Spatial Upsampling of Head-Related Transfer Functions Using a Physics-Informed Neural Network

Fei Ma, *Member, IEEE,* Thushara D. Abhayapala, *Senior Member, IEEE,*
Prasanga N. Samarasinghe, *Senior Member, IEEE* and Xingyu Chen, *Member, IEEE*

***Abstract*—Head-related transfer function (HRTF) capture the information that a person uses to localize sound sources in space, and thus is crucial for creating authentic virtual acoustic experiences. However, practical HRTF measurement systems may only measure a person's HRTFs sparsely, and this necessitates HRTF upsampling. This paper proposes a physics-informed neural network (PINN) method for HRTF upsampling. The PINN exploits the Helmholtz equation, the governing equation of acoustic wave propagation, for regularizing the upsampling process. This helps the generation of physically valid upsamplings which generalize beyond the measured HRTF. Furthermore, the size (width and depth) of the PINN is set according to the Helmholtz equation and its solutions, the spherical harmonics (SHs). This makes the PINN to have an appropriate level of expressive power and thus does not suffer from the over-fitting problem. Since the PINN is designed independent of any specific HRTF dataset, it offers more generalizability compared to pure data-driven methods. Numerical experiments confirm the better performance of the PINN method for HRTF upsampling in both interpolation and extrapolation scenarios in comparison with the SH method and the HRTF field method.**

***Index Terms*—Head-related transfer function (HRTF), physics-informed neural network (PINN), spherical harmonics, spatial audio, virtual acoustics.**

I. Introduction

HEAD-related transfer function (HRTF) is defined as the ratio between the sound pressure at a point in the ear canal and the sound pressure at the origin with the head being absent [1]. HRTF characterizes the scattering effect of a person's torso, head, and ears with respect to the direction of sound [1], and contains the information that a person uses to localize sound sources in space. Spatial audio and virtual acoustic systems rely on the knowledge of HRTF to reproduce personalized acoustic experience [2].

However, the dependence of HRTF on a person's anatomy makes HRTF highly individual, and thus accurate measurement of HRTF over a large number of directions is desirable for creating an authentic acoustic experience [1]. Nonetheless, a complete measurement of a person's HRTF is both time-consuming and expensive [1]. Practical HRTF measurement systems may only conduct the measurement over a limited number of directions due to the inconvenience of arranging loudspeakers over a whole sphere [1] or to reduce the measurement time, resulting in spatially sparse HRTF datasets. (Although, fast and continuous measurement systems can alleviate the time constraint [3], [4], the high cost of these systems and the necessary anechoic chambers make them inaccessible to most people.) Spatially sparse HRTF datasets can compromise source localization in virtual acoustic environments [5], [6], prompting researchers to upsample them into spatially dense HRTF datasets.

HRTF upsampling consists of two scenarios: interpolation and extrapolation. [1] For the interpolation scenario, HRTF is measured over a limited number of directions to reduce the measurement time, and the aim is to estimate the unknown HRTF whose direction is between those of the measured ones. Early works on interpolation were mainly based on the expansion of HRTF into some linear functions, such as spherical harmonics (SHs) [11]–[13], principle components [14]–[16], spline functions [17], and wavelet functions [18]. Recent works, on the other hand, are mainly based on non-linear modeling with neural networks (NNs) such as auto encoder [19]–[21], generative adversarial networks [22], [23], feature-wise linear modulation [24], convolutional neural network [25], and neural field [26].

To simulate sound from downstairs or the sound of foot-steps, we need HRTF for low evaluation angles. However, the presence of people within the HRTF measurement system makes the measurement at low evaluation angles difficult. This results in the HRTF extrapolation scenario, and the aim is to estimate the unknown HRTF whose direction is beyond those of the measured ones. Although HRTF extrapolation is a task of must, it is often overlooked by existing research. Up to data, there are only a few related works. Zhang *et al.* developed iterative methods [27], [28], which successively estimate the unknown HRTF for missing directions. The methods successfully recover a low order HRTF over a full sphere with one quarter of data missing [27]. Duraiswami *et al.* proposed a regularized SH method [11] which estimates the unknown HRTF at the expense of reduced accuracy in

Source code https://github.com/feima1024/PINN-for-HRTF-upsampling
This work is sponsored by the Australian Research Council (ARC) Discovery Projects funding schemes with project numbers DP200100693.

Fei Ma, Thushara D. Abhayapala, Prasanga N. Samarasinghe, and Xingyu Chen, are all with the Audio and Acoustic Signal Processing Group, College of Engineering, Computing & Cybernetics, The Australian National University, Canberra, ACT 2601, Australia (e-mail: feima.0011@gmail.com, {Thushara.Abhayapala, prasanga.samarasinghe,xingyu.chen1}@anu.edu.au).

Manuscript received xxx xxx, xxx; revised xxx xx, xxx.

[1] This paper focuses on direction related HRTF upsampling. Distance related HRTF upsampling [7]–[10] is not addressed.

representing the measured HRTF. Ahrens *et al.* proposed a non-regularized SH method [29] which estimates the unknown HRTF based on a low-order least-square fit to the measured HRTF and estimations of the unknown HRTF.

There are two limitations with above mentioned upsampling methods. First, most of the conventional linear function expansion methods, such as SH methods [11], [27], had a limited exploitation of additional information in the upsampling process. Their upsamplings are essentially transformations of the information that is contained in the measured HRTF, and thus their performance is constrained by the diversity of the measured HRTF. It was found that by exploiting the scattering field of a rigid sphere, the performance of SH methods can be improved [30]. Second, recent NN based methods [19]–[26], which try to build up implicit associations between HRTF with additional information (such as human anatomy and ear geometry), are dataset dependent [26]. This makes it difficult for them to extrapolate beyond the training data.

In recognition of these limitations, we adopt an HRTF upsampling strategy based on the physics-informed neural network (PINN) [31]–[34]. PINN is one kind of NNs which integrate physical knowledge, i.e., the governing partial differential equation (PDE) of a physical phenomenon, into its architecture [31]–[34]. The physical knowledge helps a PINN to model the physical phenomenon besides physical quantities. Since the seminal works of Raissi and his colleagues [31], [32], PINNs have been successfully applied in many areas such as earthquake modeling [35], [36], propeller-noise prediction [37], and wave-field (sound-field) modeling [38]–[42].

Owing to the principle of acoustic reciprocity [1], HRTF can be regarded as the sound-field generated by a source placed inside of the ear canal. This sound-field together with other sound-fields obey the Helmholtz equation, the governing PDE of acoustic wave propagation [43].

This fact inspires us to develop a PINN method for HRTF upsampling, and we inform the method with physical knowledge from two aspects. First, a modified form of the Helmholtz equation is used as the PDE loss. This helps the PINN to generate physically valid upsamplings which generalize beyond the training data, and relieve the burden of balancing the PDE loss and the data loss with additional parameters.

Second, we set the size of the PINN method according to the SH decomposition of HRTF and the Helmholtz equation. Although the PDE loss helps the PINN method to generate physically valid output, it also prompts the output to be zero [44]–[46], which is a valid but trivial solution of the Helmholtz equation [44]. This problem can be mitigated with improved gradient updating strategy [44], [45]. Nonetheless, the strategy [44], [45] is computationally expensive and requires expert knowledge to tune additional parameters. We found that this problem is due to over-fitting, i.e., over-parameterization of PINN methods. On recognition of the over-fitting, we design the PINN method with an appropriate level of expressive power by exploiting the solution of the Helmholtz equation in spherical coordinates, the SHs. Specifically, we set the PINN method width (the number of neurons in each hidden layers) as half of the dimensionality of HRTFs under SH decomposition [47]–[49], and the depth of it (the number of hidden layers) as three. This setup separates the proposed PINN method apart from general PINN methods in other works that suffer from the over-fitting problem due to inappropriate design of the network [44]–[46].

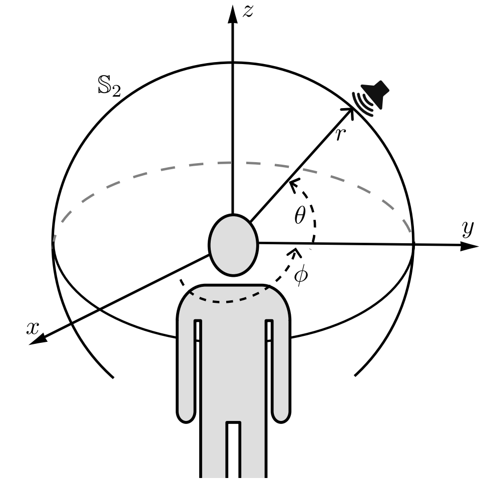


Fig. 1. Layout of a typical HRTF measurement system, which measures the HRTF between a loudspeaker placed on the sphere $\mathbb{S}_2$ and a microphones placed at a specific position inside of the person's ear, ideally close to the ear drum. The person is facing the positive x-axis. HRTF is measured from various directions by rotating either the loudspeaker or the person.

The effective exploitation of the data-independent Helmholtz equation and the SHs compensates for the lack of measured data, and grants the PINN method with extrapolation ability. The performance of the PINN method for upsampling HRTF in both interpolation and extrapolation scenarios are confirmed by numerical experiments, and compared with the SH method [11] and the HRTF field method [26].

The rest of this paper is organized as follows. The problem of interest is introduced in Sec. II. We review the SH method [11] in Sec. III and propose a PINN method in Sec. IV. In Secs. V and VI, we compare the performance of the PINN method, the SH method [11], the standard NN method, and the HRTF field method [26] using interpolation and extrapolation experiments, respectively. Section VII discusses the experiment results, points out directions of improvement, and presents limitations of the PINN method. Section VIII concludes this paper.

