Novel View Synthesis of Steering Vectors with Physically Consistent Machine Learning

Diego Di Carlo, *Member, IEEE,* , Shoichi Koyama, *Member, IEEE,* Arie Aditya Nugraha, *Member, IEEE,*, Mathieu Fontaine, *Member, IEEE,* Yoshiaki Bando, *Member, IEEE,* Kazuyoshi Yoshii, *Member, IEEE.*

***Abstract*—Steering vectors encoded the sound propagation between a sound source and a microphone array. They play a crucial role in acoustic front-end processing for automatic speech recognition as well as for creating personalized acoustics virtual environments. They are core objects in sound source localization, enhancement and binaural rendering algorithms. ...**

***Index Terms*—Physics-informed neural networks (PINN), spherical harmonics, head-related transfer function (HRTF), Gaussian process, spatial audio.**

I. Introduction

AUGMENTED listening encompasses technologies designed to enhance human auditory perception by modifying real-time soundscapes [1]. These technologies span conventional hearing aids, consumer-grade smart headphones, assistive listening devices, and emerging platforms in extended reality (XR) such as smart glasses and head-mounted displays. Augmented listening serves as the auditory parallel to augmented reality (AR), where visual overlays enhance user perception of the physical world.

Recent innovations in XR platforms, including Microsoft HoloLens 2, Meta Quest 3, and ARIA smart glasses [2], exemplify the potential of immersive XR interactions. Despite advancements in visual and tactile modalities, spatial audio processing—a critical component for immersive experiences—remains underexplored [3]. Audio integration is paramount, not only for a realistic sense of immersion [4] but also for improving accessibility [5].

Applications of augmented listening often involve creating *personalized sound zones* by manipulating sound fields through addition, removal, or modification of sources, tailored to individual preferences [6]. A typical augmented listening pipeline includes spatial audio capture (e.g., Ambisonics encoding, spatial filtering), analysis and decomposition (e.g., sound source localization and separation), and reproduction (e.g., binaural synthesis and HRTF-based rendering) [7]. For example, the SPEAR challenge [8] demonstrated the efficacy of a multichannel minimum variance distortionless response (MVDR) beamformer combined with spatial filtering for augmented reality scenarios, underscoring the importance of accurate steering vectors and noise covariance estimation [9].

Steering vectors represents the interaction of sound waves with microphone arrays [10]. They are foundational in spatial audio processing, enabling tasks such as speech enhancement [11], sound source localization [12], and sound scene synthesis [13]. While traditionally modeled as algebraic formulations in free-field scenarios, practical applications often require handling complex acoustic environments, accounting for generic acoustic propagation effects as for acoustic impulse responses (AIRs), indoor reverberation as for room impulse responses (RIR), or the anatomical effects as in head-related transfer functions (HRTFs).

In this work we consider steering vectors as multi-channel mathematical quantities encoding the anechoic sound propagation from a location to a reference point at given frequency. Specifically, we assume a multi-channel spatio-temporal representation of the sound field, expressed as a collection of acoustic impulse responses (AIRs) measured on an aperture which constitutes the spatial domain of interest, for instance the a sphere surrounding a listener head. This representation is most commonly adopted in sound field synthesis and Ambisonics reproduction, where the listener and the valid sound field are delimited by a space enclosed by a loudspeaker array with far-field sources contributing to the overall reconstructed sound pressure [14]. Note that this definition allows use consider both HRTF and multi-channel measurement of a sound field as steering vectors.

Algebraic steering vectors analytically encode the direct propagation of an acoustic transfer function or an impulse response over space and frequency or time, respectively. However, for in-the-wild scenarios, their algebraic model is limited by several impairments, and often replace with more general filters. Extended formulations include models for microphone directivity and filtering as well as sound interaction with the receiver, such as occlusion, diffraction, and scattering. In hearing aids application, the steering vectors, also known as head-related transfer function (HRTF), model the filtering effects of the user's pinnae, head, and torso.

Source code available here

This work was supported by ANR Project SAROUMANE (ANR-22-CE23-0011) and Hi! Paris Project MASTER-AI, JST PRESTO no. JPMJPR20CB, and JSPS KAKENHI nos. JP20H00602, JP21H03572, JP23K16912, JP23K16913.

Diego Di Carlo and Aditya Arie Nugraha are with the Center for Advanced Intelligence Project (AIP), RIKEN, Tokyo 103-0027, Japan (e-mail: diego.dicarlo@riken.jp; adityaarie.nugraha@riken.jp)

Shoichi Koyama is with National Institute of Informatics, Tokyo, Japan (e-mail: XXXXXXXXXXX).

Mathieu Fontaine is with LTCI, T ´el ´ecom Paris, Institut Polytechnique de Paris, France (e-mail: mathieu.fontaine@telecom-paris.fr).

Yoshiaki Bando is with the National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, 135-0064, Japan (e-mail: y.bando@aist.go.jp).

Kazuyoshi Yoshii is with the Center for Advanced Intelligence Project (AIP), RIKEN, Tokyo 103-0027, Japan, and XXXXXXXXX (e-mail: yoshii@i.kyoto-u.ac.jp).

(Corresponding author: Diego Di Carlo).

Manuscript received XXXX XX, XXXX; revised XXXX XX, XXXX.

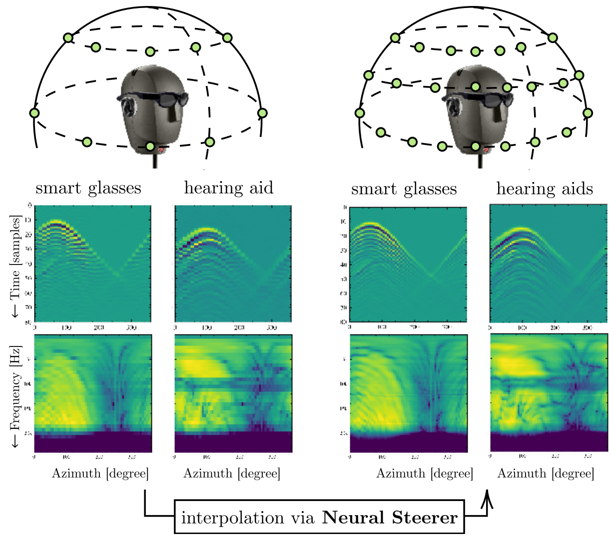

Fig. 1. Visual abstract of continuous steering vector modeling

Steering vectors can be broadly categorized based on the acoustic conditions they encode. In anechoic free-field environments, steering vectors are derived from closed-form expressions knowing the array geometry [10] or relative time-differences of arrival (TDOA) in far-field scenarios [15]. This formulation offer a simple, fast and differentiable computation of the steering vectors, a geometric interpretation, valid in every environment. However it does not model acoustics reflection, which are regarded as noise term in further downstream tasks.

