Physics Informed Neural Network for Head-Related Transfer Function Upsampling

Fei Ma, Thushara D. Abhayapala, Prasanga N. Samarasinghe, and Xingyu Chen

***Abstract*—Head-related transfer functions (HRTFs) capture the spatial and spectral features that a person uses to localize sound sources in space, and thus are vital for creating authentic virtual acoustic experience. However, practical HRTF measurement systems can only provide an incomplete measurement of a person's HRTFs, and this necessitates HRTF upsampling. This paper proposes a physics-informed neural network (PINN) method for HRTF upsampling. Unlike other upsampling methods which are based on the measured HRTFs only, the PINN method exploits the Helmholtz equation as additional information for constraining the upsampling process. This helps the PINN method to generate physically amiable upsamplings which generalize beyond the measured HRTFs. Furthermore, the width and the depth of the PINN are set according to the dimensionality of HRTFs under spherical harmonic (SH) decomposition and the Helmholtz equation. This makes the PINN to have an appropriate level of expressiveness and thus does not suffer from under-fitting and over-fitting problems. Numerical experiments confirm the superior performance of the PINN method for HRTF upsampling in both interpolation and extrapolation scenarios over several datasets in comparison with the SH methods.**

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

I. Introduction

HEAD-related transfer functions (HRTFs) denote the free field acoustic transfer functions between a point source and a position inside of a person's ear [1]. HRTFs characterize the filtering effect of a person's torso, head, and ears with respect to the direction of sound [1], and contains the spatial and spectral features that a person uses to localize sound sources in space. Spatial audio and virtual acoustic systems rely on the knowledge of HRTFs to reproduce artificial acoustic experience [2].

However, the dependence of HRTFs on a person's anatomy makes HRTFs highly individualized, and thus that accurate measurements of HRTFs over a large number of directions are desired for creating authentic acoustic experience [1]. Nonetheless, the complete measurement of HRTFs is both time-consuming and expensive, which holds most people back from taking their HRTF measurement. Practical HRTF measurement systems usually have to measure HRTFs over a limited number of directions due to the inconvenience of arranging loudspeakers over a whole sphere [1] or due to the time constraint on the measurement process, resulting an incomplete HRTF dataset. The incompleteness of HRTF datasets motivates researchers to upsample them.

HRTF upsampling consists of two scenarios: interpolation and extrapolation. (Note that we focus on direction related HRTF upsampling. Distance related HRTF upsampling [3]–[6] is not covered by this paper.) For the interpolation scenario, the HRTFs are measured over a limited number of directions due to time constraint, and the aim is to estimate the unknown HRTFs whose directions are between those of the measured HRTFs. Early works on the interpolation are mainly based on the expansion of HRTFs on linear functions, such as the spherical harmonics (SHs) [7]–[9], the principle components [10]–[12], the spline functions [13], and the wavelet functions [14]. Recent works of HRTF interpolation, on the other hand, are mainly based non-linear modeling with neural networks (NNs) such as the auto encoder [15], the generative adversarial networks [16], and the feature-wise linear modulation [17].

For the extrapolation scenario, HRTFs are measured over a limited polar angle range due to the inconvenience of arranging loudspeakers deep below and high above a person, and the aim is to estimate the unknown HRTFs beyond the range. The missing information over a range makes the extrapolation much more challenging than the interpolation. Up to date, there are only several works about HRTF extrapolation, and the majority of them are based on the SH decomposition. Zhang *et al.* developed an iterative algorithm which successively fills and estimates the unknown HRTFs [18], [19], and successfully recovered a low order HRTF over a full sphere with one quarter of data missing. Zotkin *et al.* proposed a regularized least-square (LS) fit method which estimates the unknown HRTFs at the expense of reduced accuracy for representing the measured HRTFs [20]. Ahrens *et al.* proposed a non-regularized LS fit method which estimates the unknown HRTFs based on a low-order LS fit to the measured HRTFs and estimations of the unknown HRTFs [21].

One problem shared by most of the above mentioned HRTF upsampling methods is that they estimate the unknown HRTFs based on the measured HRTFs only. Their estimations are essentially transformations of the information that is contained in the measured HRTFs. They have not effectively use additional information to further improve the accuracy of their estimations. This fact prompts us to take a different approach to HRTF upsampling: the physics-informed neural network (PINN).

The PINN is a special kind of NN which incorporates physical knowledge, i.e., the governing partial differential equation

This work is sponsored by the Australian Research Council (ARC) Discovery Projects funding schemes with project numbers DP200100693.

Fei Ma, Thushara D. Abhayapala, and Prasanga N. Samarasinghe 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: {fei.ma, Thushara.Abhayapala, prasanga.samarasinghe}@anu.edu.au).

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

(PDE) of a physical phenomenon, into its architecture [23]–[26]. The physics knowledge is the additional information that can help a PINN to model the physic phenomenon beside the physical quantities. Since the seminal works of Raissi and his colleagues [23], [24], PINN has been successfully applied to many areas such as earth quake modeling [27], [28], propeller noise prediction [29], room acoustics [30], and sound field estimation [31].

The HRTFs can be regarded as the sound field around the human head, and sound fields in space obey the Helmholtz equation, the governing equation of acoustic wave propagation [32]. This inspires us to develop a PINN method for HRTF upsampling. We inform the training and the designing of the PINN with physics knowledge from two aspects. First, we use a modified form of the Helmholtz equation as part of the loss function. This helps the PINN to generate physically amiable 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 according to SH decomposition and the Helmholtz equation. Specifically, we set the width of the PINN as half of the dimensionality of HRTFs under SH decomposition [32], [33] and the depth of as three (the order of the Helmholtz equation plus one). This sets the proposed PINN method apart from PINN methods in other works which suffers from under-fitting or over-fitting problems due to inappropriate design of the network [34]–[36]. The superior performance of the PINN method for upsampling HRTFs is confirmed by numerical experiments on several datasets, and is compared with that of the SH method, the most widely used HRTF upsampling framework.

The rest of this paper is organized as follows. We introduce the problem in Sec. II. We review the SH method in Sec. III and propose the PINN method in Sec. IV. The performance of the PINN method and the SH method are compared by extrapolation and interpolation experiments in Sec. V. Section VI concludes this paper and points out future directions of improvement.

II. Problem formulation

We present the layout of a typical HRTF measurement system in Fig. 1, where we set up a Cartesian coordinate system and a spherical coordinate system with respect to the center of a person's head, point O. We denote the Cartesian coordinates and the spherical coordinates as (x, y, z) and (r, θ, ϕ), respectively. The system measures the HRTFs between the loudspeakers which are placed on a sphere $\mathbb{S}_2$ and the microphones which are placed inside of the person's ears. We denote HRTFs as $P(\omega, r, \theta, \phi)$ in spherical coordinates or as $P(\omega, x, y, z)$ in Cartesian coordinates [22], where $\omega = 2\pi f$ is the angular frequency and f is the frequency. Hereafter, we evaluate HRTFs at a single frequency and on a single sphere, and thus we skip frequency ω and the sphere radius r when representing acoustic quantities for notation simplicity.