II. Problem formulation

Figure 1 presents the layout of a typical HRTF measurement system [1]. Let (x, y, z) and (r, θ, ϕ) denote the Cartesian coordinates and the spherical coordinates of a point with respect to the center of a person's ears. We denote HRTF as $P(\omega, r, \theta, \phi)$ in spherical coordinates or as $P(\omega, x, y, z)$ in Cartesian coordinates, where $\omega = 2\pi f$ is the angular frequency and f is the frequency. Hereafter, HRTF is evaluated on a single sphere, and thus we skip the sphere radius r when representing HRTF and related acoustic quantities for notational simplicity.

As shown in Fig. 1, due to the presence of the person's body, the measurement system can not measure HRTF for low elevation angles, i.e., $\theta < -60°$. To reduce the measurement time, the system may only measure HRTF over a limited number of directions. Both of these two scenarios will result in spatially sparse HRTF datasets, which may be insufficient for virtual acoustic applications [5], [6].

Owing to the principle of acoustic reciprocity [1], HRTF can be regarded as a sound-field [47]. This sound-field, as well as other sound-fields, obeys the governing PDE of acoustic wave propagation, i.e., the Helmholtz equation, [43]

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

where c is the speed of sound, and ∇^2 denotes the Laplacian operator. In spherical coordinates, the Laplacian is given by [43]

$$\nabla^2 P = \frac{2}{r}\frac{\partial P}{\partial r} + \frac{\partial^2 P}{\partial r^2} + \frac{\cos\theta}{r^2 \sin\theta}\frac{\partial P}{\partial \theta} + \frac{1}{r^2}\frac{\partial^2 P}{\partial \theta^2} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 P}{\partial \phi^2}, \tag{2}$$

and in Cartesian coordinates it is given by

$$\nabla^2 P = \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} + \frac{\partial^2 P}{\partial z^2}. \tag{3}$$

In this paper, by exploiting the Helmholtz equation Eq. (1) and its solution in spherical coordinates, the SHs, we aim to upsample a spatially sparse HRTF dataset $\{P(\omega, \theta_q, \phi_q)\}_{q=1}^{Q}$ or equivalently $\{P(\omega, x_q, y_q, z_q)\}_{q=1}^{Q}$ into a spatially dense HRTF dataset. (Q is the number of sampling points and q is the index of a particular sampling point.)

III. Spherical-harmonics-based method

In this section, we first briefly present the SH decomposition of HRTF and then review the regularized SH method [11] for HRTF upsampling. HRTF is expressed in spherical coordinates for the ease of SH decomposition.

HRTF can be decomposed into SHs as [43]

$$\mathbf{P} \approx \mathbf{YA}, \tag{4}$$

where $\mathbf{P} = [P(\omega, \theta_1, \phi_1), P(\omega, \theta_2, \phi_2), ..., P(\omega, \theta_Q, \phi_Q)]^\intercal$ denote the measured HRTF for directions $(\theta_q, \phi_q)_{q=1}^{Q}$, $(\cdot)^\intercal$ is the transpose operation, $\mathbf{A} = [A_{0,0}(\omega), A_{1,-1}(\omega), ..., A_{U,U}(\omega)]^\intercal$ denote the SH coefficients, and

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\theta_1, \phi_1) & Y_1^{-1}(\theta_1, \phi_1) & \dots & Y_U^U(\theta_1, \phi_1) \\ Y_0^0(\theta_2, \phi_2) & Y_1^{-1}(\theta_2, \phi_2) & \dots & Y_U^U(\theta_2, \phi_2) \\ \dots & \dots & \dots & \dots \\ Y_0^0(\theta_Q, \phi_Q) & Y_1^{-1}(\theta_Q, \phi_Q) & \dots & Y_U^U(\theta_Q, \phi_Q) \end{bmatrix}, \tag{5}$$

denotes a $Q \times (U+1)^2$ matrix whose entries are SHs $Y_u^v(\cdot, \cdot)$ of order u and degree v. SHs are defined as [50]

$$Y_u^v(\theta, \phi) \equiv \sqrt{\frac{(2u+1)(u-|v|)!}{4\pi(u+|v|)!}} \mathcal{P}_u^{|v|}(\sin\theta) e^{iv\phi}, \tag{6}$$

where $|\cdot|$ is the absolute value operator, $\mathcal{P}_u^{|v|}(\cdot)$ is the associated Legendre function of order u and degree $|v|$, $i = \sqrt{-1}$ is the imaginary unit, and $e^{(\cdot)}$ is the exponential function. SHs is the solution of the Helmholtz equation for the elevation angle θ and the azimuth angle ϕ [43].

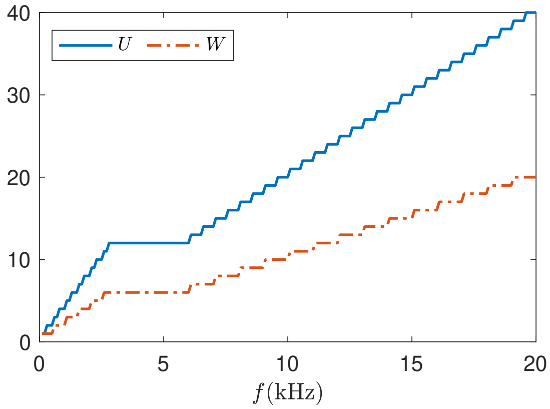


Fig. 2. Dimensionality U of HRTFs under SH decomposition and the PINN width W as functions of frequency f.

In Eqs. (4) and (5), U is the dimensionality of HRTF under SH decomposition and can be approximated as [47], [48]

$$U = \lceil 2\pi f r_\mathrm{h}/c \rceil, \tag{7}$$

where $\lceil \cdot \rceil$ is the ceiling operation, and

$$r_\mathrm{h} = \begin{cases} 0.2 \text{ m}, & f \leq 3 \text{ kHz}, \\ 0.09 \text{ m}, & f > 3 \text{ kHz}, \end{cases} \tag{8}$$

is the radius of human head, including the head-and-torso scattering effect. In this paper, for simplicity, we approximate Eq. (7) as

$$U \approx \begin{cases} \lceil f/250 \rceil, & f < 3 \text{ kHz}, \\ 12, & 3 \text{ kHz} \leq f \leq 6 \text{ kHz}, \\ \lceil f/500 \rceil, & f > 6 \text{ kHz}, \end{cases} \tag{9}$$

and present U as a function of frequency in Fig. 2 for reference. Since the dimensions of human heads are highly individual, Eq. (9) and Fig. 2 represent approximations only.

The regularized SH method [11] first estimates the SH coefficients $\hat{\mathbf{A}} = [\hat{A}_{0,0}(\omega), \hat{A}_{1,-1}(\omega), ..., \hat{A}_{U,U}(\omega)]^\intercal$ through

$$\hat{\mathbf{A}} = (\mathbf{Y}^\intercal \mathbf{Y} + \gamma \mathbf{H})^{-1} \mathbf{Y}^\intercal \mathbf{P}, \tag{10}$$

where $\mathbf{H}$ is a $(U+1)^2 \times (U+1)^2$ diagonal matrix whose diagonal entries are $h_{l,l} = 1 + u(u+1)$ and γ is the regularization parameter. The regularization limits the estimated SH coefficients $\hat{\mathbf{A}}$, especially the high-order coefficients, from taking large values [11]. To estimate the SH coefficients up to order U, the number of measured HRTFs needs to be sufficiently large, i.e., $Q > (U+1)^2$ [47], [48].

The HRTF for an arbitrary direction (θ_e, ϕ_e) can be estimated as

$$\hat{P}_\mathrm{SH}(\omega, \theta_e, \phi_e) \approx \mathbf{Y}_e \hat{\mathbf{A}}, \tag{11}$$

where $\mathbf{Y}_e = [Y_0^0(\theta_e, \phi_e), Y_1^{-1}(\theta_e, \phi_e), ..., Y_U^U(\theta_e, \phi_e)]$.

IV. Physics-informed-neural-network-based method

In this section, we first briefly introduce general PINN methods, then propose a PINN method for HRTF upsampling, and at last provide rationals for the configuration of the proposed PINN method.

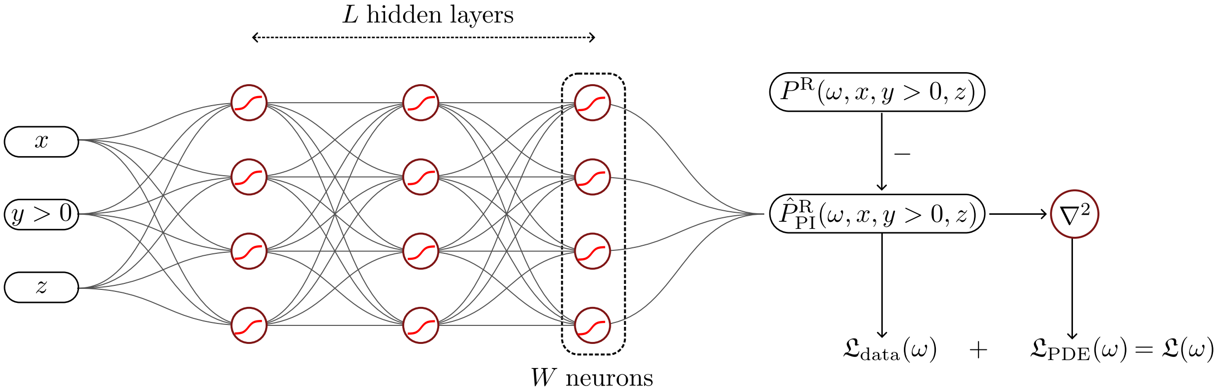

L hidden layers
x
$y > 0$
z
W neurons
$P^{\text{R}}(\omega, x, y > 0, z)$
$-$
$\hat{P}^{\text{R}}_{\text{PI}}(\omega, x, y > 0, z)$
∇^2
$\mathfrak{L}_{\text{data}}(\omega) \quad + \quad \mathfrak{L}_{\text{PDE}}(\omega) = \mathfrak{L}(\omega)$

Fig. 3. Structure of the proposed PINN method for modeling the real and left part HRTF: The inputs are Cartesian coordinates $(x, y > 0, z)$ and the output is HRTF estimation $\hat{P}^{\text{R}}_{\text{PI}}(\omega, x, y > 0, z)$. There are L hidden layers with W neurons in each hidden layer, and the activation function is $\tanh$. Data loss and PDE loss are calculated with respect to HRTF estimation $\hat{P}^{\text{R}}_{\text{PI}}(\omega, x, y > 0, z)$ and its Laplacian ∇^2, respectively.