For scenarios involving non-free-field anechoic environments, steering vectors account for additional factors such as scattering objects and microphone directivity (a.k.a., pick-up) patterns. A prominent example is the Head-Related Transfer Function (HRTF), which captures the spatial filtering effects caused by human anatomical features, such as pinnae, head and torso. Such vectors are typically estimated through either in-lab measurements or computationally-heavy simulators.

Beyond anechoic conditions, general AIR-based steering vectors represent complex propagation environment that can be computed using model based on numerical acoustic simulators or fast simulators integrating geometric acoustics, stochastic approaches or hybrid methods [16]. While these representations provide precise spatial information, their applicability is often constrained to environments where acoustic properties are well-characterized.

The methods to obtain steering vectors depend on the above condition and available resources. Algebraic approaches and some simulators offer computational efficiency, especially in simple anechoic environments, but require detailed knowledge of the array geometry and environmental parameters. Such methods, while fast and differentiable, may struggle to replicate real-world acoustic profiles, being sensitive to microphone positioning and speed of sound variations [10, Chapter 6.6]. Estimation-based methods treat steering vector acquisition as an inverse problem, utilizing tools like blind source separation [] or relative transfer function (RTF) estimation [11, Section VI.B.3]. Despite progress in this area, robust estimation remains a significant challenge in audio signal processing [17], [18].

Alternatively, steering vectors can be directly measured, which provides the most accurate representation of the actual acoustic environment [13]. However, such measurements are expensive, time-consuming, require expert oversight, and are highly dependent on spatial sampling resolution [19]. Additionally, measured vectors are often stored in lookup tables, making them susceptible to dataset-specific conventions and preprocessing requirements [20].

To reduce the time require and the complexity of measured steering vector setup and to make the method scalable, spatial upsampling have been proposed to generate high-resolution steering vectors These methods are commonly referred to as upsampling, interpolation, regression or super-resolution methods, especially in the field of HRTF synthesis and sound field reconstruction. Several methods have been proposed to spatially upsampling measurements improving the spatial resolution of arrays with fewer sensors or offering a cost-effective solution without compromising performance. Considering the steering vectors and this is basically a curve fitting problem.

In this work, we propose a novel method for upsampling measured steering vectors from limited observations, formulated as a Gaussian Process (GP) regression problem. By modeling the relationship between spatial-frequency coordinates and multichannel steering vector measurements, our approach ensures a continuous representation across the channel dimension. This method leverages strong physical priors, striking a balance between soft and hard physics-based constraints to effectively address data scarcity. Notably, our model simultaneously accounts for both the magnitude and phase of steering vectors, ensuring a comprehensive and accurate representation. We further perform an ablation study to assess the contributions of various components within the architecture, providing insights into the design choices that drive its performance.

The evaluation focuses on the challenging task of upsampling steering vectors from sparse measurements, addressing scenarios encountered in augmented listening pipelines. We present a thorough analysis of the proposed method, benchmarking it against state-of-the-art approaches, including Physics-Informed Neural Networks (PINNs), across frequency and spatial dimensions. Our results demonstrate the model's ability to achieve high fidelity in both angular and frequency domains, significantly outperforming baseline methods. Additionally, we extend the evaluation to downstream tasks such as beamforming using real world data, showcasing the method's utility in end-to-end augmented listening pipelines. These contributions position our approach as a robust solution for steering vector upsampling in both theoretical and applied contexts.

The rest of the paper is organized as follows. Section II presents the state of the art in sound field reconstruction and spatial upsampling. Section III formulate the signal models and the problem of interest. Section V presents the proposed models of continuous steering vectors. Section VII reports some implementation details and presents a series of exper-

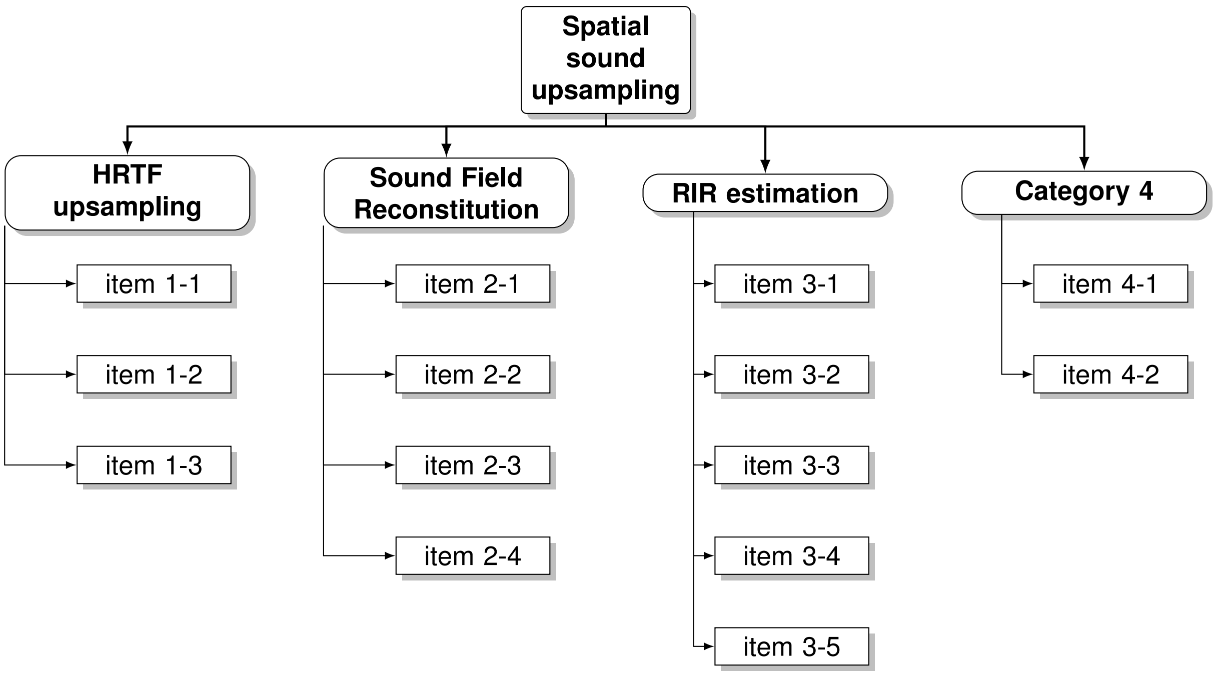

Fig. 2. Caption

iments conducted with the the proposed models and their comparison with several state-of-the-art spatial upsampling methods. This includes analysis for sound field reconstruction and the downstream task in speech enhancement. Section VIII concludes the paper.