Due to the obstruction of the person's body and the inconvenience of arranging the loudspeakers high above the person, the measurement system may only be able to measure the HRTFs over a polar range $(\theta_{\text{Low}}, \theta_{\text{High}})$, where $0 < \theta_{\text{Low}} < \theta_{\text{High}} < \pi$. Due to the time constraint, the measurement system may only be able to measure the HRTFs over limited number of directions. Both of these two scenarios will result in an incomplete HRTF datasets.

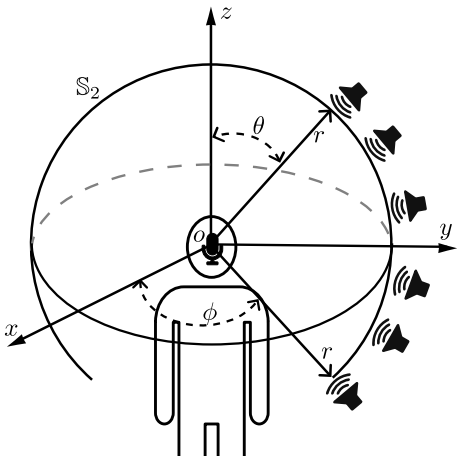


Fig. 1. Layout of a typical HRTF measurement system, which measures the HRTFs between the loudspeakers which are placed on a sphere $\mathbb{S}_2$ and the microphones which are placed inside of the person's ears.

In this paper, we aim to upsample an incomplete HRTF dataset $\{P(\theta_q, \phi_q)\}_{q=1}^{Q}$ or equivalently $\{P(x_q, y_q, z_q)\}_{q=1}^{Q}$ into a full or dense dataset.

III. Spherical harmonic methods

In this section, we first briefly present the SH decomposition of HRTFs, and then review the regularized SH method for HRTF upsampling [20]. We express HRTFs in spherical coordinates for the ease of SH decomposition.

HRTFs can be decomposed onto SHs and their coefficients as [37]

$$\mathbf{P} \approx \mathbf{Y}\mathbf{A}, \tag{1}$$

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

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\theta_1, \phi_1) & Y_1^{-1}(\theta_1, \phi_1) & ... & Y_U^U(\theta_1, \phi_1) \\ Y_0^0(\theta_2, \phi_2) & Y_1^{-1}(\theta_2, \phi_2) & ... & Y_U^U(\theta_2, \phi_2) \\ ... & ... & ... & ... \\ Y_0^0(\theta_Q, \phi_Q) & Y_1^{-1}(\theta_Q, \phi_Q) & ... & Y_U^U(\theta_Q, \phi_Q) \end{bmatrix}, \tag{2}$$

denotes a $Q \times (U+1)^2$ matrix whose entries are order u and degree v SH $Y_u^v(\cdot, \cdot)$ evaluated at $(\theta_q, \phi_q)_{q=1}^{Q}$.

In (1) and (2), U is the dimensionality of HRTFs [32], [33] under SH decomposition and is normally chosen as

$$U = \lceil 2\pi f r_{\text{h}}/c \rceil, \tag{3}$$

where $\lceil \cdot \rceil$ is the ceiling operation, $c = 343$ m/s is the speed of sound propagation, and

$$r_{\text{h}} = \begin{cases} 0.2 \text{ m}, & f \leq 3\text{kHz}, \\ 0.09 \text{ m}, & f > 3\text{kHz}, \end{cases} \tag{4}$$

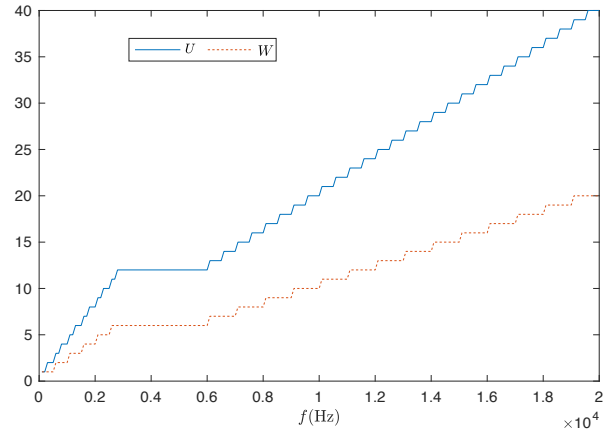


Fig. 2. Dimensionality U of HRTFs under SH decomposition and the PINN width W as functions of frequency f.

is the radius of human head (including the head and torso scattering effect) [32], [33]. In this paper, for simplicity, we choose

$$U = \lceil 2\pi f r_{\rm h}/c \rceil \approx \begin{cases} \lceil f/250 \rceil, & f \le 3{\rm kHz}, \\ \lceil f/500 \rceil, & f > 3{\rm kHz}, \end{cases} \tag{5}$$

and show the dimensionality U of HRTFs as a function of frequency in Fig. 2 for reference. In Fig. 2, we choose $U = \max\{\lceil f/250 \rceil, \lceil f/500 \rceil\}$ for 3 kHz $< f <$ 6 kHz [32]. Note that the sizes of human heads vary, and thus that (3), (5) and Fig. 2 should be regarded as a rule of thumb and are not supposed to be followed exactly.

The regularized SH method first estimates the SH coefficients through [20]

$$\hat{\mathbf{A}} = (\mathbf{Y}^\intercal \mathbf{Y} + \gamma \mathbf{H})^{-1} \mathbf{Y}^\intercal \mathbf{P}, \tag{6}$$

where $\mathbf{H}$ is a $(U+1)^2 \times (U+1)^2$ diagonal matrix whose diagonal entries are $h_{l,l} = 1 + u(u+1)$ (u is the order of corresponding SH), and γ is the regularization parameter. To accurately estimate the SH coefficients up to order U, we need the number of measured HRTFs to be sufficiently large, $Q > (U+1)^2$ or $U < \sqrt{Q} - 1$ [32], [33].

The regularized SH method then estimate the HRTF from an arbitrary direction (θ_e, ϕ_e) as

$$\hat{P}_{\rm SH}(\theta_e, \phi_e) \approx \sum_{u=0}^{U} \sum_{v=-u}^{u} \hat{A}_{u,v} Y_u^v(\theta_e, \phi_e). \tag{7}$$

The regularization in (6) prevents the estimated HRTF (7) from taking exceptionally large value by constraining the amplitudes of the estimated SH coefficients $\hat{\mathbf{A}} = [\hat{A}_{0,0}, \hat{A}_{1,-1}, \hat{A}_{1,0}, \hat{A}_{1,1}, ..., \hat{A}_{U,U}]^\intercal$ [20], especially the high order coefficients.

IV. PINN METHOD

In this section, we first briefly introduce the PINN, and then propose a PINN method for HRTF upsampling. We express HRTFs in Cartesian coordinates to simplify the calculation of the Laplacian by the PINN.