A. General PINN methods

PINN methods are commonly constructed as multi-layer fully connected feed-forward neural networks [31]–[34]. The functionality of one layer is

$$\mathcal{P}(\mathbf{x}) = [\sigma\left(\mathbf{x}^{\intercal}\mathbf{w}_1 + b_1\right), \sigma\left(\mathbf{x}^{\intercal}\mathbf{w}_2 + b_2\right), ..., \sigma\left(\mathbf{x}^{\intercal}\mathbf{w}_J + b_J\right)], \tag{12}$$

where $\mathbf{x}$ is the input variable vector, $\{\mathbf{w}_{j=1}^{J}\}$ are the weight vectors, $\{b_j\}_{j=1}^{J}$ are the biases, J is the number of neurons, and $\sigma(\cdot)$ is the activation function. The overall functionality of a PINN method is the composition of L layers

$$\hat{P}(\mathbf{x};\zeta) = \mathcal{P}_L(....(\mathcal{P}_2(\mathcal{P}_1(\mathbf{x})))), \tag{13}$$

where ζ represents the set of trainable parameters, and $\hat{P}(\mathbf{x};\zeta)$ represents the network output. ζ is adjusted by minimizing a loss function

$$\mathfrak{L}=(1-\gamma)\frac{1}{Q}\sum_{q=1}^{Q}\left(P_q - \hat{P}(\mathbf{x}_q;\zeta)\right)^2 + \lambda\mathfrak{L}_{\text{PDE}}(\mathbf{x};\zeta), \tag{14}$$

where $\{\mathbf{x}_q, P_q\}_{q=1}^{Q}$ are input-output training data pairs which are obtained by testing and measuring a physical system, $\mathfrak{L}_{\text{PDE}}(\mathbf{x};\zeta)$ corresponds to the residual of the governing PDE, and λ is a regularization parameter which balances the contributions of two loss terms to the total loss $\mathfrak{L}$.

B. Proposed PINN method

We use four PINN methods to model the HRTF for one ear of a person at a single frequency ω. Specifically, for

1) the real and left part $P^{\text{R}}(\omega, x, y > 0, z)$;
2) the real and right part $P^{\text{R}}(\omega, x, y < 0, z)$;
3) the imaginary and left part $P^{\text{I}}(\omega, x, y > 0, z)$;
4) the imaginary and right part $P^{\text{I}}(\omega, x, y < 0, z)$.

The superscripts $(\cdot)^{\text{R}}$ and $(\cdot)^{\text{I}}$ denote the real part and the imaginary part of a value, respectively. With the person facing the positive x-axis as shown in Fig. 1, $y > 0$ and $y < 0$ denote the left side HRTF and the right side HRTF, respectively.

The structure of the PINN method for modeling the real and left part HRTF $P^{\text{R}}(\omega, x, y > 0, z)$ is shown in Fig. 3. The loss function is given by

$$\mathfrak{L}(\omega) = \underbrace{\frac{1}{Q}\sum_{q=1}^{Q}(P^{\text{R}}(\omega, x_q, y_q > 0, z_q) - \hat{P}^{\text{R}}_{\text{PI}}(\omega, x_q, y_q > 0, z_q))^2}_{\mathfrak{L}_{\text{data}}(\omega)}$$
$$\underbrace{+\frac{1}{D}\sum_{d=1}^{D}(\frac{\nabla^2\hat{P}^{\text{R}}_{\text{PI}}(\omega, x_d, y_d > 0, z_d)}{(w/c)^2} + \hat{P}^{\text{R}}_{\text{PI}}(\omega, x_d, y_d > 0, z_d))^2,}_{\mathfrak{L}_{\text{PDE}}(\omega)} \tag{15}$$

where the Laplacian operator ∇^2 is given by Eq. (3), $\{x_q, y_q > 0, z_q\}_{q=1}^{Q}$ are Cartesian coordinates of the measured HRTF $P^{\text{R}}(\omega, x_q, y_q > 0, z_q)$, $\{x_d, y_d > 0, z_d\}_{d=1}^{D}$ denote the Cartesian coordinates (or directions) of the unknown HRTF we want to upsample and is a super set of $\{(x_q, y_q > 0, z_q)\}_{q=1}^{Q}$, and $\mathfrak{L}_{\text{data}}(\omega)$ and $\mathfrak{L}_{\text{PDE}}(\omega)$ denote data loss and PDE loss, respectively. Note that we regard HRTF as a sound-field around a human head and thus the Cartesian coordinates in Eq. (15) correspond to $(r_{\text{h}}, \theta, 0 \leq \phi < \pi)$.

Except the training data and the output, the structures of PINN methods for modeling other three parts and the loss functions are identical to Fig. 3 and Eq. (15), respectively. Once trained, we combine the outputs of four PINN methods to arrive at the complex value HRTF for a single ear of a person at a single frequency, i.e.,

$$\begin{aligned}\hat{P}_{\text{PI}}(\omega, x, y, z) = \hat{P}^{R}_{\text{PI}}&(\omega, x, y > 0, z)\\ \cup\, i \cdot \hat{P}^{I}_{\text{PI}}&(\omega, x, y > 0, z)\\ \cup\, \hat{P}^{R}_{\text{PI}}&(\omega, x, y < 0, z)\\ \cup\, i \cdot \hat{P}^{I}_{\text{PI}}&(\omega, x, y < 0, z),\end{aligned} \tag{16}$$

where $\cup$ is the union operator. Denote (θ_e, ϕ_e) as an arbitrary direction, the PINN method estimates HRTF for that direction as $\hat{P}_{\text{PI}}(\omega, x_e, y_e, z_e)$, where (x_e, y_e, z_e) correspond to $(r_{\text{h}}, \theta_e, \phi_e)$.

We use four PINN methods to model HRTF for two reasons. First, when the loudspeaker is at the same side with the ear the magnitude of HRTF tends to be larger than the magnitude of HRTF when the loudspeaker is at the opposite side to the ear. The magnitude difference subsequently affects their contributions to the loss function, making a single PINN method difficult to attain consistent upsampling accuracy for both sides. Second, training real-valued neural networks is simpler compared to their complex-valued counterparts [51].

C. Configuration rationals

The configuration rationals for the proposed PINN method are provided below:

1) **Cartesian coordinates vs spherical coordinates:**
For the PINN method, HRTF is expressed in Cartesian coordinates instead of spherical coordinates for two reasons. First, the Laplacian in spherical coordinates, Eq. (2), can be numerically unstable due to the $\sin\theta$ term in denominators. Second, HRTF is evaluated on a single sphere, and thus there is no variations along the radial direction. This makes the PINN unable to estimate the first-order and second-order radial gradient used for calculating the Laplacian in spherical coordinates, Eq. (2).

2) **Frequency-wise upsampling:**
The proposed method focuses on upsampling HRTF for each frequency. This allows better control of the training process with respect to frequency as shown in **Helmholtz equation** and ***PINN width*** below.
To model HRTF of a person at all frequencies and at two ears, we need to build $L_\omega \times 2 \times 4 = 8\ L_\omega$ PINN methods, where L_ω is the number of frequencies of interest.

3) **Loss:**
The data loss $\mathfrak{L}_{\text{data}}(\omega)$ prompts the PINN method output to approximate the measured HRTF, i.e., $\hat{P}_{\text{PI}}(\omega, x_q, y_q, z_q) \approx P(\omega, x_q, y_q, z_q)$ for $q \in [1, Q]$.
The PDE loss $\mathfrak{L}_{\text{PDE}}(\omega)$ regularizes the PINN method output to conform with the Helmholtz equation at $\{(x_d, y_d, z_d)\}_{d=1}^{D}$, a super set of $\{(x_q, y_q, z_q)\}_{q=1}^{Q}$. This helps the PINN method to generate physically valid output beyond the training data. The regularization in Eq. (10), on the other hand, may not enable the SH method to generate physically valid output as shown in the experiment section.

4) **Helmholtz equation:**
Generally speaking, PINN methods are trying to solve multiple loss optimization problems [52]. Additional parameters, such as λ in Eq. (14), are normally used to balance different loss terms [52]. Although tuning additional parameters may improve the performance of PINN methods, we decided not to do so.
Instead, we modify the Helmholtz equation Eq. (1) to be

$$\frac{\nabla^2 P}{(\omega/c)^2} + P = 0. \tag{17}$$

Corresponding modifications are made to the PDE loss in Eq. (15). Under the modification, the PDE loss will have the same physical unit as the data loss, and hence balancing is not needed. Without the need for tuning loss-balancing parameters [52], the training of the PINN method is simplified.