II. Related works

To upsample acoustics measurement around the human head, various techniques have been developed that have been propose in the field of HRTF upsampling [], microphone pickup and source directivity estimation [], sound field reconstruction []. HRTF upsampling methods however commonly focus on the magnitude of a minimum-phase representation of the HRTF, whereas the phase is reconstructed and compensated by algebraically computed phase difference. HRTF upsampling methods consider the contribution of the 2 ears equal or independent. HRTF upsampling methods usually preprocess the measurements. Sound field reconstruction methods instead focus on more general use cases. Sound filed always uses physical models. Sound field reconstruction methods may uses specific array geometry (spherical microphone array).

A. HRTF upsampling

Authors of [21] recently reviewed the state-of-the-art techniques of for HRTF measurement interpolatio, while [22] compared nearest neighbors methods with interpolation carried out in spherical harmonics domain concluding comparable performances ...

. In such work, the rich literature is categorized in 3 classes: nearest-neighbor approaches that return a weighed combination of neighboring measurement, functional methods that models a mathematical function of frequency and direction, and methods based on neural networks. Here we propose a different classification based on how much explicit physical knowledge is employed to super-resolve HRTF measurements: data-driven and knowledge-driven approaches, the latter being subdivided in physics-driven and approximated parametric methods. HRTF representaion also influnces interpolation performances: the results of [23] indicated that a separate interpolation of the magnitude and unwrapped phase of the HRTF performed better than an interpolation of the complex spectrum. This fact, also confirmed in [22], applied for the task of virtual acoustic environments, not for other downstream task, such as beamforming, as investigate in this work.

a) Data-driven methods: these methods relies only on information contained in the observations, in prepared training sets or provided from data coming from multi-modalities. Data-driven methods could be local measurements (linear, bilinear, trilinear, barycentric) interpolation or exploit data-driven knowledge from a global set of measurements using methods like PCA, wavelets, or DNN.

Weighted interpolation is the most straightforward approach. This method typically assumes noiseles measurements acquired on a regular grid. These methods has been shown to produce a sufficiently good agreement between measured and interpolated HRTFs when a relatively large number of measurements are still present [24]. In the case where the low-resolution HRTF contains 320 or more source positions, it is preferable to use barycentric interpolation; [25] As explained in [25], the acoustic measurement [26], [27] is still considered the gold standard of these different approaches. The downside to performing acoustic measurements is the expensive custom setup required and the time it takes. This method has been shown to produce good results when the HRTFs contain a relatively large number of IRs [24], for example, with an angular distance of 10–15° between measurements; however, it becomes much less reliable when interpolating sparser measurements (e.g. each 30–40°) [25]. Methods in this category are uses inverse distance [23], bilinear interpolation [28], [29], tetrahedral interpolation [24], barycentric interpolation in spherical interpolation [30] or natural neighbor interpolation [22]. The authors of [31] extends the this methods to non-uniform grid. This methods have been recently evaluated with real data measured on a Kemar mannequin in [21], [32]. According to [21] the best performing methods is the bilinear interpolation to upsample ... not reported??.

Subspace methods aims to reduce the dimensionality of a data set while retaining the primary variation in the data, rather than rather than improve interpolation in case of sparse measurements [22]. In this way, fewer coefficients are interpolated instead of the HRTF itself. Principal Components Analysis (PCA) is a statistical algorithm for deriving spectral shape basis functions and decomposing HRTFs (see [33, Chapter 6.2] for a review). Authors of [34] proved that PCA of the resulting 5300 HRTF magnitude functions revealed that the HRTFs can be modeled as a linear combination of five basic spectral shapes (basis functions), and that this representation accounts for approximately 90% of the variance in the original HRTF magnitude functions. Later works such as [35], [36] that interpolated via bilinear interpolation or multivariate polynomial filtering [37]. Authors of [38] introduce the Spatial PCA which was later used by [39] in conjunction with a deep neural network that process anthropometric features. The work of [35] tackle the problem in continuous settings proposing a model based on the Karhunen-Loéve transform, while in [40] proposed the interpolation of wavelet coefficients.

Deep Neural Network have been extensively used for HRTF upsampling. This approach is appealing because of good generalization, multi-modal learning and quick inference. A recent review is provided in [21] and [7]. In contrast with local methods, DNN are able to extract both global and local properties (features) that can be used for a fast inference. In the HRTF literature, we can identify the following approaches: DNN architectures, like UNet, inspired from the one used in image super-resolution or inpainting, usually change 2D-CNN layer with TCN. [25], [41]–[43]; or architecture that uses other modalities, such as anthropometric features [44], [45], or image of the pinnae [46], [47] with autoencoders models. Finally, a recent trends that focus on upsampling sparse measurements utilize Generative Adversarial Networks for spatial upsampling of HRTF [25], [43], [48]. Another approach exploint the natural interpolation effect due to the spectral bias of coordinate-based neural network [49]. These architecture, called *neural fields*, have been recently proposed in computer vision to model physical *fields*, such as radiance field for novel view synthesis of 3D objects (e.g., in NeRF) and physical quantities (e.g. PINNs) [50]. Such approach has been used also for HRTF interpolation [45], implicit auralization of audio signal [42] and different HRTF datasets with different grid conventions across different subjects [20].

b) Knowledge-driven methods: methods of this group leverage on prior knowledge to super-resolve measurements. The problem is typically formulated as a regression problem whose solutions are constrained by a parametric model or regularized. Methods in this group can be further classified as geometric-based methods that uses geometric reasoning to weights the interpellant coefficients [51]–[59], DSP-based methods that approximate the shape of HRTF's spectrum with simple fiters whose parameter space can be smoothly interpolated [60]–[68]. Finally, Physics-driven methods use the Helmholtz equation or its free-field parameterized solution, e.g. the Green's function, to regularize [69] or constraint the solution [70]–[80].