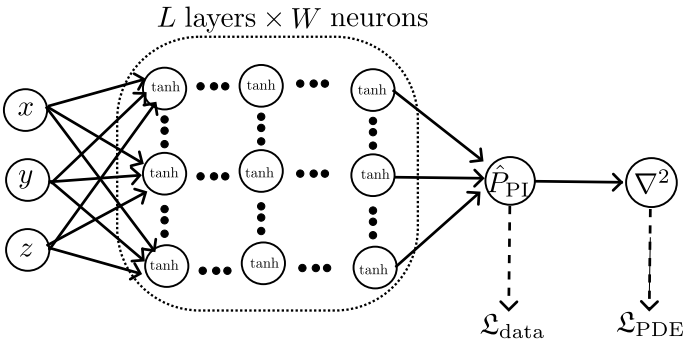


Fig. 3. Structure of the PINN: the inputs are the Cartesian coordinates, the outputs are the HRTF estimations $\hat{P}_{\rm PI}$, there are L hidden layers with W neurons on each hidden layer, we calculate the data loss and the PDE loss with respect the HRTF estimations $\hat{P}_{\rm PI}$ and their Laplacian ∇^2.

We normally build a PINN as a multiple layer fully connected feed-forward neural networks [23]–[26]. The functionality of one layer is

$$\mathfrak{P}(\mathbf{x}) = \sigma\left(\mathbf{x}^T \mathbf{w} + b\right), \tag{8}$$

where $\mathbf{x}$ is the input variable vector, $\mathbf{w}$ is the weight vector, b is the bias and σ is the activation function. The overall functionality of the PINN is the composition of L layers

$$\Phi(\mathbf{x}; \zeta) = \left(\mathfrak{P}_L \circ \circ \circ \mathfrak{P}_2 \circ \mathfrak{P}_1\right)(\mathbf{x}), \tag{9}$$

where ζ represents the set of all trainable parameters. We adjust the parameters ζ by minimizing a cost function

$$\mathfrak{L} = \frac{1}{Q} \sum_{q=1}^{Q} \left(p_q - \Phi(\mathbf{x}_q; \zeta)\right)^2 + \lambda \mathfrak{L}_{\rm PDE}(\mathbf{x}; \zeta), \tag{10}$$

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, $\mathfrak{L}_{\rm PDE}(\mathbf{x}; \zeta)$ corresponds to the residual of the governing PDE, and λ is a regularization parameter.

For HRTF upsampling, we design a PINN whose structure is shown in Fig. 3, where there are L hidden layers with W neurons on each hidden layer, with the inputs being the Cartesian coordinates, the activation function being $\tanh$, and the output being HRTF estimation $\hat{P}_{\rm PI}(x, y, z)$. We adjust the trainable parameters by minimizing the following cost function

$$\mathfrak{L} = \underbrace{\frac{1}{Q} \sum_{q=1}^{Q} \| P(x_q, y_q, z_q) - \hat{P}_{\rm PI}(x_q, y_q, z_q) \|_2^2}_{\mathfrak{L}_{\rm data}} + \underbrace{\frac{1}{D} \sum_{d=1}^{D} \| \frac{1}{(w/c)^2} \nabla^2 \hat{P}_{\rm PI}(x_d, y_d, z_d) + \hat{P}_{\rm PI}(x_d, y_d, z_d) \|_2^2}_{\mathfrak{L}_{\rm PDE}}, \tag{11}$$

where $\|\cdot\|_2$ is the 2-norm, $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplacian operator [37], $\{x_q, y_q, z_q\}_{q=1}^{Q}$ are the Cartesian coordinates of the measured HRTFs, $\{x_d, y_d, z_d\}_{d=1}^{D}$ is a super set of $\{(x_q, y_q, z_q)\}_{q=1}^{Q}$, $\mathfrak{L}_{\rm data}$ and $\mathfrak{L}_{\rm PDE}$ denote the data loss and the PDE loss, respectively.

Once trained, the PINN can estimate the HRTF from an arbitrary direction (θ_e, ϕ_e) as $\hat{P}_{\mathrm{PI}}(x_e, y_e, z_e)$. Note that we regard the HRTFs as the sound field around the human head, and thus that the Cartesian coordinates in (11) correspond to $(r_{\mathrm{h}}, \theta, \phi)$. Below we explain the design and training of the PINN in detail:

1) **Loss**: The loss function (11) consists of two parts, the data loss $\mathfrak{L}_{\mathrm{data}}$ and the PDE loss $\mathfrak{L}_{\mathrm{PDE}}$. The data loss $\mathfrak{L}_{\mathrm{data}}$ makes the PINN output to fit the measured HRTFs, or $\hat{P}_{\mathrm{PI}}(x_q, y_q, z_q) \approx P(x_q, y_q, z_q)$ for $q \in [1, Q]$.

The PDE loss $\mathfrak{L}_{\mathrm{PDE}}$ regularizes the PINN output to conform with the Helmholtz equation, the governing equation of acoustic wave propagation, at $\{(x_d, y_d, z_d)\}_{d=1}^{D}$. This helps the PINN to generate physical amiable output at and beyond the training data. The regularization in (6), on the other hand, does not necessarily make the SH methods to generate physical amiable output as shown in the experiment section.

2) **Helmholtz equation:** As shown in (10), PINNs are multiple target optimization problems, and one would balance different loss terms with additional parameters such as λ. Although tuning the additional parameter may improve the performance of the PINN, we decide not to do so because the tuning process can be tedious [40].

Instead, we use a special form of the Helmholtz equation as the PDE loss in (11), where the $(\omega/c)^2$ term is used as the denominator for the Laplacian ∇^2 rather than as a multiplier for the PINN output $\hat{P}_{\mathrm{PI}}$. The modification makes the magnitude of the PDE loss comparable with that of data loss, and more importantly reveals a different point of view of the Helmholtz equation. That is, the Helmholtz equation can be regraded as a fitting to the HRTFs with the Laplacian.

With the magnitudes of the PDE loss and the data loss comparable with each other and without any apparent reason to prefer the data fitting or the Laplacian fitting, we simply add the two losses together without balancing them with additional parameters. In this way, the training of the PINN is greatly simplified and does result in a good HRTF upsampling performance.

3) **Training 1:** In Fig. 4 (a), Fig. 5 (a), and Fig. 6 (a), we present the normalized amplitudes of HRTFs at three different frequencies. These figures show that the amplitudes of the left parts where $\phi < \pi$ tend to be different from the amplitudes of the right parts where $\phi > \pi$. This is typical for HRTFs due to the head shadowing effect, and makes the PINN unable to estimate the two parts with the same accuracy because of their different levels of contributions to the loss (8). This fact informs us to train the PINN for the left part and right part separately, and merge the results afterwards.

4) **Training 2:** The HRTFs are complex values which are difficult to model with scalar activation functions. To simplify the training process, we train the PINN with the real part and the imaginary part of HRTFs separately, and merge the results afterwards.

5) **PINN size:** We provide guidance on the size of the PINN, specifically its width W (the number of neurons on each hidden layer) and depth L (hidden layer number).

First, we set the width W based on the dimensionality U of HRTFs under SH decomposition. For the PINN method, the width W is the number of components that are needed to model the HRTFs [41]. For the SH method, the dimensionality U is the maximum complexity of HRTFs under SH decomposition. The similar roles of the width W in the PINN method and the dimensionality U in the SH method and the fact that the HRTF is fundamentally a characteristic of the human subject irrespective of the basis function inspire us to set the width W according to U [41].