5) **PINN method width:**
Based on our knowledge of SHs, we provide guidance on the PINN method width, the number of neurons in each hidden layer.
Although HRTF is defined over all directions, the most interesting directions are on the xy plane [1], where the SH decomposition of HRTF reduces to

$$\begin{aligned} P(\omega, \theta = 0, \phi) &\approx \sum_{u=0}^{U} \sum_{v=-u}^{u} A_{u,v}(\omega) Y_u^v(0, \phi) \\ &= \sum_{u=0}^{U} \sum_{v=-u}^{u} A_{u,v}(\omega) \\ &\quad \times \sqrt{\frac{(2\mu + 1)(u - |v|)!}{4\pi(u + |v|)!}} \mathcal{P}_u^{|v|}(0) e^{iv\phi} \\ &= \sum_{v=-U}^{U} A_v(\omega) e^{iv\phi}, \end{aligned} \tag{18}$$

and

$$A_v(\omega) = \sum_{u=|v|}^{U} A_{u,v}(\omega) \sqrt{\frac{(2u + 1)(u - |v|)!}{4\pi(u + |v|)!}} \mathcal{P}_u^{|v|}(0), \tag{19}$$

is obtained by manipulating the second and third lines of Eq. (18). Eq. (18) indicates that HRTF on the xy plane can be expressed by $2U + 1$ basis functions with related weights $\{e^{iv\phi}, A_v(\omega), v \in [-U, U]\}$.
Based on Fig. 3, from the output's point of view, HRTF on the xy plane can be expressed as

$$\hat{P}_{\text{PI}}(\omega, x, y, z = 0) = \sum_{j=1}^{W} \sigma_j \mathrm{w}_j + \sigma_0 b, \tag{20}$$

where σ_j denotes the value of the j-th neuron on the last hidden layer, w_j is the corresponding weight, and the bias b can be regarded as the weight for a linear activation function σ_0. $\{\sigma_j\}_{j=0}^{W}$ are implicit functions of Cartesian coordinates.
If the PINN method learns the underlying HRTF, then for corresponding ϕ and (x, y) there must be

$$P(\omega, \theta = 0, \phi) \approx \hat{P}_{\text{PI}}(\omega, x, y, z = 0), \tag{21}$$

and hence

$$\sum_{v=-U}^{U} A_v(\omega) e^{iv\phi} \approx \sum_{j=1}^{W} \sigma_j \mathrm{w}_j + \sigma_0 b, \tag{22}$$

though they are expressed in different coordinates and with different basis functions. Based on Eq. (22) and the minimum description length principle [53], we choose the PINN method width as

$$W + 1 = 2U + 1, \quad \text{or} \quad W = 2U. \tag{23}$$

In Sec. IV-B, four identical PINN methods are used for modeling the left/right and real/imaginary parts of HRTF. This prompts us to arrive at the final choice for the width

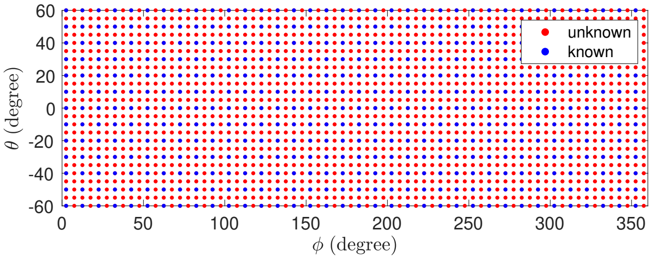


Fig. 4. Directions of the known HRTF and the unknown HRTF for the interpolation experiment.

(the number of neurons in each hidden layer) of a PINN method as

$$\begin{aligned} W &= 2U/4 \\ &= U/2 \\ &\approx \begin{cases} \lceil f/500 \rceil, & f < 3\,\text{kHz}, \\ 6, & 3\,\text{kHz} \le f \le 6\,\text{kHz}, \\ \lceil f/1000 \rceil, & f > 6\,\text{kHz}. \end{cases} \end{aligned} \tag{24}$$

The width W as a function of frequency is presented in Fig. 2 for reference. Note that similar to Eqs. (7) and (9), Eq. (24) represents approximations only.

6) **PINN depth:**
For the proposed PINN method, with no additional parameters for different loss terms in Eq. (15), the activation function fixed to be $\tanh$, and the width set to be $W = U/2$, the last parameter that could determine its performance is the depth L (the number of hidden layers), other than the training data. We found that a depth of $L = 3$ is a suitable choice, which balances the upsampling accuracy with the model complexity. This may be because the Helmholtz equation is a second-order PDE or three variables (x, y, z) are needed to determine the HRTF for a direction. In future studies, we plan to further analyze the Helmholtz equation and its solutions to determine an optimal depth L.

V. Interpolation experiment

Numerical experiments were conducted in this section to interpolate unknown HRTF whose direction is between those of measured ones.

A. Data processing

Experiments were conducted on the HUTUBS dataset [50], subject 11, 12, ... 50, left-ear HRTF. Based on SH coefficients up to 35-th order [50], HRTF for 330 directions, where $\theta \in [-60°, -48°, ..., 60°]$ and $\phi \in [4°, 16°, ..., 352°]$, is calculated according to Eq. (4) and used as the known HRTF; HRTF for 930 directions, where $\theta \in [-60°, -54°, ..., 60°]$ and $\phi \in [4°, 10°, ..., 358°]$, is calculated according to Eq. (4) and used as the unknown HRTF (ground-truth). Directions of the known HRTF and the unknown HRTF for the interpolation experiment were shown in Fig. 4. The magnitudes of all HRTFs were normalized to be within $[0, 1]$.

The 256 tap head-related impulse response [50], sampled at 44100 Hz, was transformed into frequency domain through discrete Fourier transform, resulting corresponding HRTF. HRTF was evaluated at [2067, 4134, 6202, 8269, 10336, 12403, 14470] Hz, which approximate multiples of the base frequency $44100/256$ Hz. Hereafter, we referred to these frequencies as 2.1, 4.1, 6.2, 8.2, 10.3, 12.3, and 14.4 kHz for notational simplicity. HRTF was not evaluated in higher frequencies ($f > 14.4$ kHz) because they do not contribute significantly to the perception of source location [1], and was not evaluated in lower frequencies ($f < 2.1$ kHz) because as shown in Sec. V-E and Sec. VI-E the simple SH method achieved better performance than other methods.

B. Implementation

SH method:
We implemented the SH method [11] following Eqs. (4) - (11), and set $\gamma = 0$ in Eq. (10) according to a trial-and-error process.

PINN method:
The known HRTF and corresponding Cartesian coordinates were the training data pairs for calculating the data loss $\mathfrak{L}_{\text{Data}}(\omega)$. Cartesian coordinates of the known HRTF and the unknown HRTF were combined and used for calculating the PDE loss $\mathfrak{L}_{\text{PDE}}(\omega)$.

To investigate how depth L and width W influence the interpolation performance, we implemented the PINN method in seven cases:

1) $L = 2$, $W = U/2$;
2) $L = 2$, $W = U$;
3) $L = 3$, $W = U/2$;
4) $L = 3$, $W = U$;
5) $L = 4$, $W = U/2$;
6) $L = 4$, $W = U$;
7) $L = 4$, $W = 50$.

For the first six cases, the width W varied with frequencies. For the last case, the width W was fixed.

We used the Tensorflow library for training, initialized the trainable parameters according to the Xavier initialization [54], set the activation function to be $\tanh$, chose the ADAM optimizer with a learning rate of 0.001, and trained the PINN methods for 10^6 epochs.

NN method:
Another method was implemented identical to the PINN method, except that the PDE loss $\mathfrak{L}_{\text{PDE}}(\omega)$ was not calculated and thus was not used for regularizing the network output. Hereafter, this method was denoted as the NN method. The NN method was implemented with different numbers of hidden layers and neurons same to the first four cases of the PINN method.

HRTF field method:
An additional HRTF field method [26] was implemented using the Sinusoidal Representation Network (SIREN) architecture, which is also a multi-layer fully connected feed-forward neural network but using sin as the activation function. The method was trained three times. In the first time, left-ear HRTF of

subject 1-10 and 25-96 was used as training set, and left-ear HRTF of subject 11-24 was used as testing set. In the second time, left-ear HRTF of subject 1-24 and 40-96 was used as training set, and left-ear HRTF of subject 25-39 was used as testing set. In the third time, left-ear HRTF of subject 1-35 and 51-96 was used as training set, and left-ear HRTF of subject 36-50 was used as testing set. The three time training was conducted because we found that the method was unable to converge if trained with less than 70 subjects' HRTF data. For each time, the training data consisted of 330 known HRTFs and corresponding Cartesian coordinates from the test set, along with 1260 known plus unknown HRTFs and corresponding Cartesian coordinates from the training set; the testing data consisted of 1260 known and unknown HRTFs and corresponding Cartesian coordinates from the test set. We set the learning rate following [26] and trained for 240 epochs. Please refer to [26] for the theory and implementation of the HRTF field method.

C. Performance evaluation

The performance of each method was evaluated based on the interpolation error

$$\mathcal{E}(\omega)=20\log_{10}\frac{\sum_{e=1}^{930}|P(\omega,\theta_e,\phi_e)-\hat{P}(\omega,\theta_e,\phi_e)|}{\sum_{e=1}^{930}|P(\omega,\theta_e,\phi_e)|},\quad (25)$$

where $P(\omega,\theta_e,\phi_e)$ and $\hat{P}(\omega,\theta_e,\phi_e)$ were the unknown HRTF (ground-truth) and its estimation generated by different methods at directions $\{(\theta_e,\phi_e)\}_{e=1}^{930}$, respectively.

D. Result: at a frequency for one subject

We presented an interpolation result in Fig. 5, which showed magnitudes of the left-ear HRTF of subject 40 at 14.4 kHz, ground-truth, interpolations and interpolation errors of different methods.

Note that, in captions of Figs. 5, 6, 8, and 9, we used "P, $L=3, W=U/2$, -8.4 dB" to denote that the PINN method with depth $L=3$ and width $W=U/2$ achieved interpolation error of -8.4 dB. Other captions could be interpreted similarly.

