Whereas the methods based on implicit neural representation are regarded as unsupervised learning, supervised-learning-based methods may have the capability to capture properties of acoustic environments and high accuracy even when the number of available microphones is particularly small. In the physics-informed convolutional neural network [17], the PDE loss is computed by using bicubic spline interpolation. In [16], a loss function based on the boundary integral equation of the Helmholtz equation, called the Kirchhoff–Helmholtz integral equation [1], is proposed in the context of acoustic imaging using a planar microphone array, which is also regarded as a supervised-learning approach.

The inference process of neural networks involves multiple linear and nonlinear transformations; therefore, if a network is designed in the frequency domain, it cannot be implemented as an FIR filter in the time domain. The kernel-ridge-regression-based method using the kernel function adapted to acoustic environments using neural networks enables the linear operation of inference while preserving the constraint of the Helmholtz equation and the high flexibility of neural networks. In [19], [43], the directional weight w in (14) is separated into the directed and residual components. The directed component is modeled by a weighted sum of von Mises–Fisher distributions (15) and the residual component is represented by a neural network to separately represent highly directive sparse sound waves and spatially complex late reverberation. The parameters of these components are jointly optimized by a steepest-descent-based algorithm. Experimental results in the frequency domain are demonstrated in Fig. 8, where the kernel method using (18), a plain neural network, PINN, PCNN with plane wave decomposition, and the adaptive kernel method proposed in [19], [43] are compared by numerical simulation. It can be observed that the adaptive kernel method performs well over the investigated frequency range.

As described above, various attempts have been made to incorporate physical properties into function interpolation and regression techniques for sound field estimation. The expansion representation introduced in Section III-A leads to the closed-form linear estimator in the frequency domain based on the linear and kernel ridge regression techniques. In exchange for their simple implementation and

TABLE I

COMPARISON OF CURRENT SOUND FIELD ESTIMATION METHODS BASED ON PIML USING NEURAL NETWORKS WITH TRAINING DATA REQUIRED OR NOT REQUIRED, LINEAR OR NONLINEAR ESTIMATOR, TIME OR FREQUENCY DOMAIN, AND PENALIZED OR CONSTRAINED PHYSICAL PROPERTY.

	Supervised/unsupervised	Estimator	Domain	Physical property
Shigemi et al. 2022 [17]	Supervised	Nonlinear	Frequency	Penalized
Karakonstantis et al. 2023 [18] and Lobato et al. 2024 [32]	Supervised	Linear	Frequency	Constrained
Chen et al. 2023 [12] and Ma et al. [13]	Unsupervised	Nonlinear	Frequency	Penalized
Olivieri et al. 2024 [14] and Karakonstantis et al. 2024 [15]	Unsupervised	Nonlinear	Time	Penalized
Ribeiro et al. 2023 [19], [43]	Unsupervised	Linear	Frequency	Constrained

small computational cost, these techniques do not have sufficient adaptability to the acoustic environment because of their fixed estimator. Methods for adaptive estimators, e.g., using sparse optimization and kernel learning techniques, have also been studied. Recent neural-network-based methods having both high flexibility and physical validity have been investigated for both supervised- and unsupervised-learning approaches. Many of them make the estimator nonlinear; thus, the inference process for each time or time–frequency data becomes the forward propagation of the neural network for the supervised methods and both the training and inference of the neural network for the unsupervised methods; this process requires a high computational load. However, several methods integrating the expansion representation and neural networks enable estimators to be linear, which will be suitable for real-time applications. The current PIML-based sound field estimation methods using neural networks are compared in Table I.

V. OUTLOOK

PIML for sound field estimation is a rapidly growing research field and has successfully been applied to some application scenarios. However, the current methods still have some limitations. In particular, the major problems that should be resolved are 1) the preparation of training data, 2) the mismatch between training and test data, and 3) the neural network architecture design.