Geometric-based methods uses simplified spatial model to achieve fast interpolations, assuming the spatial distribution of the directional HRTF on a sphere centered on the user head. The most popular methods is the Vector-Based And Panning (VBAP) [51] which is used in ISO/IEC MPEG-H 3D Audio standard for reproducing immersive audio coding with multiple loudspeakers. This method creates spatialized sound by distributing the audio signal between loudspeakers in a way that gives the impression of a sound source being at an arbitrary position. By considering HRTF as directional audio data, VBAP can be employed to spatialization HRTF using the known measurement as anchor points. VBAP is considered as a local panning technique, because it only drives a small number of loudspeakers (at most three) close to the target direction, as opposed to global panning techniques such as Ambisonics amplitude pannning, which is based on approximated physical modeling, as explaied below. Authors of [52] proposed an extension of VBAP, recasting it as an l1 optimization problem applied to global panning. Alternative global geometric methods relies of smooth interpolation on a sphere [53], [54] using spherical thin-plate splines [81]. This approach was used by [55] to interpolate the PCA coefficient instead of the actual measurements for fast global interpolation. Interestingly, the authors of [56] propose a shallow neural network extending the RBF basis function to represent spherical data, accepting the quering direction as input and returning the value of HRTF for given frequency. The proposed "von Mises Basis Function" is the natural constraint of periodicity and singularity at the poles. This idea could be considered as a precursor of the neural field approach presented below. Gaussian Process have been also applied to continuously model HRTF over direction and frequencies [57]. In particular the authors proposed a stationary covariance function based on a kernel based on the chordal distance to model the sources and a inverse-quadratic kernel function to model the correlation among frequencies, which model of an exponentially decreasing process in the time-domain. To our best knowledge this is the only work that explicitly consider a continuous model (and smoothness) over frequency. This approach was later extended to aggregate heterogeneous HRTF dataset in [82], a task that is later studied with neural fields in [20]. Finally, in the deep learning community, some studies focused on extending CNN layer to accomodate the spherical geometry of HRTF data [58], [59]. Spherical convulutional layers have been proposed to executes rotation-equivariant feature transforms. Interestingly, [59] place the upsampling neural network along side with Neural Gaussian Process to provides uncertainty estimates on the upsampled regions and such estimates could inform the sequential decision problem of acquiring as few correcting HRTF data points as needed to meet a desired level of HRTF individualization accuracy. While the authors claims promising results, this method in inherently discrete, performing upsampling on a given resolution. While this is not issues in real life application, the problem is that such approaches require annotated training data. Interestingly, this work analyses the performances for very few observation, within 5 to 100 ranges.

DSP-driven methods use signal processing techniques and properties to reduce the complexity of the problem. Most of the works model HRTF as cascade of zeros and/or poles filters [60]–[64] to model the peak and notches of HRTFs spectrum. Noticing smoothness in the variation of such parameters with respect to direction, they apply interpolation over the space of the filter's parameters. Later this techniques was extended to more expressive IIR filters with parametric representation [65], [68]. In particular, the authors of [68] propose a neural field for the prediction of such coefficients, using then a cascade of differentiable IIR filters in the spirit of Differentiable DSP neural architecture [83].

Physic-driven methods models regards the HRTF as the sound field around the human head. Physics-driven methods can be broadly subdivided in two categories: physics-constarained and physics-informed methods. The former methods aims at the reconstruction of the sound fields typically relying on the interpolation of the projection of the measured set onto a linear combination of spatial basis functions, a linear combination of some spherical harmonics or plane waves that satisfy of the Helmholtz equation. Therefor the solution is constrained to be in a specific phiscal space. These basis functions represent sound propagation in a homogeneous medium where each sound field component is unknown and obtained as part of an optimization problem. This is commonly brought about as a plane wave expansion, considered an implicit, truncated solution to the homogeneous Helmholtz. A sound field can also be then approximated as a finite sum of spherical harmonics. Estimation of the weighting coefficients of the basis function is achieved by direct integration of the measurements [71] or solving regularized least-square optimization problems [72]–[75] with spherical basis function or using pre/post-processing techniques to further reduce the number of measurements needed for good interpolation with LS fitting by spherical harmoncs [76]–[78]. As the SH basis functions form a spatially continuous set of solutions of the wave equation, an interpolation in the SH domain yields a physically correct and spatially continuous HRTF representation as long as $N \geq \kappa r$ [84], with $\kappa = \omega c$, ω the frequency and c the propagation velocity of sound. Meaning that a minimum of Consequently, the interpolation of sparse HRTF sets, i.e., HRTF sets which are measured with a low spatial resolution results in an incomplete description of the spatial and spectral properties and leads to order-limitation artifacts affecting high-frequency components and binaural cues, Effect of truncation order error in [85]. Sampling scheme of the measurement on the sphere also affects the results. While equiangular, Gaussian, Lebedev, or Fliege schemes yelds to similar results [86], it is know that random spare measurement lead to a poor interpolation performances [77]. Also, it is a specific property of SH interpolation that by transforming the sampled functions to the SH domain, errors or inaccuracies of one measured point on the sphere unavoidable affect the entire SH representation [22]. Therefore, Regularization or more advance optimization techniques are required, for instance, the coefficients can be predicted by a neural network [**?**] or optimized using Bayesian variational inference [80], or preprocessing, e.g. time-aliment [76], [87] or directional equalization to compensation linear phase component and diffraction [75], [77], [78]. See [78], [87] for reviews. To reduce spatial aliasing and order truncation errors.

Finally the physics-informed methods uses physics as data augmentation or "soft" regularization terms as a bias to drive the solution towards a good balance between data and physics using the paradigm of Physics-informed Neural Network [69].

B. *Sound Field Reconstruction*

The objective of sound field reconstruction problem is to continuously reconstruct the pressure field in a given position within a target region [88] to that it is possible to reproduce sound signal at arbitrary position. The applications spans from VR/AR, interpolation of acoustic transfer function, acoustic imaging (or holography), and active noise control [89]. Because of its deep connection with physics and the strict hi-fi requirement, the majority of the work uses prior physical knowledg. An exception is made for few pure DNN-based works based on the recent development in deep learning, e.g. [90]–[94] inspired from models of image super-resolution in computer vision. The work in this field can be broadly classified into two [95]: non-parametric or expansion-based models [88], [96]–[101] and parametric models [102]–[110] and a traversal hybrid approach based on deep learning [14], [88], [90]–[95], [111]–[119] and Gaussian Processes [101], [120]–[122]. The formers aims at estimating the acoustic field

C. *Other related field of research*

a) Source directivity pattern estimation:

b) Spatial RIR upsampling: and generation that can be purely data-driven [123], [124], [125]. Maybe Mathieu can help here.