Based on our experience about SH analysis, we find two possible values for the width W: $(U+1)^2$ and $2U+1$. $(U+1)^2$ is the total number of SHs up to order U, and $2U+1$ is the number of SHs of order U [33]. The SHs form an orthogonal function set where different orders of SHs represent different levels of complexity of HRTFs [32], [33]. The $\tanh$ functions with different arguments are not orthogonal with respect to each other, and thus that less number of $\tanh$ functions together with their arguments maybe sufficient to model HRTFs. The difference between SHs and the $\tanh$ function prompts us to choose the less value of the two as the width $W = 2U + 1$. Further considering that **Training 1** and **Training 2** each halves the complexity of HRTFs, we arrive at the final choice for width W [32], [33]

$$W = \frac{2U+1}{2\times 2} \approx U/2 \approx \begin{cases} \lceil f/500 \rceil, & f \le 3\text{kHz}, \\ \lceil f/1000 \rceil, & f > 3\text{kHz}. \end{cases} \quad (12)$$

We present the width W as a function of frequency in Fig. 2 for reference, where we let $W = \max\{\lceil f/500 \rceil, \lceil f/1000 \rceil\}$ for 3 kHz $< f <$ 6 kHz [32]. Note that similar to (3) and (5), the width (12) should be regraded as a rule of thumb and is not supposed to be followed exactly. Indeed, for HRTF upsampling, we are facing a data insufficient condition, and to avoid the over-fitting problem we may need to slightly reduce the width W.

Second, with the width of the PINN set to be $W = U/2$, we find that a depth of $L = 3$ is sufficient and necessary for HRTF modeling. This maybe because the Helmholtz equation is a second order PDE which requires three conditions to specify a unique solution.

Further theoretical investigations are necessary to justify the choices and provide better guidance for the PINN design. This will be one of our future works.

V. Numerical experiments

In this section, we compare the PINN method performance with that of the SH method which has provided valuable inspirations to the design of the PINN and has been widely used for both extrapolation and interpolation scenarios.

Let (θ, ϕ) be the direction of a HRTF. We transfer the spherical coordinates $(r_{\mathrm{h}}, \theta, \phi)$ into Cartesian coordinates. We use these Cartesian coordinates as the inputs to the PINN. We normalize the amplitudes of HRTFs to be within $[-1, 1]$. We conduct upsampling on the amplitude of HRTFs rather on the magnitude to avoid potential impacts on perceptual features of

HRTFs. We evaluate the performance of all methods by the upsampling error

$$\mathcal{E}=10\log_{10}\frac{\sum_{e=1}^{E}\|P(\theta_e,\phi_e)-\hat{P}(\theta_e,\phi_e)\|_2^2}{\sum_{e=1}^{E}\|P(\theta_e,\phi_e)\|_2^2}, \quad (13)$$

where $P(\theta_e,\phi_e)$ and $\hat{P}(\theta_e,\phi_e)$ are the unknown HRTFs and are their estimations at $\{(\theta_e,\phi_e)\}_{e=1}^{E}$, respectively.

A. Extrapolation

In this section, we aim to extrapolate the unknown HRTFs whose directions are beyond those of the measured HRTFs.

We conduct experiments on HRTFs from the 3D3A dataset [38]. We set the simulated HRTFs of subject 37 left ear at 100, 200, ..., 16000 Hz as the ground truth. We use $\{P(\theta_q,\phi_q)\}_{q=1}^{1368}$ whose polar angle $0.2\pi \le \theta_q \le 0.8\pi$ as the measured HRTFs. We aim to extrapolate the unknown HRTFs $\{P(\theta_e,\phi_e)\}_{e=1}^{362}$ where $\theta_e < 0.2\pi$ or $\theta_e > 0.8\pi$.

We implement the SH method [20] following (1) - (7), and set $\gamma = 0.02$ in (6) according to a trial-and-error process.

For the PINN method, we initialize the trainable parameters with the Xavier initialization [42]. We train the PINN for 10^7 epochs with a learning rate of 10^{-5} using the ADAM optimizer. We evaluate the data loss $\mathfrak{L}_{\rm data}$ with respect to the 1368 measured HRTFs, and the PDE loss $\mathfrak{L}_{\rm PDE}$ with respect to the Cartesian coordinates of all 1368+362=1730 HRTFs.

We implement an additional method which is similar to the PINN method, except that we do not add the PDE loss $\mathfrak{L}_{\rm PDE}$ to the loss function (11). Hereafter, we refer to this method as the NN method.

We first conduct the experiment at 8 kHz. We present the real part of the ground truth HRTF in Fig. 4 (a), where the dashed lines $\theta = 0.2\pi$ and $\theta = 0.8\pi$ are the boundaries between the measured and the unknown HRTFs. Figure 4 (b) denotes the estimation obtained from the SH method with $U = 16$. Figure 4 (c) and (d) denote the estimations obtained from the NN method with $L = 3, W = 16$ and $L = 3, W = 8$, respectively. Figure 4 (e) and (f) denote the estimations obtained from the PINN method with $L = 3, W = 16$ and $L = 3, W = 8$, respectively. Figure 4 (g) and (h) denote the estimations obtained from the PINN method with $L = 2, W = 8$ and $L = 4, W = 8$, respectively.

From Fig. 4 (b), we can see that the upsampling accuracy of the SH method for the unknown HRTFs, where $\theta_e < 0.2\pi$ or $\theta_e > 0.8\pi$, is poor. From Fig. 4 (c), we can see that the NN method with width $W = 16$ estimates the unknown HRTFs with spurious values. This problem is mitigated in Fig. 4 (d) by reducing the width to be $W = 8$. However, the estimation is still not satisfying as the estimated unknown HRTFs show a curve around $\theta = 0.9\pi$ that does not present in the ground truth. Figure 4 (b), (c), and (d) reveal the problem of the SH method and the NN method. That is, they have no control over the estimations and thus can assign arbitrary values to the unknown HRTFs.

The PINN method, on the other hand, has some control over the estimations, because the PDE loss $\mathfrak{L}_{\rm PDE}$ contains the contribution from the unknown HRTFs. Figure 4 (e) shows that the PINN method with $W = 16$ tends to assign zero to the unknown HRTFs. Zero is a valid but trivial solution of the Helmtholtz equation [34]. Large area of zero value estimation indicates that the PINN method with $W = 16$ has more expressiveness than it needs to be, causing the over-fitting problem. Figure 4 (f) shows that the PINN method with an appropriate level of expressiveness, or width $W = 8$, can accurately estimate the unknown HRTFs. Figure 4 (g) shows that the PINN method with less depth $L = 2$ and hence less expressiveness, omits some details shown in the ground truth HRTFs. This is the under-fitting problem. Figure 4 (h) shows that the PINN method with more depth $L = 4$ and hence more expressiveness, suffers from the over-fitting problem as Fig. 4 (e) because it also tends to assign zero to the unknown HRTFs.