The methods using only the observation data, i.e., unsupervised methods, are useful when the acoustic environment to be measured is unpredictable. Supervised methods have the potential to extract properties of an acoustic environment by using a set of training data and to improve estimation accuracy when

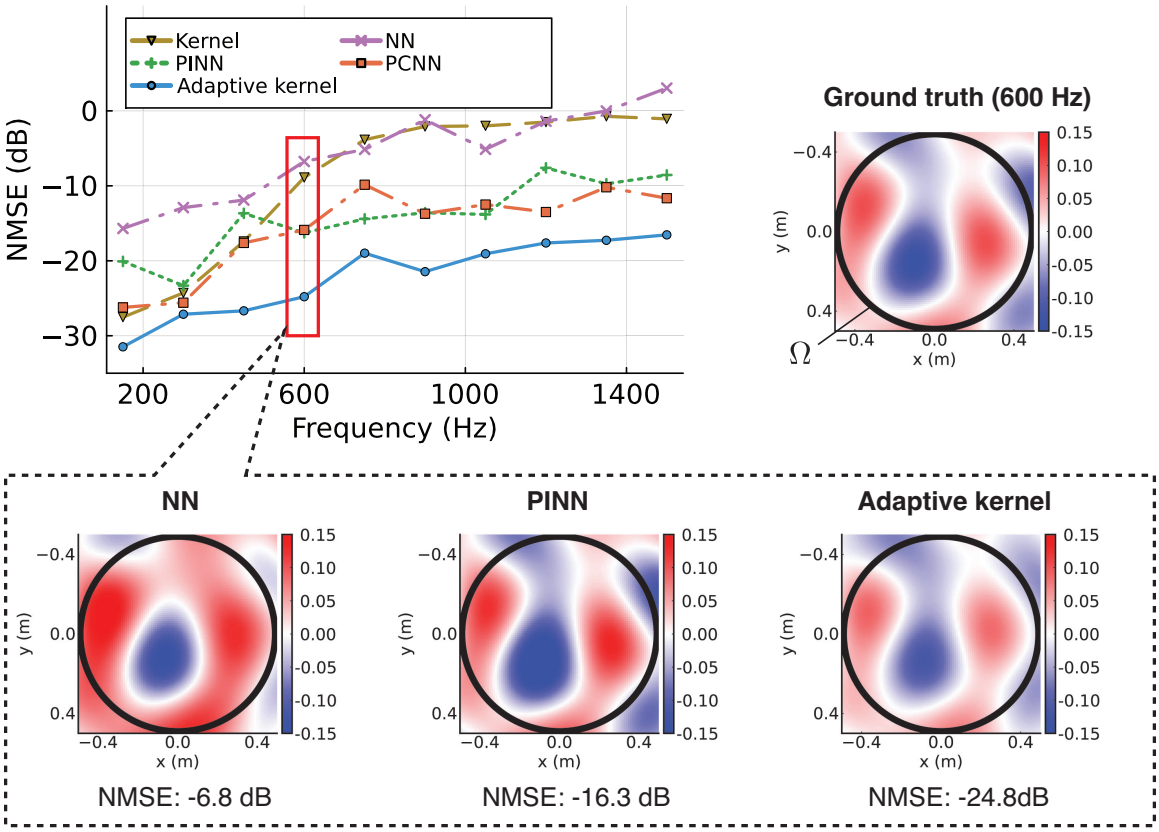


Fig. 8. Normalized mean square error (NMSE) and reconstructed pressure distribution inside a sphere of the target region in the frequency domain. A single point source was located outside the target region, and RIRs in a shoebox-shaped room were simulated by the image source method so that the reverberation time T_{60} becomes 400 ms. $M = 41$ microphones were placed on two spherical surfaces with radii of 0.50 and 0.49 m. The kernel method using (18), the neural network without any physics (NN), PINN, PCNN with plane wave decomposition, and the adaptive kernel method proposed in [19], [43] are compared. The number of evaluation points of the PDE loss in PINN was $N = 56$. The architectures of NN, PINN, PCNN, and the adaptive kernel are basically the same (fully connected layers with 3 entries for input, 3 hidden layers with 8 neurons for the first, 8 neurons for neural ordinary differential equation (NODE) layer [44], and 5 neurons for the third hidden layer, and 1 output neuron. The rowdy activation function [45] is used for the first three layers and ReLU for the last layer on PCNN and adaptive kernel to guarantee nonnegativity.). Further details of the experimental setting are presented in [43].

the number of microphones is particularly small. However, the training data should be a set of sound field data with high spatial (or spatiotemporal) resolution. Since the acoustic environment can have large variations of, e.g., source positions, directivity, and reflections, the first problem is how to prepare the training data, even though physical properties could help prevent overfitting in the case of a small number of training data. Currently, it is considered impractical to collect all the variations of sound field data by actual measurements. Therefore, it will be necessary to integrate multiple training datasets measured in

different environments and also include some simulated data. Thus, the second problem of the mismatch between training and test data arises. The measurement systems for acoustic fields could have some variations. Even for the wave-based acoustic simulation techniques requiring a high computational load, there are some deviations between real and simulated data. Furthermore, the governing PDEs still include given model parameters, e.g., sound speed, which can be different from those in the training data. It is still unclear how these deviations can affect the estimation performance and how to mitigate those effects. The third problem is generally common to neural-network-based techniques, but there is currently no clear methodology for architectural design. Therefore, at present, architectures must be designed by trial and error depending on the task and data.