Problems related to steering vectors upsampling:

- HRTF. pro: similar data, accounts for anthropomorphic features and perceptual evaluation. Magnitude and phase, make the learning unstable. Two channel are considered independent. Mostly consider only Magnitude interpolation, phase is reconstructed with IPD for minimum phase filters.
- Source directivity upsampling: [126], [127].
- Microphone directivity (antenna radiation) pattern upsampling: [**?**], [96], [115]: cons: use of spherical microphones.
- Sound field reconstruction: pro:; cons: limited to single frequencies; no dependencies between frequencies; only few works focus on scattering objects. Channel are independent. Many microphones for one source, or many sources for one microphones.

- Spatial (early) RIR upsampling: cons: mostly focus on the distribution of early echoes and the global RIR shape. pro: the multichannel dependencies. Many microphones.
- Up to the current knowledge of the authors there a no current techniques specific to steering vectors upsampling.

Common approaches to tackle such problems. We can broadly classify on as data-to-prior-knowledge spectrum. Data-based methods only leverage on the information contained in the data or other modalities whose dependency are difficult to model in closed-form. Knowledge-driven methods helps to improve accuracy, generalization, or address scarcity of data or noisy observation. Purely data-driven methods, like nearest neighbors are preferable in case of dense noise-free observation. While smoothness enforced my knowledge-based method can improve accuracy for unseen data, the output of this methods can detrimentally affect downstream application (such as inverse filtering for beamforming).

As explained in [25], the acoustic measurement [26], [27] is still considered the gold standard of these different approaches. The downside to performing this acoustic measurement is the expensive custom setup required and the time it takes.

Therefore, in the case where the low-resolution HRTF contains 320 or more source positions, it is preferable to use barycentric interpolation; [25]

Most of the works, operates with magnitude of HRTF, reconstructing the minimum phase response [128].

Binaural reproduction relies on the head-related transfer function (HRTF) which represents the scattering effect of human anatomy with respect to the direction of sound. [27] An analysis of HRTFs indicates that the functions exhibit periodicity in amplitude along the azimuthal angle [129].

A direct practical application of HRTFs is the creation of virtual acoustic environments. In this case, the input signal of a sound source is filtered with the HRTF corresponding to a given desired source position. When the resulting ear signals are presented to a listener – typically using headphones – the sensation of a sound source in the desired position is evoked.

c) HRTF upsampling: From the literature of HRTF (which offer some models spanning the entire data-to-knowledge-driven spectrum). Most of the this works focus on magnitude of HRTF only of minimal phase representation, phase reconstructed later knowing the ITD.

- Data-based models: model train from data only
  - Interpolation: natural neighbors: see [22] for a comparative review and different HRTF representation, weighted interpolation [?], [23], [24], [28], [30]–[32], [130]. This method has been shown to produce a sufficiently good agreement between measured and interpolated HRTFs when a relatively large number of measurements are still present [24]. This method has been shown to produce good results when the HRTFs contain a relatively large number of IRs [24], for example, with an angular distance of 10–15° between measurements; however, it becomes much less reliable when interpolating sparser measurements (e.g. each 30–40°).
  - Subspace : PCA/KL expansion [?], [29], [34]–[39], Wavelets [40]
  - DNN with side-information: autoencoders [?], [44], GANs [?], [?], [43], [48], CNN [?], [41]–[43], [47] Deep Believe Network [131]
  - DNN with Neural Fields (natural interpolation thanks to the spectral bias) [?], [20], [42], [45], [66],
  - Manifold Learning [132]–[134]
  - DNN anthropometric [39], [45], or aggregating multiple subjects [20], [25], [42], [79]. Estimating implicitly from speech observation [42], [66].
  - works that uses only HRTF magnitude [20], [45], [134]. Study of the phase [130]
  - works that focus on very sparse measurements [25], [38], [135]
  - local [51] vs global interpolation [52]
- Model-based models: improve interpolation using a-priori knowledge
  - Geometrical smoothness: panning-based methods [51], [52], Spherical thin-plate splines [?], [53]–[55] using spherical splines [81], Radial basis function neural network [56] Gaussian processes with chordal distance [57], Spherical extension of CNN [?], [59].
  - DSP: IIR filters [?], [?], [60]–[65].
  - Pattern matching: [?], [135]
  - Physics
    - Informed: [?], [?] using PINNs [136], [137].
    - Constrained: regularized linear regression, azimuthal harmonics [70] SH [71]–[78], [138] kernel ridge regression using SH, DNN [?] Bayesian [80]
      [78], [87] reviews for pre- and post- processing HRTF techniques for the SH harmonics for order reduction, such as [75]–[78]. To reduce spatial aliasing and order truncation errors. Effect of truncation order error in [85].

d) Sound Field Reconstruction: From the literature of SFR (which is more focused of the physical reconstruction of the field). It is related to the literature of room impulse response estimation, but here we focus on the problem of interpolation at unknown position. [?] for a review. Parametric solution: On the one hand, parametric solutions [11–15] rely on simplified parametric models of the sound field. Non-parametric solution: On the other hand, non-parametric methods [16–18,20–26] aim to numerically estimate the acoustic field. see [93] for a recent review.

- Natural interpolation [96]
- IIR/FIR [102]
- DSP / sparse approximation approach [?], [103]–[108] using modeling of the sound, typically using plain wave decomposition.
- Regularized Linear Regression with SH [97], [98] and PWD [99], [139]
- Kernel Ridge Regression for PWD [88], [100], extended to GP in [101].
- Parametric model [125], [140]
- Statistical form of the scattering field [141]
- DNN [90]–[93]

	Data-driven methods	Knowledge-driven methods				
		Heuristic models		Physics model		
		Math	DSP	Physics-informed	Physics-consistent	Physics-constrained
HRTF upsampling	**Exemplar-based methods:** Nearest neighbors [review in [22]] **Smooth interpolation:** Bilinear interpolation [?], [?], [22]–[24], [30] **Subspace methods:** PCA [36], [37], [39] Wavelets [40]					
Sound Field Reconstruction						
RIR upsampling						

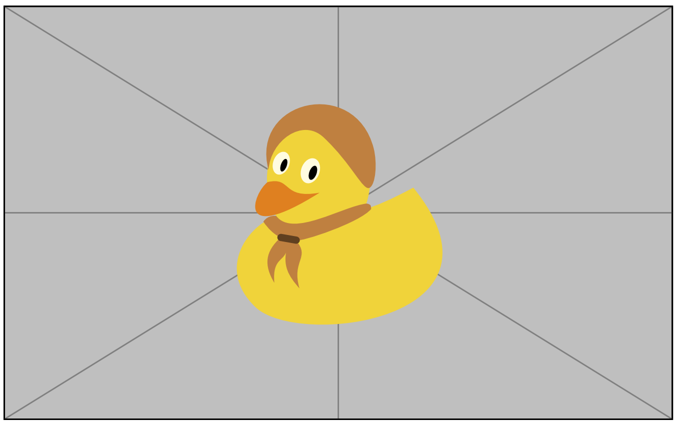

Fig. 3. Measurement grid and reference system used in this work

- Physics-informed Neural Network [?], [?], [?], [?], [113], [114], [116], [127]
- Deep Neural Operator, DeepONet) [?], [117]
- Physics-constrained Neural Network [?], [14], [88], [119]
- Optimal transport [124]
- Diffusion models [94]
- GP regression for RIR reconstruction [?], [?], [?], [121], [122]

In some application, it is preferable to have measure the trade-off between the observed data and the interpolation. Methods like linear and splines interpolation allows for perfect recall of the "training" data.