In the case, the dimensionality of HRTF under SH decomposition was $U=\lceil 2\pi f r_{\mathrm{h}}/c\rceil=29$. With only $330\ll(29+1)^2$ known HRTFs, the SH method was unable determine the SH coefficients up to $U=30$, and hence was unable to accurately interpolate the unknown HRTF.

Referring to Fig. 4 and Fig. 5 (c) - (f), we saw the NN methods assigned physically invalid values to the unknown HRTFs, and the interpolation errors were larger than 0 dB in all four cases.

Thanks to the regularization of the PDE loss, the PINN methods did not assigned physically invalid values to the unknown HRTFs as shown in Fig. 5 (g) - (m). Nonetheless, the PINN methods tended to assign zero to the unknown HRTFs, especially the deeper and wider PINN method shown in Fig. 5 (m). This indicated over-fitting and demonstrated the difficult of training a PINN [44]–[46]. Nonetheless, the PINN method with depth $L=3$ and width $W=U/2$ showed the least interpolation error of -8.4 dB.

As shown in Fig. 5 (n), the HRTF field method failed to capture the structure of HRTF, and the interpolation error was about -1.3 dB.

E. Result: across all frequencies and subjects

Figure 6 showed the interpolation errors across all frequencies and subjects.

The interpolation errors of the SH method were the lowest among all methods in lower frequency range, where $f=2.1, 4.1$ kHz. With the increment of frequency, the SH method interpolation error increased as well as the interpolation errors all other methods.

The NN method interpolation errors of four cases were similar across all frequencies. Unlike other methods, the interpolation errors of NN methods could go above 0 dB.

For frequency $f\leq 4.1$ kHz, the PINN method interpolation errors were similar to those of the NN methods. However, for $f\geq 8.2$ kHz, the PINN method interpolation errors were lower than those of the NN methods. in high-frequency range where $f\geq 10.3$ kHz, the PINN method with $L=3$ and $W=U/2$ demonstrated the smallest average interpolation errors among the 40 subjects for about -14 dB at 10.3 kHz, about -12.5 dB at 12.3 kHz, and about -9 dB at -14.4 kHz.

The interpolation errors of the HRTF field method showed the least variance across the frequencies. However, it did not achieved the smallest interpolation errors at any frequency.

VI. Extrapolation experiment

Numerical experiments were conducted in this section to extrapolate the unknown HRTF whose direction is beyond those of the measured ones.

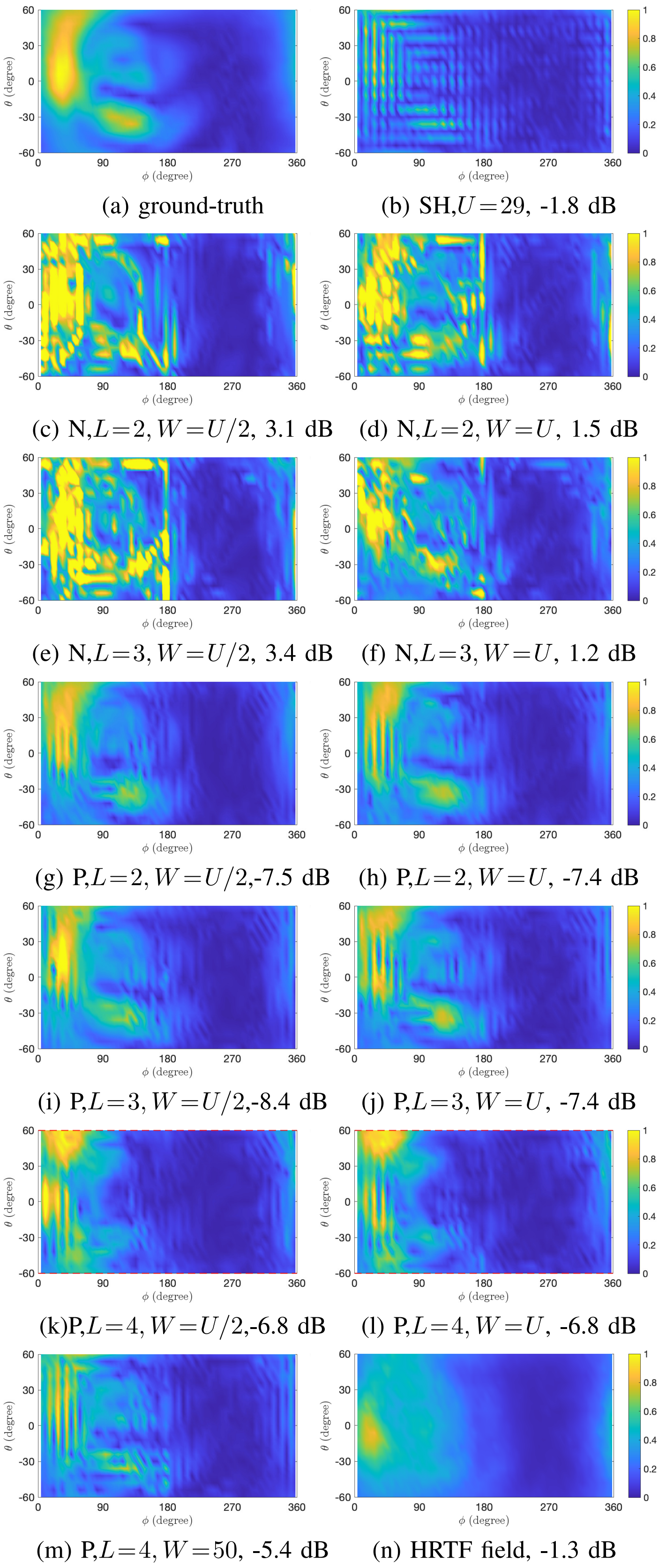
A. Data processing

Experiments were also conducted on the HUTUBS dataset [50], subject 11 to 50, left-ear HRTF. Based on SH coefficients up to 35-th order [50], HRTF for 675 directions, where $\theta\in[-56^\circ,-48^\circ,...,56^\circ]$ and $\phi\in[4^\circ,12^\circ,...,356^\circ]$, were calculated according to Eq. (4) and used as the known HRTF; HRTF for 270 directions, where $\theta\in[-80^\circ,-72^\circ,-64^\circ,64^\circ,72^\circ,80^\circ]$ and $\phi\in[4^\circ,12^\circ,...,356^\circ]$, were calculated according to Eq. (4) and used as the unknown HRTF (ground-truth). Directions of the known HRTF and the unknown HRTF for the extrapolation experiment were shown in Fig. 7. HRTF was evaluated at the same frequencies as in Sec. V. The magnitudes of all HRTFs were normalized to be within $[0,1]$.

B. Implementation

SH method:
We implemented the SH method [11] following Eqs. (4) - (11), and set $\gamma=10^{-6},10^{-4},10^{-1},10^{-1},10^{-2},10^{-1},10^{-3}$ for $f=2.1$, 4.1, 6.2, 8.2, 10.3, 12.3, 14.4 kHz in Eq. (10) according to a trial-and-error process.

PINN method and NN method:
The implementations of the PINN method and the NN method were identical to Sec. V B, except the training data and output.



(a) ground-truth (b) SH,$U=29$, -1.8 dB

(c) N,$L=2, W=U/2$, 3.1 dB (d) N,$L=2, W=U$, 1.5 dB

(e) N,$L=3, W=U/2$, 3.4 dB (f) N,$L=3, W=U$, 1.2 dB

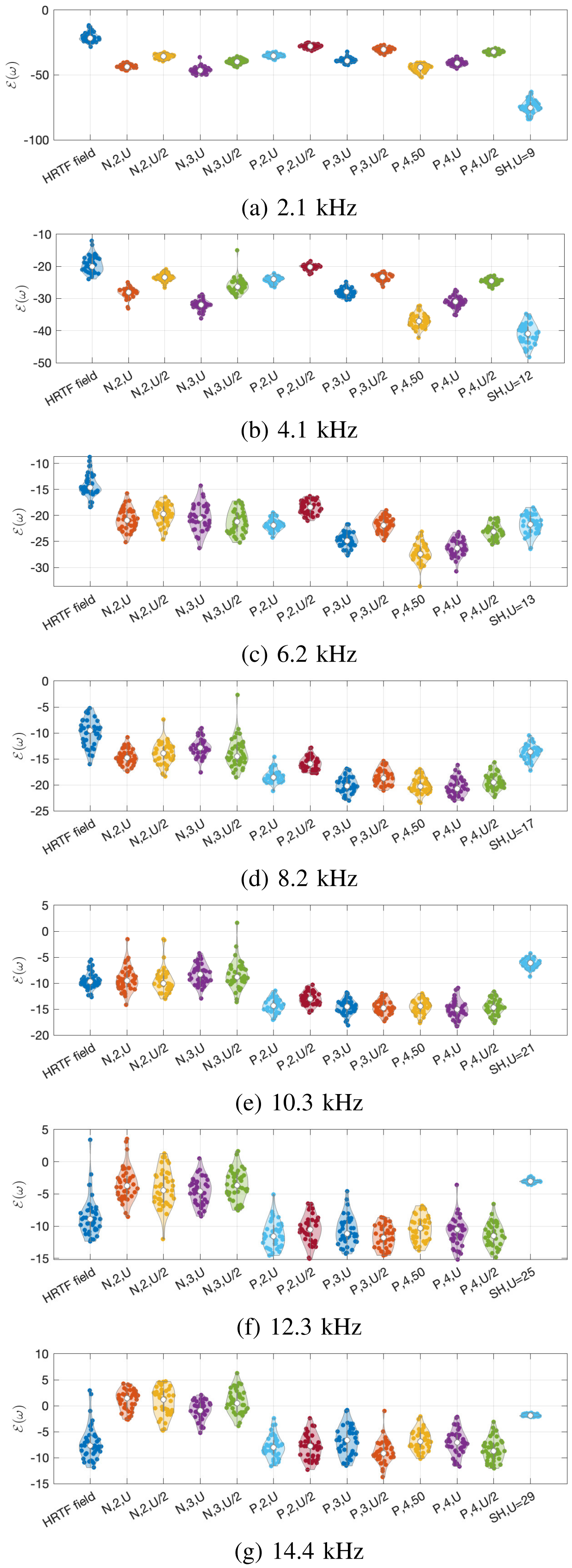
(g) P,$L=2, W=U/2$,-7.5 dB (h) P,$L=2, W=U$, -7.4 dB

(i) P,$L=3, W=U/2$,-8.4 dB (j) P,$L=3, W=U$, -7.4 dB

(k)P,$L=4, W=U/2$,-6.8 dB (l) P,$L=4, W=U$, -6.8 dB

(m) P,$L=4, W=50$, -5.4 dB (n) HRTF field, -1.3 dB

Fig. 5. Interpolation results: left-ear HRTF of subject 40 at 14.4 kHz, magnitudes of the ground-truth and interpolations by different methods.