In addition to the problems listed above, there are several other unsolved issues. First, the target region of estimation is usually assumed not to include any sound sources or scattering objects. This assumption allows us to represent the interior sound field by basis functions introduced in Section III-A. However, in some applications, the target region is not necessarily source-free; thus, the governing PDE of the sound field to be estimated can become inhomogeneous. Several methods have been proposed for estimating a sound field including sound sources or scatterers. Furthermore, the methods using the PDE loss, such as PINN, can also be applied in this setting. However, further investigations of these methods in practical applications are still necessary.

Second, the estimation accuracy of the sound field estimation methods is highly dependent on the number of microphones and their placement. It is often necessary to cover the entire target region with the smallest possible number of microphones. On the basis of boundary integral equations, such as Kirchhoff–Helmholtz integrals, a sound field inside a source-free interior region can be predicted by using pressure and its normal derivative on its surface. However, the interior sound field cannot be uniquely determined at some frequencies solely from the pressure measurements on the boundary surface, which is called the *forbidden frequency problem* [1]. Although it is necessary to place microphones inside the target region, its optimal placement is not straightforward, especially for broadband measurements. Several techniques for the optimal placement of microphones for sound field estimation have been proposed; however, their effectiveness has not been sufficiently evaluated in practical applications. It may also be possible to incorporate auditory properties and/or multimodal information to reduce the number of microphones.

The PIML-based sound field estimation techniques, particularly linear-ridge-regression and kernel-ridge-regression-based methods, have already been applied to various applications, e.g., the spatial audio recording for binaural reproduction [46], [47], spatial interpolation of HRTFs [48], and spatial active noise control [49], [50], owing to their simple implementation and small computational cost. In binaural reproduction, the signals reaching both ears, when a listener is present in the target region, are reproduced

by estimating the expansion coefficients of the sound field at the listening position. The HRTFs are also required in the binaural reproduction, but measuring them for each listener is costly. The interpolation of HRTFs can be formulated as an exterior sound field estimation problem based on reciprocity. Active noise control in a spatial region requires estimating a sound field of primary noise using microphones in a target region and synthesizing an anti-noise sound field using loudspeakers in real time. On the other hand, applications of neural-network-based methods are still limited to offline applications, e.g., acoustic imaging, owing to the above-mentioned challenges as well as the issue of relatively high computational cost. It is expected that these methods will also be applied to various applications in the future because of their high flexibility and interpolation capability.

VI. Conclusion

PIML-based sound field estimation methods are overviewed. Our focus was the interior problem, but several techniques can also be applied to the exterior problem and estimation in a region containing some sources or scattering objects. There exist several techniques that can be used to embed physical properties in interpolation techniques. In particular, the governing PDE of the sound field is considered useful prior information. Neural-network-based methods, which have been successfully applied in various domains, are expected to perform well for sound field estimation owing to their high expressive power. How to integrate physical prior knowledge with the high expressive power of neural networks will continue to be one of the most important topics in sound field estimation and its applications. Although there are still many unresolved issues, such as the preparation of training data, we expect further development of sound field estimation methods and their applications in various applied technologies in the future.

References

[1] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[2] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer, 2015.

[3] N. Bertin, L. Daudet, V. Emiya, and R. Gribonval, "Compressive sensing in acoustic imaging," in *Compressed Sensing and its Applications*, H. Boche, R. Calderbank, G. Kutyniok, and J. Vybiral, Eds. Springer, 2015.

[4] N. Antonello, E. De Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 10, pp. 1929–1941, 2017.

[5] N. Murata, S. Koyama, N. Takamune, and H. Saruwatari, "Sparse representation using multidimensional mixed-norm penalty with application to sound field decomposition," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3327–3338, 2018.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[7] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikari, S. Wang, and L. Yang, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, p. 422–440, 2021.

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

[9] S. Cuomo, V. S. D. Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli, "Scientific machine learning through physics–informed neural networks: Where we are and what's next," *J. Sci. Comput.*, vol. 92, 2022.