The upsampling errors of the SH method Fig. 4 (b), the NN method with $W = 16$ Fig. 4 (c), and $W = 8$ Fig. 4 (d) are 1.6 dB, -1.4 dB, and -10.8 dB, respectively. The upsampling errors of the PINN methods are -5.7 dB, -17.5 dB, -12.3 dB, and -14.5 dB for Fig. 4 (e), (f), (g), and (h), respectively.

This experiment demonstrates the superior HRTF upsampling performance of the PINN method. However, the good performance can only be achieved with the incorporation of physics knowledge and proper design of the PINN.

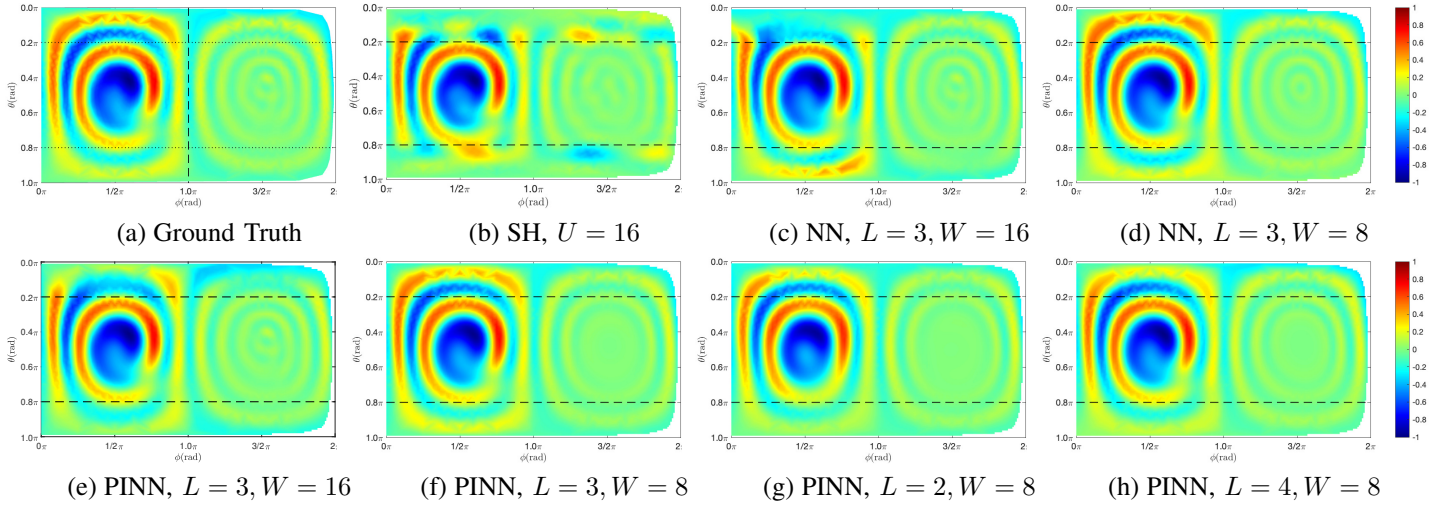
Hereafter, unless otherwise stated, we assume the U for the SH method is calculated through (3), and the depth and the width for both the NN method and the PINN method are $L = 3$ and $W = U/2$, respectively.

We next repeat the experiment on the same 3D3A HRTFs but at 16 kHz. We present the real part of the ground truth HRTF in Fig. 5 (a). Figure 5 (b), (c), and (d) denote the estimations obtained from the SH method, the NN method, and the PINN method, respectively. In this case, the SH method fails to estimate to the unknown HRTFs. The NN method, without any control over the estimation process, assigns spurious values to the unknown HRTFs. The PINN method, on the other hand, can estimate the unknown HRTFs with a better accuracy. The upsampling errors of the SH method, the NN method, and the PINN method are 3.1 dB, 5.4 dB, and -4.8 dB, respectively.

We present the HRTFs upsampling errors of the SH method and the PINN method for the 3D3A HRTFs over a broad frequency range in Table 1. (The upsampling errors of the NN method are consistently larger than that of the PINN methods, and thus are not shown.) As shown in Table 1, the upsampling errors of the PINN method is smaller than that of the SH method over the whole range. The experiment results for upsampling the imaginary part of the 3D3A HRTFs using the three methods are similar to Fig. 4, Fig. 5, and Table 1, and thus are not shown for brevity.

TABLE I
UPSAMPLING ERRORS OF THE SH METHOD AND THE PINN METHOD FOR THE 3D3A HRTFS

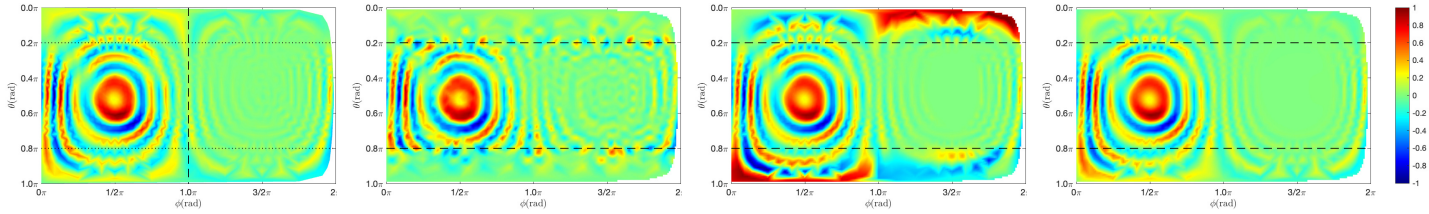
f (kHz)	2	4	6	8
SH	-34.9 dB	-12.5 dB	-1.3 dB	1.6 dB
PINN	-35.8 dB	-30.4 dB	-28.3 dB	-17.5 dB
f (kHz)	10	12	14	16
SH	1.4 dB	2.6 dB	2.3 dB	3.1 dB
PINN	-14.8 dB	-9.3 dB	-5.5 dB	-4.8 dB



(a) Ground Truth (b) SH, $U = 16$ (c) NN, $L = 3, W = 16$ (d) NN, $L = 3, W = 8$

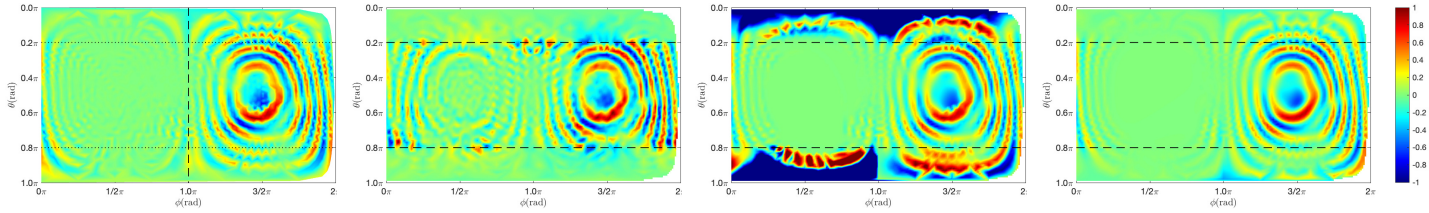
(e) PINN, $L = 3, W = 16$ (f) PINN, $L = 3, W = 8$ (g) PINN, $L = 2, W = 8$ (h) PINN, $L = 4, W = 8$

Fig. 4. Extrapolation: 3D3A dataset, HRTF 8 kHz: (a) Ground truth, (b) SH, $U = 16$, (c) NN, $L = 3, W = 16$, (d) NN, $L = 3, W = 8$, (e) PINN, $L = 3, W = 16$, (f) PINN, $L = 3, W = 8$, (g) PINN, $L = 2, W = 8$, (h) PINN, $L = 4, W = 8$.