(a) 2.1 kHz

(b) 4.1 kHz

(c) 6.2 kHz

(d) 8.2 kHz

(e) 10.3 kHz

(f) 12.3 kHz

(g) 14.4 kHz

Fig. 6. Interpolation errors across all frequencies and subjects.

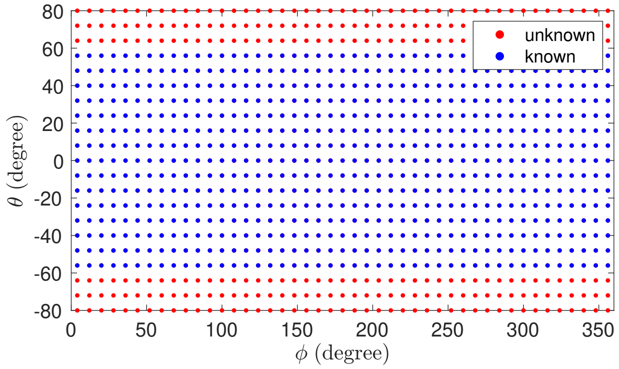


Fig. 7. Directions of the known HRTF and the unknown HRTF for the extrapolation experiment.

HRTF field method:
The implementation of the HRTF field method [26] was similar to Sec. V B, except the training and testing data. Training data consisted of the 675 known HRTFs and corresponding Cartesian coordinates from the test set, along with 945 known plus unknown HRTFs and corresponding Cartesian coordinates from the training set. Testing data consisted of 945 known plus unknown HRTFs and corresponding Cartesian coordinates from the test set. Please refer to [26] for the theory and implementation of the HRTF field method.

C. *Performance evaluation*

The performance of each method was evaluated based on the extrapolation error

$$\mathcal{E}(\omega) = 20\log_{10}\frac{\sum_{e=1}^{270}|P(\omega,\theta_e,\phi_e) - \hat{P}(\omega,\theta_e,\phi_e)|}{\sum_{e=1}^{270}|P(\omega,\theta_e,\phi_e)|},\tag{26}$$

where $P(\omega,\theta_e,\phi_e)$ and $\hat{P}(\omega,\theta_e,\phi_e)$ were the unknown HRTF (ground-truth) and its estimation generated by different methods at directions $\{(\theta_e,\phi_e)\}_{e=1}^{270}$, respectively.

D. *Result: at a single frequency for one subject*

We presented an extrapolation result in Fig. 8, which showed the magnitudes of left-ear HRTF of subject 20 at 14.4 kHz, ground-truth, extrapolations and extrapolation errors of different methods.

In the case, the SH method, the NN methods, and the HRTF field method all failed to extrapolate the unknown HRTF. Owing to the regularization of Eq. (10), the SH method [11] was unable to accurately represent the known HRTF. Similar to Fig. 6, the NN methods assigned physically invalid values to the unknown HRTFs.

As shown in Fig. 8 (g) - (m), without reducing the accuracy of representing the known HRTF and without assigning physically invalid values to the unknown HRTF, the PINN methods showed better extrapolation results than the SH method, the NN methods, and the HRTF field method. The PINN methods with width $W = U/2$ with depth $L = 3$ and $L = 4$ achieved the least extrapolation error of -5.6 dB and -6.0 dB, respectively.

E. *Result: across all frequencies and subjects*

Figure 9 showed the extrapolation errors across all frequencies and subjects. Comparing Fig. 9 with Fig. 6, we observed that at the same frequency the extrapolation errors were larger than the interpolation errors for all methods.

Small extrapolation error $\mathcal{E}(\omega) < -20$ dB of the SH method could only be achieved at frequency $f = 2.1$ kHz. For frequency above 4.1 kHz, the extrapolation errors of the SH method downgraded to be around 0 dB.

Below 8.2 kHz, the extrapolation errors of the NN methods were comparable to those of the PINN methods. However, above 10.3 kHz, the PINN method extrapolation errors were consistently smaller than corresponding NN methods. At $f = 12.3, 14.3$ kHz, the PINN method with depth $L = 3$ and width $W = U/2$ achieved the smallest average extrapolation errors of -4.8 dB and -4.7 dB, respectively.

The HRTF field method did not exhibit any extrapolation ability.

VII. Discussion

A. *Experiment results*

From the experiment results shown in Secs. V and VI we saw that all methods' performance downgraded with the increment of frequency.

The SH method's performance was the best in low-frequency range, but its performance down-gradation was significant in high-frequency range.

The NN methods achieved performance comparable to those of PINN methods for $f \leq 6.2$ kHz. Without regularization to the output, the NN methods assigned physically invalid values to the unknown HRTF in high-frequency range, $f \geq 12.3$ kHz.

The PINN methods demonstrated the least upsampling errors in high-frequency range, $f \geq 10.3$ kHz, where the upsampling was the most challenging.

The HRTF field method was useful for the interpolation scenario only.

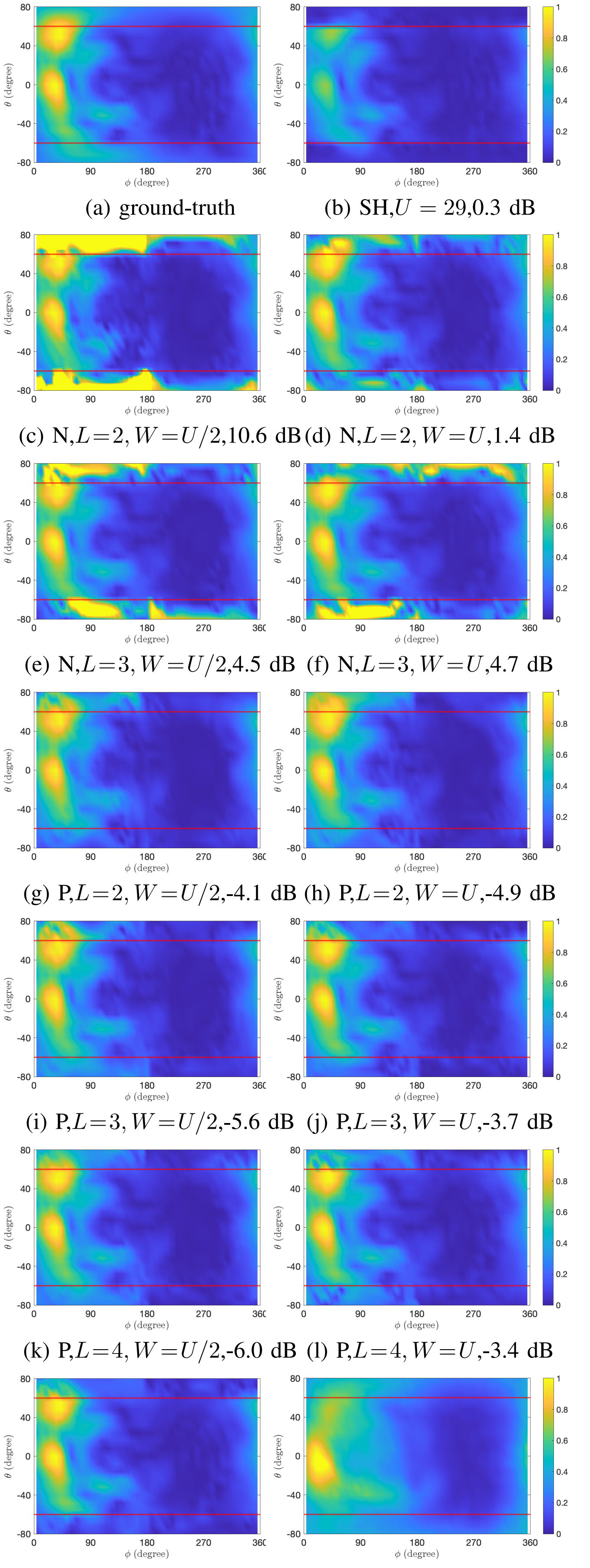
Based on these results, it is recommended to employ the SH method for HRTF upsampling in low-frequency range, the NN method in mid-frequency range, and the PINN method in high-frequency range, respectively.

B. *Regularization*

Fundamentally, HRTF upsampling is an ill-conditioned problem. To avoid generating physically invalid upsamplings like the NN methods, all upsampling methods have to exploit additional information to regularize the upsampling process.

The SH method [11] achieved the regularization by constraining the amplitudes of the estimated SH coefficients of high-orders. However, the high-order SH coefficients are necessary to represent the fine details of HRTF in high-frequency range. Thus, constraining their amplitudes will inevitably downgrade the upsampling performance.