III. Problem formulation

A. Free-field sound propagation

In the frequency domain, the homogeneous Helmholtz equation describes the evolution of the complex acoustics pressure field $h \in \mathbb{C}$ as a function of position $\mathbf{q}$ and the angular frequency ω as

$$\nabla^2_{\mathbf{q}} h(\mathbf{q}, \omega) + \frac{\omega^2}{c^2} h(\mathbf{q}, \omega) = 0, \tag{1}$$

where $\nabla^2_{\mathbf{q}}$ is the 3-dimensional Laplacian operator and c is the speed of sound with respect to the space. This equation is linear with respect to h, implying that the pressure field is the sum of the pressure fields resulting from multiple sound sources.

Assuming free space propagation, a single frequency point source at position $\mathbf{s} \in \mathbb{R}^3$ emits a pressure wave in the form of

$$h(\mathbf{q}, \omega \mid \mathbf{s}) = \frac{1}{\sqrt{4\pi r}} e^{-\jmath\omega r/c} \tag{2}$$

where $r = \|\mathbf{q} - \mathbf{s}\|_2$ is the distance between the source and the measurement location and $\jmath = \sqrt{-1}$. This equation is the solution of equation (1) in ideal free-field propagation, which is also known as *Green's function*, or free-field *acoustic impulse response* in the signal processing vocabulary.

Therefore, the acoustic sound field $x(\mathbf{q}, \omega)$ measured at $\mathbf{q}$ produced by a sound source emitting a signal $s(\mathbf{s}, \omega)$ at location $\mathbf{s}$, can be computed as

$$x(\omega, \mathbf{q}) = h(\omega, \mathbf{q} \mid \mathbf{s}) s(\omega, \mathbf{s}). \tag{3}$$

B. Spherical harmonics representation

The sound pressure field can represented as a linear combination of basis functions. In the case of free-field sound propagation, spherical harmonics are solutions of the PDE in (1) and are commonly used to describe a pressure field. A function H(, ,) that is square integrable on the surface of a sphere that is centered around the coordinate origin can be represented by the coefficients H nm() of a series of spherical harmonics Ynm(,) as [142]. In the present case, we may apply the Helmholtz reciprocity principle and assume that we observe the sound pressure of a sound source in the ear of the subject on the surface of a sphere [8].

$$x(\omega, \mathbf{q}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm}(\omega, \mathbf{q}_0) \bar{Y}_l^m(\omega, \mathbf{q} - \mathbf{q}_0) \tag{4}$$

where $\kappa = \omega/c$ is the wave number. Here c_{lm} are the expansion coefficient of order l and degree m, $\mathbf{q}_0$ is the expansion center, and $\bar{Y}_m^l$ are modified spherical harmonics of order l and degree m.

Depending on the nature of the sound field of interest, the modified spherical harmonics can be further developed as follows

$$\bar{Y}_l^m(\omega, \mathbf{q}) = \begin{cases} Y_m^l\left(\frac{\mathbf{q}}{\|\mathbf{q}\|_2}\right), & \textit{in general case} \\ h_l^1(\kappa\|\mathbf{q}\|_2)Y_l^m\left(\frac{\mathbf{q}}{\|\mathbf{q}\|_2}\right), & \textit{exterior field} \\ j_l^1(\kappa\|\mathbf{q}\|_2)Y_l^m\left(\frac{\mathbf{q}}{\|\mathbf{q}\|_2}\right), & \textit{interior field} \end{cases} \tag{5}$$

where $\bar{Y}_m^l$ are spherical harmonics of order l and degree m accepting as argument the azimuthal and the polar coordinate of the unit vector.

The surface spherical harmonics $Y_l^m(\cdot)$ are a complete and orthonormal set that can be defined as

$$Y_n^m(\theta, \phi) = (-1)^m \sqrt{\frac{(2l+1)}{4\pi}\frac{(n-|m|)!}{(n-|m|)!}} P_n^{|m|}(\cos\phi)e^{\jmath m\theta} \tag{6}$$

where $P_l^m(\cdot)$ denotes mth-order the associated Legendre function of n-th degree, ϕ denotes the azimuth and θ the colatitude.

Steering vectors encode the acoustic transfer function from a sound source to a microphone. From a j-th sound source at position $\mathbf{s}_j$ to the i-th microphone at position $\mathbf{m}_i$, in the frequency domain, the algebraic model for anechoic steering vector at frequency f is expressed as

$$svect_{ij}(f) := d_{ij}(f) = \exp(\jmath\omega_f r_{ij}/cF_s)/\sqrt{4\pi r_{ij}^2} \tag{7}$$

We consider the measurements of the steering vectors as sound pressure describing a continuous sound field.

Let us consider a spherical microphone array of radius r composed of I microphones located a $\mathbf{m}_i = [x_i, y_i, z_i]^T \in \mathbb{R}^3$, or, equivalently, in the spherical coordinates $\bar{\mathbf{m}}_i = [r, \theta_i, \varphi_i]$. where θ and φ are the polar (angle with respect to positive polar axis) and the azimuthal coordinate.[1]

IV. Sound field regression

Let $y_n := y(\mathbf{z}_n)$ denote a noise measurement of the sound field h generated from a single source, such that

$$y_n = h(\mathbf{z}_n) + \varepsilon_n \tag{8}$$

where ε_n models noise and $\mathbf{z}_n = [\omega_n, \mathbf{q}_n] \subset \mathbb{R} \times \mathbb{R}^3$ is the vector of measurement collocation point, frequency and space. Let $\mathbf{y} = [y(\mathbf{z}_1), \dots, y(\mathbf{z}_N)]^\mathsf{T} \in \mathbb{C}^{FI}$ and $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]\mathsf{T} \in \mathbb{R}^{N\times 4}$ be the vector of measurements the sound field measured and coordinates at F frequencies and I positions, respectively. Given $\mathbf{y}$ and $\mathbf{Z}$, we here consider the problem of estimating the underling continuous function h, or, similarly, the reconstruction of the sound field at another frequency and position $\mathbf{z}_*$. We will refer to this problem as regression, that is interpolation in presence of noise [**?**].