[10] Y. Lin, J. Theiler, and B. Wohlberg, "Physics-guided data-driven seismic inversion: Recent progress and future opportunities in full-waveform inversion," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 115–133, 2023.

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

[12] X. Chen, F. Ma, A. Bastine, P. Samarasinghe, and H. Sun, "Sound field estimation around a rigid sphere with physics-informed neural network," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, 2023, pp. 1984–1989.

[13] F. Ma, S. Zhao, and I. S. Burnett, "Sound field reconstruction using a compact acoustics-informed neural network," arXiv:2402.08904, 2024.

[14] M. Olivieri, X. Karakonstantis, M. Pezzoli, F. Antonacci, A. Sarti, and E. Fernandez-Grande, "Physics-informed neural network for volumetric sound field reconstruction of speech signals," *EURASIP J. Audio, Speech, Music Process.*, 2024, (in press).

[15] X. Karakonstantis, D. Caviedes-Nozal, A. Richard, and E. Fernandez-Grande, "Room impulse response reconstruction with physics-informed deep learning," *J. Acoust. Soc. Amer.*, vol. 155, no. 2, p. 1048–1059, 2024.

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

[17] K. Shigemi, S. Koyama, T. Nakamura, and H. Saruwatari, "Physics-informed convolutional neural network with bicubic spline interpolation for sound field estimation," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Sep. 2022.

[18] X. Karakonstantis and E. Fernandez-Grande, "Generative adversarial networks with physical sound field priors," *J. Acoust. Soc. Amer.*, vol. 154, no. 2, pp. 1226–1238, 2023.

[19] J. G. C. Ribeiro, S. Koyama, and H. Saruwatari, "Kernel interpolation of acoustic transfer functions with adaptive kernel for directed and residual reverberations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Rhodes island, Greece, Jun. 2023.

[20] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer, 2019.

[21] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, 2013.

[22] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[23] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[24] A. Laborie, R. Bruno, and S. Montoya, "A new comprehensive approach of surround sound recording," in *Proc. 114th AES Conv.*, Amsterdam, Netherlands, 2003.

[25] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, 2005.

[26] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[27] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 135–139, 2018.

[28] ——, "Directionally weighted wave field estimation exploiting prior information on source direction," *IEEE Trans. Signal Process.*, vol. 69, pp. 2383–2395, 2021.

[29] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons, 2009.

[30] S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnström, "MeshRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2021, pp. 151–155.

[31] A. Hirose, *Complex-Valued Neural Networks*. Springer-Verlag, 2012.

[32] T. Lobato, R. Sottek, and M. Vorländer, "Using learned priors to regularize the Helmholtz equation least-squares method," *J. Acoust. Soc. Amer.*, vol. 155, no. 2, p. 971–983, 2024.

[33] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: a survey," *J. Mach. Learn. Res.*, pp. 1–43, 2018.

[34] F. Lluís, P. Martínez-Nuevo, M. B. Møller, and S. E. Shepstone, "Sound field reconstruction in rooms: Inpainting meets super-resolution," *J. Acoust. Soc. Amer.*, vol. 148, no. 2, 2020.

[35] M. Pezzoli, D. Perini, A. Bernardini, F. Borra, F. Antonacci, and A. Sarti, "Deep prior approach for room impulse response reconstruction," *Sensors*, vol. 22, no. 7, 2022.

[36] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, "Generative models for sound field reconstruction," *J. Acoust. Soc. Amer.*, vol. 153, no. 2, p. 1179–1190, 2023.

[37] F. Miotello, L. Comanducci, M. Pezzoli, A. Bernardini, F. Antonacci, and A. Sarti, "Reconstruction of sound field through diffusion models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Seoul, Republic of Korea, Apr. 2024.

[38] X. Zhao, Z. Gong, Y. Zhang, W. Yao, and X. Chen, "Physics-informed convolutional neural networks for temperature field prediction of heat source layout without labeled data," *Eng. Appl. Artif. Intell.*, vol. 117, p. 105516, 2023.

[39] M. Pezzoli, F. Antonacci, and A. Sarti, "Implicit neural representation with physics-informed neural networks for the reconstruction of the early part of room impulse responses," in *Proc. Forum Acusticum*, Sep. 2023.

[40] D. Caviedes-Nozal, N. A. B. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," *J. Acoust. Soc. Amer.*, vol. 149, no. 2, 2021.

[41] A. Luo, Y. Du, M. J. Tarr, J. B. Tenenbaum, A. Torralba, and C. Gan, "Learning neural acoustic fields," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2022.