(a) (b) (c) (d)

Fig. 5. Extrapolation: 3D3A dataset, HRTF 16 kHz: (a) Ground truth, (b) SH estimation, (c) NN estimation, (d) PINN estimation.



(a) (b) (c) (d)

Fig. 6. Extrapolation: HRSN dataset, HRTF 20.25 kHz: (a) Ground truth, (b) SH estimation, (c) NN estimation, (d) PINN estimation.

We repeat the experiment on the high resolution spherical nearfield (HRSN) dataset [39]. The dataset contains HRTFs of the Neumann KU100 dummy head which is measured at different distances. We use the right ear HRTFs measured with a loudspeaker array on a 1.0 m radius sphere at frequencies 375, 750, ..., 20250 Hz as the ground truth. We use $\{P(\theta_q, \phi_q)\}_{q=1}^{2157}$ whose polar angle $0.2\pi \leq \theta_q \leq 0.8\pi$ as the measured HRTFs. We aim to extrapolate the unknown HRTFs $\{P(\theta_e, \phi_e)\}_{e=1}^{545}$ where $\theta_e < 0.2\pi$ or $\theta_e > 0.8\pi$. The implementations of the SH method, the NN method, and the PINN method are similar to their implementations for the 3D3A HRTFs.

We show the upsampling results at a high frequency $f = 20.25$ kHz in Fig. 6, where Fig. 6 (a) shows the real part of the ground truth, and Fig. 6 (b), (c), and (d) show the estimations obtained from the SH method, the NN method, and the PINN method, respectively. Comparing Fig. 6 with Fig. 5, we can see that the upsampling result of three methods for the 3D3A HRTFs at 16 kHz and for the HRSN HRTFs at 20.25 kHz are similar. The SH method fails the upsampling task. The NN method assigns spurious values to the unknown HRTFs. The upsampling accuracy of the PINN method is the best among the three methods. The upsampling errors of the SH method, the NN method, and the PINN method are 4.6 dB, 10.7 dB, and -3.2 dB, respectively.

We present the HRTFs upsampling errors of the SH method and the PINN method for the HRSN HRTFs over a broad frequency range in Table 2. (The upsampling errors of the NN methods are larger than that of the PINN method, and thus are not shown.) Comparing Table 1 and Table 2, we can see that the upsampling errors of both methods for the measured HRSN HRTFs are larger than the corresponding errors for the simulated 3D3A HRTFs. For these two datasets,

TABLE II
UPSAMPLING ERRORS OF THE SH METHOD AND THE PINN METHOD FOR THE HRSN HRTFS

f (kHz)	2.25	4.125	6.0	8.25	10.125
SH	-12.8 dB	-3.3 dB	0.5 dB	3.6 dB	5.5 dB
PINN	-16.4 dB	-12.1 dB	-12.3 dB	-11.7 dB	-10.3
f (kHz)	12.0	14.25	16.125	18.0	20.25
SH	3.4 dB	6.6 dB	4.3 dB	5.1 dB	4.6 dB
PINN	-8.8 dB	-6.3 dB	-5.5 dB	-4.8 dB	-3.2 dB

TABLE III
UPSAMPLING ERRORS OF PINN METHOD FOR THE CHEDAR DATASET AND HUTUBS DATASET AT 10 KHZ FOR 5 SUBJECTS

CHEDAR Subject	3	18	29	57	89
PINN	-11.9 dB	-9.8 dB	-11.7 dB	-8.9 dB	-10.1 dB
HUTUBS Subject	7	35	46	78	99
PINN	-11.8 dB	-14.0 dB	-11.8 dB	-12.1 dB	-12.5 dB

the upsampling errors the PINN methods are consistently smaller than that of the SH method. The experiment results for upsampling the imaginary part of the HRSN HRTFs using the three methods are similar to Fig. 6, and Table 2, and thus are not shown for brevity.

The extrapolation experiments are further conducted on the CHEDAR dataset [44] and the HUTUBS dataset [45]. We randomly select five subjects from these two datasets, and conduct HRTF upsampling at 10 kHz. Similar to the 3D3A case and the HRSN case, we extrapolate the unknown HRTFs beyond $0.2\pi \leq \theta \leq 0.8\pi$ based on the HRTFs within the range. We present the upsampling errors of the PINN method together with subject number in Table III. As shown in Table III, for these two datasets, the PINN method can achieve upsampling errors around -10 dB and -12 dB, respectively. The upsampling errors of the PINN method on other frequencies for these two datasets are similar to the Table I and Table II, and thus are not shown for brevity.

B. Interpolation

In this section, we aim to interpolate the unknown HRTFs whose directions are between those of the measured HRTFs.

We conduct the experiment on same 3D3A HRTFs. We use HRTFs $\{P(\theta_q, \phi_q)\}_{q=1}^{1368}$ whose polar angle $0.2\pi \leq \theta_q \leq 0.8\pi$ as the ground truth. We randomly select one third (456) of the ground truth HRTFs as the measured HRTFs and the rest (912) as the unknown HRTFs. We show the arrangement of the measured HRTFs and the unknown HRTFs in Fig. 7 for reference. Note that as shown in Fig. 7, there are regions where no HRTFs are measured, and this make Fig. 7 not necessary the best arrangement for testing the performance of the PINN method. Further investigation of the optimal arrangement of the measured HRTFs is beyond the scope of this paper, and will be one of our future works.

The implementations of the SH method and the PINN method are the same as in Sec. V-A, except that for the SH method we set $\gamma = 0$ according to a trial-and-error process. We do have implemented the NN method, but the same as in

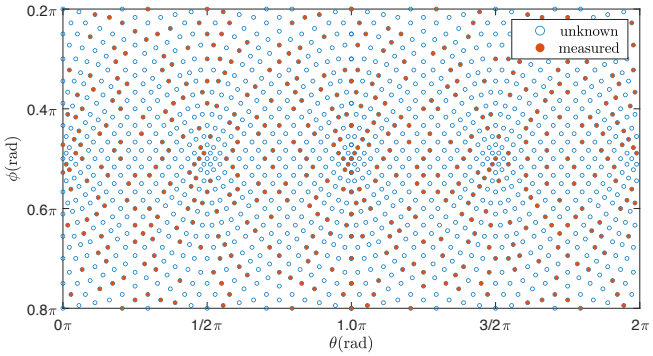


Fig. 7. Arrangement of the measured and the unknown HRTFs.