The HRTF field method [26], as well as other learning based methods [19]–[25], achieved the regularization through learning implicit associations between direction (or ear geometry) and HRTF. Their dependence on the training data makes them lacks extrapolation ability as shown in Figs. 8 and 9.



(a) ground-truth (b) SH,$U = 29$,0.3 dB

(c) N,$L\!=\!2, W\!=\!U/2$,10.6 dB (d) N,$L\!=\!2, W\!=\!U$,1.4 dB

(e) N,$L\!=\!3, W\!=\!U/2$,4.5 dB (f) N,$L\!=\!3, W\!=\!U$,4.7 dB

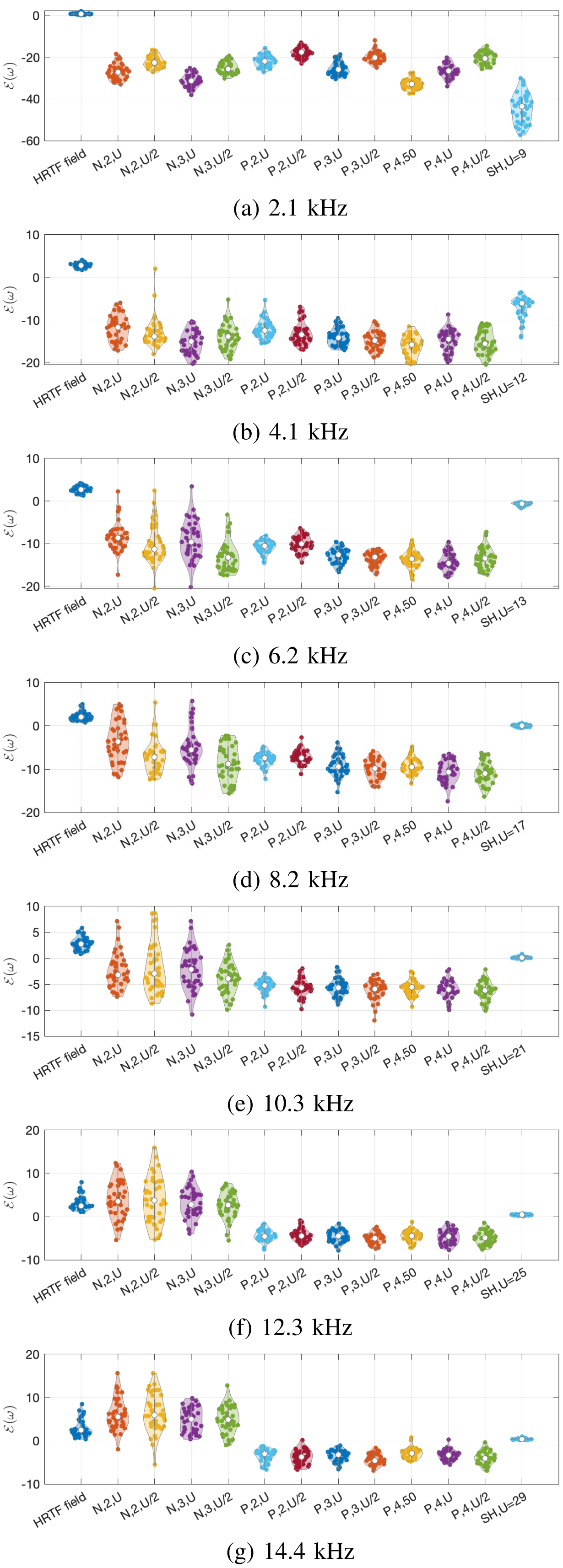
(g) P,$L\!=\!2, W\!=\!U/2$,-4.1 dB (h) P,$L\!=\!2, W\!=\!U$,-4.9 dB

(i) P,$L\!=\!3, W\!=\!U/2$,-5.6 dB (j) P,$L\!=\!3, W\!=\!U$,-3.7 dB

(k) P,$L\!=\!4, W\!=\!U/2$,-6.0 dB (l) P,$L\!=\!4, W\!=\!U$,-3.4 dB

(m) P,$L\!=\!4, W\!=\!50$,-3.7 dB (n) HRTF field, 3.4 dB

Fig. 8. Extrapolation results: left-ear HRTF subject 20 at 14.4 kHz, magnitudes of the ground-truth and extrapolations by different methods. The red lines denote the boundaries between the known HRTF and the unknown HRTF.



(a) 2.1 kHz

(b) 4.1 kHz

(c) 6.2 kHz

(d) 8.2 kHz

(e) 10.3 kHz

(f) 12.3 kHz

(g) 14.4 kHz

Fig. 9. Extrapolation result across all frequencies and subjects.

The proposed PINN method achieved the regularization through the PDE loss which is based on the Helmholtz equation. The Helmholtz equation dictates that sound pressure is proportional to its Laplacian

$$P = -\frac{1}{(\omega/c)^2}\nabla^2 P. \tag{27}$$

As shown in Figs. 5 and 8, this constrained the HRTF upsamplings from taking physically invalid values like the NN methods. Furthermore, the Helmholtz equation does not depend on any HRTFs, and thus that the PINN method can extrapolate the unknown HRTFs whose directions are beyond those of the known HRTFs.

C. PINN width and depth

To accurately model the training data, researchers tended to design PINN methods to be deep and wide [31]–[34]. As shown in Figs. 5, 6, 8, and 9, a deep and wide PINN method not necessarily achieved the best upsampling results, especially in high-frequency range. This indicated over-fitting [44]–[46].

To avoid over-fitting, we drew inspiration from the SH decomposition of HRTFs, and recommended to set the PINN width as $W = U/2$. As shown in Figs. 6 and 9, with the same depth L, most cases the PINN methods with width $W = U/2$ performed better than the PINN methods with width $W = U$ in high-frequency range $f \geq 10.3$ kHz, and slightly worse in low-frequency range $f \leq 8.2$ kHz.

Figures 6 and 9 also showed that, with the same width, the PINN methods of depth $L = 3$ performed better than the PINN methods of depth $L = 2$ in most cases, and comparable or slightly better than the PINN methods of depth $L = 4$ in high-frequency range $f \geq 10.3$ kHz.

Further considering that, as shown in Table I, the number of trainable parameters of the PINN method with depth $L = 3$ and width $W = U/2$ was the second least, we recommended to set the PINN method depth as $L = 3$ and width as $W = U/2$, especially in high-frequency range $f \geq 10.3$ kHz. This could reduce the training time.

Nonetheless, the deeper and wider PINN method with depth $L = 4$ and width $W = 50$ did achieve the least upsampling errors below 6.2 kHz as shown in Figs. 6 and 9. This indicated that a deep and wide PINN method may be less susceptible to over-fitting in low-frequency range where HRTFs are smoother.

The design of the PINN method, specifically its width and depth, was still empirical. Further theoretical investigation is needed to provide better guidance on the PINN method design. This will be one of our future works.

D. HRTF extrapolation

Comparing Fig. 9 with Fig. 6, we saw that extrapolation was much more challenging than interpolation. The -5 dB extrapolation error of the PINN methods for $f \geq 10.3$ kHz was smaller than those of other methods, but may not be enough for accurate spatial audio reproduction. This indicated that the Helmholtz equation regularization alone was insufficient to help the PINN method to generate accurate extrapolations in high-frequency range.

TABLE I
NUMBER OF TRAINABLE PARAMETERS OF THE PINN METHODS.

depth and width	Number of trainable parameters	$f \leq 14.4$ kHz,$U \leq 29$
$L = 2$, $W = U/2$	$U^2/4 + 3U + 1$	≤ 316
$L = 2$, $W = U$	$U^2 + 6U + 1$	≤ 1081
$L = 3$, $W = U/2$	$U^2/2 + 7U/2 + 1$	≤ 556
$L = 3$, $W = U$	$2U^2 + 7U + 1$	≤ 2011
$L = 4$, $W = U/2$	$3U^2/4 + 7U/2 + 1$	≤ 781
$L = 4$, $W = U$	$3U^2 + 7U + 1$	≤ 2911
$L = 4$, $W = 50$	7901	

Exploration of additional information, such as human anatomy, ear geometries, and cross dataset knowledge, is necessary to further improve the performance of the PINN method for HRTF extrapolation. This will be one of our future works.

E. Limitations

Error metric:
In this paper, we evaluated the upsampling performance in terms of the MSE of HRTF magnitudes only. Error metrics, such as phase error of the upsampling and comparisons of HRTF magnitude spectra over frequency between the ground-truth and the upsampling in sagittal planes, would also be helpful to assess the upsampling performance. Furthermore, it is unclear to which extent the MSE shown in Figs. 5, 6, 8, and 9 are perceivable. These limitations will be addressed in a future work.

Computational complexity:
As shown in Table I, the number of trainable parameters of the PINN methods is small. However, as mentioned in Sec. IV-C 1), $8L_\omega$ PINN methods are needed to the model the HRTFs of a person for two ears and for L_ω frequencies. The computational complexity will be large. The computational complexity can be reduced by building a single PINN method that models HRTFs for both ears over all L_ω frequencies like the HRTF field method [26]. This will be one of our future works.

VIII. CONCLUSION

This paper proposed a PINN method for HRTF upsampling. The proposed method exploited the Helmholtz equation, the governing PDE of acoustic wave propagation, for constraining the upsampling process and generating physically valid outputs. Furthermore, based on the dimensionality of HRTF under SH decomposition and the Helmholtz equation, we set the PINN with an appropriate width and depth. This helped the PINN method to avoid the over-fitting problem. The Helmholtz equation regularized PINN method with a suitable width and depth outperformed the SH method, the NN method, and the HRTF field method in both interpolation and extrapolation experiments.