[42] Z. Liang, W. Zhang, and T. D. Abhayapala, "Sound field reconstruction using neural processes with dynamic kernels," *EURASIP J. Audio, Speech, Music Process.*, vol. 13, 2024.

[43] J. G. C. Ribeiro, S. Koyama, and H. Saruwatari, "Sound field estimation based on physics-constrained kernel interpolation adapted to environment," TechRxiv, 2023.

[44] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2018.

[45] A. D. Jagtap, Y. Shin, K. Kawaguchi, and G. E. Karniadakis, "Deep kronecker neural networks: A general framework for neural networks with adaptive activation functions," *Neurocomputing*, vol. 468, pp. 165–180, 2022.

[46] N. Iijima, S. Koyama, and H. Saruwatari, "Binaural rendering from microphone array signals of arbitrary geometry," *J. Acoust. Soc. Amer.*, vol. 150, no. 4, pp. 2479–2491, 2021.

[47] L. Birnie, T. Abhayapala, V. Tourbabin, and P. Samarasinghe, "Mixed source sound field translation for virtual binaural application with perceptual validation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 1188–1203, 2021.

[48] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of hrtfs [head related transfer functions]," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2004, pp. IV, 45–48.

[49] J. Zhang, T. D. Abhayapala, W. Zhang, P. N. Samarasinghe, and S. Jiang, "Active noise control over space: A wave domain approach," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 4, pp. 774–786, 2018.

[50] S. Koyama, J. Brunnström, H. Ito, N. Ueno, and H. Saruwatari, "Spatial active noise control based on kernel interpolation of sound field," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 3052–3063, 2021.

Shoichi Koyama (koyama.shoichi@ieee.org) received the B.E., M.S, and Ph.D. degrees from the University of Tokyo, Tokyo, Japan, in 2007, 2009, and 2014, respectively. He is currently an Associate Professor at the National Institute of Informatics (NII), Tokyo, Japan. Prior to joining NII, he was a Researcher at Nippon Telegraph and Telephone Corporation (2009–2014), and Research Associate (2014–2018) and Lecturer (2018–2023) at the University of Tokyo, Tokyo, Japan. He was also a Visiting Researcher at Paris Diderot University (Paris 7), Institut Langevin, Paris, France (2016–2018), and a Visiting Associate Professor at Research Institute of Electrical Communication, Tohoku University, Miyagi, Japan (2020–2023). His research interests include audio signal processing/machine learning, acoustic inverse problems, and spatial audio.

Juliano G. C. Ribeiro Juliano G. C. Ribeiro received his B.E. degree in electronic engineering from the Instituto Militar de Engenharia (IME), Rio de Janeiro, Brazil, in 2017, and the M.S. and Ph.D. degrees in information science and technology from the University of Tokyo, Tokyo, Japan, in 2021 and 2024, respectively. Currently a Research Engineer with Yamaha corporation. His research interests include spatial audio, acoustic signal processing and physics-informed machine learning.

Tomohiko Nakamura (tomohiko.nakamura.jp@ieee.org) received B.S., M.S., and Ph.D. degrees from the University of Tokyo, Japan, in 2011, 2013, and 2016, respectively. He joined SECOM Intelligent Systems Laboratory as a researcher in 2016 and moved to the University of Tokyo as Project Research Associate in 2019. He is currently a Senior Researcher with the National Institute of Advanced Industrial Science and Technology (AIST). His research interests include signal-processing-inspired deep learning, audio signal processing, and music signal processing.

Natsuki Ueno (natsuki.ueno@ieee.org) received the B.E. degree in engineering from Kyoto University, Kyoto, Japan, in 2016, and the M.S. and Ph.D. degrees in information science and technology from the University of Tokyo, Tokyo, Japan, in 2018 and 2021, respectively. He is currently an Associate professor at Kumamoto University, Kumamoto, Japan. His research interests include spatial audio and acoustic signal processing.

Mirco Pezzoli (mirco.pezzoli@polimi.it) received the M.S. degree (cum laude), in 2017, in computer engineering from the Politecnico di Milano, Italy. In 2021 he received the Ph.D. degree in information engineering at Politecnico di Milano. After two years as a Postdoctoral Researcher, he joined the Department of Electronics, Information and Bioengineering of the Politecnico di Milano as Junior Assistant Professor. His main research interests are multichannel audio signal processing, sound field reconstruction and musical acoustics.