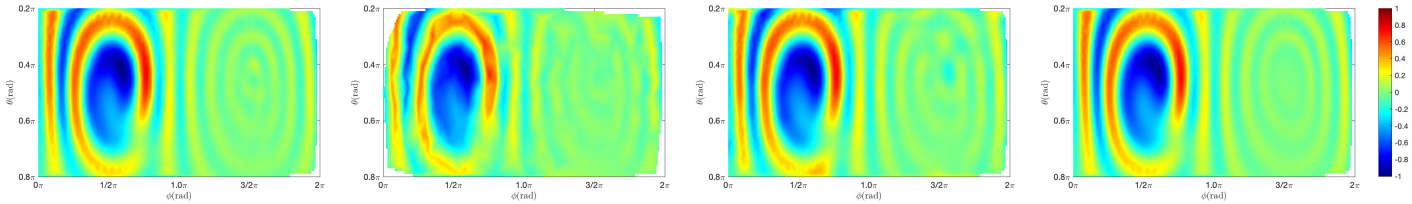
Sec. V-A, the performance of the NN method is inferior to that of the PINN method. Thus, the results of the NN method are not shown for brevity.

We compare the performance of two methods at 8 kHz in Fig. 8, where Fig. 8 (a) shows the real part of the ground truth, and Fig. 8 (b) shows the measured HRTFs. Figure 8 (c) and (d) denote the estimations obtained from the SH method and from the PINN method, respectively. In this case, the dimensionality of the HRTFs under SH decomposition is $U = 16$, which is less than $\sqrt{456} - 1$ [32], [33]. The measured HRTFs provides the SH method with enough information to accurately estimate the SH coefficients up to order $U = 16$. The upsampling error of the SH method is -14.9 dB. Comparing Fig. 8 (c) and (d), we can see that the PINN method can estimate the unknown HRTFs with a better accuracy. The upsampling error of the PINN method is -23.4 dB.

check in the figure θ and ϕ is wrong.

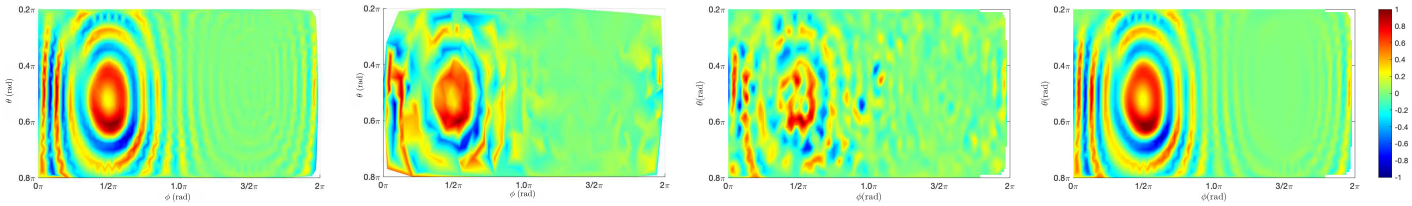
We compare the performance of two methods at 16 kHz in Fig. 9, where Fig. 9 (a) shows the real part of the ground truth, and Fig. 9 (b) shows the measured HRTFs. Figure 9 (c) and (d) denote the estimations obtained from the SH method and the PINN method, respectively. Comparing Fig. 9 (a) and Fig. 8 (a), we can see that the HRTFs at 16 kHz is much more complex then the HRTFs at 8 kHz. Figure 9 (b) shows more discontinues regions than Fig. 8 (b), and this makes it more challenging to estimate the unknown HRTFs. In this case, the dimensionality of the HRTF under SH decomposition is $U = 32$, which is more than $\sqrt{456} - 1$ [32], [33]. The measured HRTFs do not provide enough information for the SH method to accurately estimate the SH coefficients up to order $U = 32$. Fig. 9 (c) shows that the SH method estimation has some dark and bright spots which do not present in the ground truth, and the upsampling error is -0.8 dB. Fig. 9 (d) shows that the PINN method has accurately estimated the ground truth HRTFs, and the upsampling error is -14.5 dB.

We present the HRTFs upsampling errors of the SH method and the PINN method for the 3D3A HRTFs over a broad frequency range in Table IV. As shown in Table IV, the upsampling errors of the PINN method is consistently smaller than that of the SH method. The experiment results for upsampling the imaginary part of the 3D3A HRTFs using the two methods are similar to Table IV, and thus are not shown



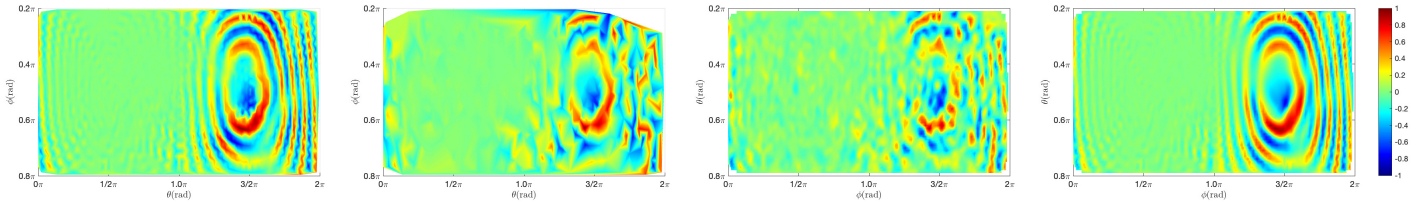
(a) (b) (d) (d)

Fig. 8. Interpolation: 3D3A dataset, HRTF 8 kHz: (a) Ground truth, (b) the measured HRTFs, (c) SH estimation, (d) PINN estimation.



(a) (b) (d) (d)

Fig. 9. Interpolation: 3D3A dataset, HRTF 16 kHz: (a) Ground truth, (b) the measured HRTFs, (c) SH estimation, (d) PINN estimation.



(a) (b) (d) (d)

Fig. 10. Interpolation: HRSN dataset, HRTF 20.25 kHz: (a) Ground truth, (b) the measured HRTFs, (c) SH estimation, (d) PINN estimation.

TABLE IV
UPSAMPLING ERRORS OF THE SH METHOD AND THE PINN METHOD FOR THE 3D3A HRTFS

f (kHz)	2	4	6	8
SH	-21.8 dB	-18.4 dB	-14.1 dB	-14.9 dB
PINN	-34.0 dB	-26.8 dB	-21.5 dB	-23.4 dB
f (kHz)	10	12	14	16
SH	-3.4 dB	0.7 dB	-0.3 dB	-0.8 dB
PINN	-18.1 dB	-15.6 dB	-14.9 dB	-14.5 dB

TABLE V
UPSAMPLING ERRORS OF THE SH METHOD AND THE PINN METHOD FOR THE HRSN HRTFS

f (kHz)	2.25	4.125	6.0	8.25	10.125
SH	-17.8 dB	-13.0 dB	-12.1 dB	-7.9 dB	-5.5 dB
PINN	-28.8 dB	-21.1 dB	-18.7 dB	-15.8 dB	-15.6 dB
f (kHz)	12.0	14.25	16.125	18.0	20.25
SH	-1.3 dB	-0.5 dB	-2.8 dB	-1.1 dB	-1.6 dB
PINN	-15.9 dB	-14.0 dB	-15.5 dB	-15.3 dB	-13.1 dB

for brevity.