The estimation of the continuous function $f : \mathbb{R}^P \to \mathbb{K}$ ($\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$) from discrete observation $\mathbf{y} \in \mathbb{K}^N$ at the sampling points $\{\mathbf{z}_i\}_{n=1}^N$ is achieve by representing f with some model with parameters $\boldsymbol{\theta}$ and solving the following optimization problem

$$\arg\min_{\boldsymbol{\theta}} \mathcal{L}\left(\mathbf{y}, \mathbf{f}(\{\mathbf{z}_i\}_{n=1}^N; \boldsymbol{\theta})\right) + \mathcal{R}(\boldsymbol{\theta}) \tag{9}$$

where $\mathbf{f}(\{\mathbf{z}_i\}_{n=1}^N; \boldsymbol{\theta}) = [f(\mathbf{z}_1; \boldsymbol{\theta}), \dots, f(\mathbf{z}_N; \boldsymbol{\theta})]^\mathsf{T} \in \mathbb{K}^N$ is the vector of the discretized function f represented by $\boldsymbol{\theta}$. $\mathcal{L}$ is a loss function evaluating the distance between $\mathbf{x} andf$ at $\{\mathbf{z}_i\}_{n=1}^N$, and $\mathcal{R}$ is a regulation term for $\boldsymbol{\theta}$ to prevent overfitting.

A. Regression with basis expansion

$$f(\mathbf{z}; \boldsymbol{\gamma}) = \sum_{l=1}^{L} \boldsymbol{\gamma}_l \psi_l(\mathbf{z}), \tag{10}$$

where $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_L]^\mathsf{T} \in \mathbb{K}^L$ and $\psi(\mathbf{z}) = [\psi_1, \dots, \psi_L]^\mathsf{T} \in \mathbb{K}^L$, for instance the spherical wave function expansion.

If the squared error loss function and a ℓ_2 penalty are used, (9) yield the following closed-form solution

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \|\mathbf{y} - \boldsymbol{\Psi}\boldsymbol{\gamma}\|_2^2 + \lambda\|\boldsymbol{\gamma}\|_2^2 \tag{11}$$

$$= \left(\boldsymbol{\Psi}^\mathsf{H}\boldsymbol{\Psi} + \lambda\mathbf{I}\right)^{-1}\boldsymbol{\Psi}^\mathsf{H}\mathbf{y}, \tag{12}$$

where $\boldsymbol{\Psi} = [\psi(\mathbf{x}_1), \dots, \psi(\mathbf{x}_N) \in \mathbb{K}^{N\times L}$, $\mathbf{I}$ is the identity matrix and $\cdot^\mathsf{H}$ denotes Hermitian transposition. Then, f is a linear combination of the basis function $\{\psi_l\}_l$ by construction.

B. Kernel regression with RKHS

$$f(\mathbf{z}; \boldsymbol{\alpha}) = \sum_{n=1}^{N} \alpha_n k(\mathbf{z}, \mathbf{z}_n) \tag{13}$$

$$= \mathbf{k}(\mathbf{z})^\mathsf{T}\boldsymbol{\alpha}, \tag{14}$$

where $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]^\mathsf{T} \in \mathbb{K}^N$ and the weight coefficient are $\mathbf{k}(\mathbf{z}) = [k(\mathbf{z}, \mathbf{z}_1), \dots, k(\mathbf{z}, \mathbf{z}_N)]^\mathsf{T} \in \mathbb{K}^N$ is the vector of kernel functions. In the kernel ridge regression [], the estimated of $\boldsymbol{\alpha}$ is compute as

$$\hat{\boldsymbol{\alpha}} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y}, \tag{15}$$

with the Gram matrix $\mathbf{K} \in \mathbb{K}^{N\times N}$ defined as

$$\mathbf{K} = \begin{bmatrix} k(\mathbf{z}_1, \mathbf{z}_1) & \dots & k(\mathbf{z}_1, \mathbf{z}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{z}_N, \mathbf{z}_1) & \dots & k(\mathbf{z}_N, \mathbf{z}_N) \end{bmatrix} \tag{16}$$

Then f is interpolated by substituting $\hat{\boldsymbol{\alpha}}$ in (14)

C. Gaussian process interpolation

A Gaussian process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution. Given any finite set of n input $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$ and the corresponding set of latent function values $\mathbf{y} = \{y(\mathbf{z}_1), \dots, y(\mathbf{z}_n)\}$, the relationship between the input data $\mathbf{x}_n$ and the observed noisy target y_n are given by

$$y_n = f(\mathbf{z}_n) + \varepsilon_n, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{17}$$

where ε is the zero-mean Gaussian noise with variance σ^2.

The prior distribution over the latent function can be written as

$$p(\mathbf{f} \mid \mathbf{Z}) \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}) \tag{18}$$

[1]Spherical coordinates follows the in ISO 80000-2:2019 convention.

where the $\boldsymbol{\mu} = m_{\boldsymbol{\theta}}(\mathbf{Z}) \in \mathbb{K}^N$ is the mean vector computed with the mean function $m(\cdot)$ and $\mathbf{K}$ is the Gram matrix whose element $\mathbf{K}_{nn'} = k_{\boldsymbol{\theta}}(\mathbf{z}_n, \mathbf{z}_{n'})$ is based on the kernel function. Both the mean and the kernel function may have hyper-parameters, denoted here as $\boldsymbol{\theta}$. The kernel function $k(\cdot,\cdot)$ controls the smoothness of the GP.

For a The predictive distribution, also known as posterior distribution, of the function values $\mathbf{f}_\star$ at the test set $\mathbf{z}_\star$ is

$$p(\mathbf{f}_\star \mid \mathbf{z}_\star, \mathbf{y}, \mathbf{Z}) \sim \mathcal{N}(\boldsymbol{\mu}_\star, \boldsymbol{\Sigma}_\star) \tag{19}$$

where the mean $\boldsymbol{\mu}_\star$ and the covariance $\boldsymbol{\Sigma}_\star$ are calculated as

$$\boldsymbol{\mu}_\star = m(\mathbf{z}_\star) + \mathbf{k}_\star^\mathsf{T}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - \boldsymbol{\mu}) \tag{20}$$

$$\boldsymbol{\Sigma}_\star = k(\mathbf{z}_\star, \mathbf{z}_\star) - \mathbf{k}_\star^\mathsf{T}(\mathbf{K} + \sigma_n^2\mathbf{I})\mathbf{k}_\star \tag{21}$$

where $\mathbf{k}_\star = [k(\mathbf{z}_\star, \mathbf{z}_1), \dots, k(\mathbf{z}_\star, \mathbf{z}_n)]$ is the vector of covariances between the test point and the N training points.