References

[1] S. Li and J. Peissig, "Measurement of head-related transfer functions: a review", *Appl. Sci.*, vol. 10, no. 14, pp. 5014, 2020.

[2] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: a review of spatial audio recording and reproduction", *Appl. Sci.*, vol. 7, no. 6, pp. 532, May 2017.

[3] J. G. Richter and J. Fels, "On the influence of continuous subject rotation during high-resolution head-related transfer function measurements", *IEEE/ACM Trans. on Audio, Speech, and Lang. Proces.*, vol. 27, no. 4, pp. 730-741, April 2019.

[4] G. Enzner, "3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering", *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, pp. 325-328, 2009.

[5] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Efficient representation and sparse sampling of head-related transfer functions using phase-correction based on ear alignment", *IEEE/ACM Trans. on Audio, Speech, and Lang. Proces.*, vol. 27, no. 12, pp. 2249-2262, Dec. 2019.

[6] J. M. Arend, C. Pörschmann, S. Weinzierl, and F. Brinkmann, "Magnitude-Corrected and Time-Aligned Interpolation of Head-Related Transfer Functions", arXiv preprint arXiv:2303.09966.

[7] L. S. Zhou, C. C. Bao, M. S. Jia, and B. Bu, "Range extrapolation of head-related transfer function using improved higher order ambisonics", *APSIPA ASC 2014*, pp. 1-4, 2014.

[8] H. Gamper, "Head-related transfer function interpolation in azimuth, elevation, and distance", *J. Acoust. Soc. Am.*, vol. 134, no. 6, pp. 533–547, 2013.

[9] M. Pollow, K. V. Nguyen, O. Warusfel, T. Carpentier, M. Muller-Trapet, M. Vorlander, and M. Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition", *Acta. Acustica united with Acustica*, vol. 98, no. 1, pp. 72–82, 2012.

[10] S. Spors and J. Ahrens, "Interpolation and range extrapolation of head-related transfer functions using virtual local sound-field synthesis", *130th Conv. AES*, May 2011.

[11] R. Duraiswaini, D. N Zotkin, and N. A Gumerov, "Interpolation and range extrapolation of head related transfer functions", *IEEE Int. Conf. on Acoust. Speech, and Signal Proces. (ICASSP)*, vol. 4, pp. iv–iv, 2004.

[12] M. J. Evans, J. A. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics", *J. Acoust. Soc. Am.*, vol. 104, no. 4, pp. 2400–2411, 1998.

[13] M. Aussal, F. Alouges, and B. Katz, "HRTF interpolation and ITD personalization for binaural synthesis using spherical harmonics", *Journal of Audio Engineering Society*, 2012.

[14] B. Xie, "Recovery of individual head-related transfer functions from a small set of measurements", *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 282–294, 2012.

[15] L. Wang, F. Yin, and Z. Chen, "Head-related transfer function interpolation through multivariate polynomial fitting of principal component weights", *Acoust. Sci. Tech.*, vol. 30, no. 6, pp. 395–403, 2009.

[16] M. Zhang, Z. Ge, T. Liu, X. Wu, and T. Qu, "Modeling of individual HRTFs based on spatial principal component analysis", *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 28, pp. 785-797, 2020.

[17] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions", *Proc. 16th Int. Audio Eng. Soc. Conf. Spatial Sound Reproduction*, pp. 319–329, 1999.

[18] J. C. B. Torres and M. R. Petraglia, "HRTF interpolation in the wavelet transform domain", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 293-296, 2009.

[19] T.-Y. Chen, T.-H. Kuo, and T.-S. Chi, "Autoencoding HRTFs for DNN based HRTF personalization using anthropometric features", *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2019)*, pp. 271-275, May 2019.

[20] R. Miccini and S. Spagnol, "HRTF individualization using deep learning", *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 390-395, Mar. 2020.

[21] Y. Ito, T. Nakamura, S. Koyama, and H. Saruwatari, "Head-related transfer function interpolation from spatially sparse measurements using autoencoder with source position conditioning", *International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 1-5, 2022.

[22] P. Siripornpitak, I. Engel, I. Squires, S. J. Cooper, and L. Picinali, "Spatial up-sampling of HRTF sets using generative adversarial networks: a pilot study", *Frontiers in Signal Processing*, vol. 2, 2022.

[23] A. O. Hogg, M. Jenkins, H. Liu, I. Squires, S. J. Cooper, and L. Picinali, "HRTF upsampling with a generative adversarial network using a gnomonic equiangular projection", *arXiv preprint arXiv:2306.05812*.

[24] J. W. Lee, S. Lee, and K. Lee, "Global hrtf interpolation via learned affine transformation of hyper-conditioned features", *IEEE Int. Conf. on Acoust. Speech and Signal Proces. (ICASSP)*, pp. 1-5, Jun. 2023.

[25] B. Zhi, D. N. Zotkin, and R. Duraiswami, "Towards fast and convenient end-to-end HRTF personalization", *IEEE Int. Conf. on Acoust. Speech and Signal Proces. (ICASSP)*, pp. 441-445, 2022.

[26] Y. Zhang, Y. Wang, and Z. Duan, "HRTF field: unifying measured HRTF magnitude representation with neural fields", *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp 1–5, Jun. 2023.

[27] W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Iterative extrapolation algorithm for data reconstruction over sphere", *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 3733–3736, Mar. 2008.

[28] U. Elahi, Z. Khalid, and R. A. Kennedy, "An improved iterative algorithm for band-limited signal extrapolation on the sphere", *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, pp. 4619-4623, Mar. 2018.

[29] J. Ahrens, M. R. P. Thomas, and I. Tashev, "HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data", *APSIPA*, Dec. 2012.

[30] C. Pörschmann, J. M. Arend, and F. Brinkmann, "Directional equalization of sparse head-related transfer function sets for spatial upsampling", *IEEE/ACM Transa. on Audio, Speech, and Lang. Proces*, vol. 27, no. 6, pp. 1060-1071, 2019.

[31] M. Raissi, P. Perdikaris, and G. Em Karniadakis, "Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations", *arXiv preprint arXiv:1711.10566*.

[32] M. Raissi, P. Perdikaris, and G. Em Karniadakis, "Physics informed deep learning (part II): data-driven discovery of nonlinear partial differential equations", *arXiv preprint arXiv:1711.10566*.

[33] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, May 2021.

[34] S. Cuomo, V. D. Di 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", *arXiv preprint arXiv:2201.05624*.

[35] C. Song, T. Alkhalifah, and U. B. Waheed, "Solving the frequency-domain acoustic VTI wave equation using physics-informed neural networks", *Geophys. J. Int.*, vol. 225, no. 2, pp. 846-859, 2021.

[36] P. Ren, C. Rao, H. Sun, and Y. Liu, "SeismicNet: physics-informed neural networks for seismic wave modeling in semi-infinite domain", *arXiv preprint arXiv:2210.14044*.

[37] Y. Wang, K. Wang, and M. Abdel-Maksoud, "NoiseNet: a neural network to predict marine propellers' underwater radiated noise", *Ocean Engineering*, vol. 236, pp. 109542, 2021.

[38] K. Shigemi, S. Koyama, T. Nakamura, and H. Saruwatari, "Physics-informed convolutional neural network with bicubic spline interpolation for sound-field estimation", *arXiv preprint arXiv:2207.10937*.

[39] B. Moseley, A. Markham, and T. Nissen-Meyer, "Solving the wave equation with physics-informed deep learning", *arXiv preprint arXiv:2006.11894*.

[40] N. Borrel-Jensen, A. P. Engsig-Karup, and C. H. Jeong, "Physics-informed neural networks for one-dimensional sound-field predictions with parameterized sources and impedance boundaries", *Jasa Express Lett.*, 2021.

[41] M. Rasht-Behesht, C. Huber, K. Shukla, and G. E. Karniadakis, "Physics-informed neural networks for wave propagation and full waveform inversions", *Journal of Geophysical Research: Solid Earth*, 2022.

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

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

[44] R. Leiteritz, and D. Pflüger, "How to avoid trivial solutions in physics-informed neural networks", *arXiv preprint arXiv:2112.05620.*

[45] S. Wang, Y. Teng and P. Perdikaris, "Understanding and mitigating gradient pathologies in physics-informed neural networks", *SIAM Journal on Scientific Computing*, vol. 43, no. 5, pp. 3055-3081, 2021.

[46] F. M. Rohrhofer, S. Posch, C. Gößnitzer, and B. C. Geiger, "Understanding the difficulty of training physics-informed neural networks on dynamical systems", *arXiv preprint arXiv:2203.13648*.

[47] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, “Insights into head-related transfer function: spatial dimensionality and continuous representation”, *J. Acoust. Soc. Amer.*, vol. 127, pp. 2347–2357, 2010.

[48] D. B. Ward and T. D. Abhayapala, “Reproduction of a plane-wave sound-field using an array of loudspeakers”, *IEEE Trans. Speech Audio Process.*, vol. 9, no. 66, pp. 697–707, 2001.

[49] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, “The expressive power of neural networks: a view from the width”, *Adv. Neural Inf. Process. Syst.*, pp. 6231–6239, 2017.

[50] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, “A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses”, *AES*, 2019.

[51] C. Lee, H. Hasegawa, and S. Gao, “Complex-valued neural networks: A comprehensive survey”, *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 8, pp. 1406-1426, 2022.

[52] O. Sener and V. Koltun, “Multi-task learning as multi-objective optimization”, *Adv. Neural Inf. Process. Syst.,* pp. 525–536, 2018.

[53] P. D. Grünwald, *The minimum description length principle*, MIT press, 2007.

[54] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks”, *AISTATS*, pp. 249–256, 2010.