We repeat the interpolation experiment on the HRSN HRTFs. We use $\{P(\theta_q, \phi_q)\}_{q=1}^{2157}$ whose polar angle $0.2\pi \leq \theta_q \leq 0.8\pi$ as the ground truth. Similar to the interpolation for the 3D3A HRTFs, we randomly select one third (719) of the ground truth HRTFs as the measured HRTFs and the rest (1438) as the unknown HRTFs. The arrangement of the measured HRTFs and the unknown HRTFs is similar to Fig. 7, and thus is not shown for brevity. The implementations of the SH method and the PINN method are similar to their implementations for the 3D3A HRTFs. We show the upsampling results at a high frequency $f = 20.25$ kHz in Fig. 10, where Fig. 10 (a) shows the real part of the ground truth, Fig. 10 (b) shows the measured HRTFs. Figure 10 (c) and (d) denote the estimations obtained from the SH method and the PINN method, respectively. The upsampling errors of the SH method and the PINN method are -1.6 dB and -13.1 dB, respectively.

We present the HRTFs upsampling errors of the SH method and the PINN method for the HRSN HRTFs over a broad frequency range in Table V. Similar to Table IV, the upsampling errors of the PINN method is smaller than that of the SH method over the broad frequency range. The experiment results for upsampling the imaginary part of the HRSN HRTFs using the two methods are similar to Table V, and thus are not shown for brevity.

The interpolation experiments are further conducted on the HUTUBS dataset [45] and the IRCAM dataset [46]. We randomly select five subjects from these two datasets, and conduct HRTF upsampling at 10 kHz. Similar to the 3D3A case and the HRSN case, we use the HRTFs where $0.2\pi \leq \theta \leq 0.8\pi$ as the ground truth. We randomly select one

TABLE VI
UPSAMPLING ERRORS OF PINN METHOD FOR THE HUTUBS DATASET AND IRCAM DATASET AT 10 KHZ FOR 5 SUBJECTS

HUTUBS Subject	5	35	47	73	89
PINN	-25.1 dB	-20.6 dB	-23.3 dB	-22.0 dB	-23.7 dB
IRCAM Subject	9	25	36	44	56
PINN	-11.9 dB	-9.8 dB	-11.7 dB	-8.9 dB	-10.1 dB

third of the ground truth as the measured HRTFs and the rest as the unknown HRTFs. We aim to estimate the unknown HRTFs based on the measured ones. We present the upsampling errors of the PINN method together with subject number in Table VI. As shown in Table VI, for the HUTUBS dataset, the PINN method can achieve upsampling errors less than -20 dB. The upsampling errors for the IRCAM dataset are around -10 dB, which is not that satisfying. This maybe because unlike other datasets which contain either simulated or measured HRTFs of artificial heads, the IRCAM dataset contains measured HRTFs of human subjects from 1680 directions. It is unclear whether human subjects can keep still over the measurement process. The upsampling errors of the PINN method on other frequencies for these two datasets show similar trends as in Table IV and Table V. That is, the upsampling errors increase with the increment of frequencies. The results are not shown for brevity.

VI. CONCLUSION

This paper proposed a PINN method for upsampling HRTFs. The performance of most existing HRTF upsampling methods is limited by the fact that they use the information of the measured HRTFs only. The proposed PINN method exploits the Helmholtz equation, the governing differential equation of acoustics, as additional information to improve the HRTF upsampling accuracy. Furthermore, based on the SH decomposition of the HRTFs and the Helmholtz equation, we set the PINN with an appropriate width and depth. This helps the PINN to avoid under-fitting and over-fitting problems. The additional information provided by the Helmholtz equation and a suitable size help the PINN to outperform the SH method in both extrapolation and interpolation scenarios.

The design of the PINN is still empirical, and we need further theoretical investigation to determine an optimal size of the PINN. Another interesting extension of this work would be incorporating even more physics information, such as head and ear geometry, into the design and training of the PINN. This will be one of our future works.

REFERENCES

[1] S. Li and J. Peissig, "Measurement of head-related transfer functions: a review", *Appl. Sciences*, 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, May 2017.

[3] 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.

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

[5] 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.

[6] 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.

[7] R. Duraiswaini, D. N Zotkin, and N. A Gumerov, "Interpolation and range extrapolation of head related transfer functions", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. iv–iv, 2004.

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] J. C. B. Torres, 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.

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

[16] 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*, 54.

[17] J. W. Lee, S. Lee, and K. Lee, "Global HRTF interpolation vi learned affine transformation of hyper-conditioned features", arXiv preprint arXiv:2204.02637.

[18] W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Iterative extrapolation algorithm for data reconstruction over sphere", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3733–3736, Mar. 2008.

[19] U. Elahi, Z. Khalid, and R. A. Kennedy, "An improved iterative algorithm for band-limited signal extrapolation on the sphere", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4619-4623, Mar. 2018.

[20] D. N. Zotkin, R. Duraiswami, N. Gumerov, et al., "Regularized HRTF fitting using spherical harmonics", *WASPAA'09. IEEE Workshop on*, pp. 256, 2009.

[21] 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 2012*, Dec. 2012.

[22] B. Rafaely, *Fundamentals of Spherical Array Processing*, (New York NY USA:Springer, 2015), vol. 8.

[23] M. Raissi, P. Perdikaris, and G. Em Karniadakis, "Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations", 2017, arXiv:1711.10566.

[24] M. Raissi, P. Perdikaris, and G. Em Karniadakis, "Physics informed deep learning (part II): data-driven discovery of nonlinear partial differential equations", 2017, arXiv:1711.10566.

[25] 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.

[26] 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", 2022, arXiv:2201.05624.

[27] 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.

[28] 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.

[29] 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.

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

[31] 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.

[32] 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.

[33] 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.

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

[35] S. Wang, Y. Teng and P. Perdikaris, "Understanding and mitigating gradient pathologies in physics-informed neural networks", arXiv:2001.04536, 2020.

[36] 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.

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

[38] R. Sridhar, J. G. Tylka, and E. Y. Choueiri, "A database of head-related transfer function and morphological measurements", *143th Conv. AES*, Oct. 2017.

[39] J. M. Arend, A. Neidhardt, and C. Pörschmann, "Measurement and perceptual evaluation of a spherical near-field HRTF set", 29 th Tonmeistertagung, 2016.

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

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

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

[43] Y. Ito, T. Nakamura, S. Koyama, and H Saruwatari, "Head-related transfer function interpolation from spatially sparse measurements using autoencoder with source position conditioning", *arXiv preprint*, arXiv:2207.10967.

[44] G. Slim, B. Xavier, and S. Renaud, "Computed hrirs and ears database for acoustic research", *148th Conv. AES*, May 2020.

[45] 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.

[46] W. Olivier, http://recherche.ircam.fr/equipes/salles/listen/download.html