Kernel functions for sound field reproduction: The GP prior covariance function encodes the assumed constraints on the latent function h. In particular the correlation between any subset of points, that is, the smoothness of the interpolation, is fully specified by the GP prior as function over the input domain and hyper-parameters. In [57] the joint spatial-frequency covariance function is spefied throught single GP covariance as the product of a Ornstein-Uhlenbec density (suitable to model exponentially-decaying process, or the continuous-time analogue of the discrete-time auto-regressive AR1 process) and and exponential covariance function of the chordal distance. The associated Gram matrix for a measurement set as a Cartesian outer-product $\mathbf{Z} = \mathbf{Z}^{(\omega)} \times \mathbf{Z}^{(\mathbf{q})}$ reads $\mathbf{K} = \mathbf{K}^{(\omega)} \otimes \mathbf{K}^{(\mathbf{q})}$. Choise for the covariance matrix

In [57]

$$k = k_\omega(\omega_f, \omega_{f'}) k_\mathbf{q}(\varphi_i, \varphi_j, \theta_i, \theta_j) \tag{22}$$

$$k_\omega(\omega_f, \omega_{f'}) = \exp\left(\frac{|\omega_f - \omega_{f'}|^2}{2\ell_\omega^2}\right) \tag{23}$$

$$k_\omega(\omega_f, \omega_{f'}) = \frac{\alpha_f}{\ell_\omega^2 + (\omega_f - \omega_{f'})^2} \tag{24}$$

$$k_\mathbf{q}(\varphi_i, \varphi_j, \theta_i, \theta_j) = \exp\left(-\frac{C_{ij}}{\ell_\mathbf{q}^2}\right) \tag{25}$$

$$k_\mathbf{q}(\varphi_i, \varphi_j, \theta_i, \theta_j) = \left(1 + \frac{\sqrt{3}C_h}{\ell}\right) \exp\left(-\frac{\sqrt{3}C_h}{\ell}\right) \tag{26}$$

$$k_\mathbf{q}(\varphi_i, \varphi_j, \theta_i, \theta_j) = \sum_{l=0}^{\infty} b_l \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_l^m(\varphi_i, \theta_i) {Y'}_l^m(\varphi_j, \theta_j) \tag{27}$$

where $C_h = 2\sqrt{\sin^2\left(\frac{\theta_j - \theta_i}{2}\right) + \sin\theta_i \sin\theta_j \sin^2\left(\frac{\varphi_i - \varphi_j}{2}\right)}$ is the chordal distance. In general, the length-scale parameters $\ell_{(\cdot)}$ is the the distance for function values to become uncorrelated.

a) Choice of the collocation points:

D. Regression with Neural Fields

A neural field $\mathcal{F}_{\boldsymbol{\theta}} : \mathbb{R}^d \to \mathbb{R}, \mathbf{z} \mapsto y$ is a neural network that maps points $\mathbf{z}$ to the function value y. Let be $\hat{y}(\cdot; \boldsymbol{\theta}) = \mathcal{F}_{\boldsymbol{\theta}}(\cdot)$ the network returned value. The network parameters $\boldsymbol{\theta}$ are commonly optimized by minimizing the loss function

Input layer	$\gamma(\mathbf{z}) = [\sin(2\pi\mathbf{B}\mathbf{z}), \cos(2\pi\mathbf{B}\mathbf{z})]$	Random Fourier Feature
	$\gamma(\mathbf{z}) = [\sin(2\pi\mathbf{W}_1\mathbf{z} + \mathbf{b}_1)]$	SIREN
	$\gamma(\mathbf{z}) = [\sin(2\pi\mathbf{W}_1\mathbf{z} + \mathbf{b}_1)]$	SF space
Output layer	$y = \mathbf{g}[1] + \jmath\mathbf{g}[2]$	Real-imaginary output encod
	$y = \exp(\mathbf{g}[1]\|) \exp\left(\jmath 2\pi \arctan(\frac{\mathbf{g}[2]}{\mathbf{g}[3]})\right)$	Magnitude-phase output enc
	$y = g$	Magnitude-phase output enc

TABLE I
CAPTION

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{n=1}^{N} (y_n - \hat{y}(\mathbf{z}_n; \boldsymbol{\theta}))^2 + \mathcal{R}(\boldsymbol{\theta}) \tag{28}$$

where the first term is the empirical risk function and $\mathcal{R}$ is a regularization function that prevents overfitting.

Natural signals, such as shapes, images and sounds, contains rich high-frequency content. Due to spectral bias, standard neural network (e.g., MLP architectures), fails to learn high-frequency function from low dimensional data [], and generate blurry or over-smooth version of the target quantity. To address this issues two main approaches have been proposed:

a) Random Fourier Features: the composition $\mathcal{F}_{\boldsymbol{\theta}} \circ \gamma$ of a neural field and a random Fourier feature (RFF) encoding γ helps to overcome the spectral bias, enabling the neural network of represent signals with high-frequency components. The RFF encoding $\gamma : \mathbb{R}^P \to \mathbb{R}^{2D}$ with $D \gg P$ is defined as

$$\gamma(\mathbf{z}) = [\sin(2\pi\mathbf{B}\mathbf{z}), \cos(2\pi\mathbf{B}\mathbf{z})], \tag{29}$$

where $\mathbf{B} \in \mathbb{R}^{D\times P}$ whose elements are randomly drawn from the normal distribution $\mathcal{N}(0, \sigma_{\text{RFF}}^2)$. RFF features are not optimized during training.

b) SIREN network: Sinusoidal activation functions have been recently proposed as an effective alternative way to overcome the spectral bias. The proposed model share the same fundamental structure of a MLP, for which the r-th layer reads

$$\phi_r(\mathbf{z}_r) = \sin(g_r \mathbf{W}_r \mathbf{z}_r + \mathbf{b}_r) \tag{30}$$

where $mbz_r, \mathbf{W}_r, \mathbf{b}_r, g_r$ are the input vector, the weight, the biases and a hyperparameter of the the r-th layer, respectively. Finally, the SIREN architecture is a composition of L layers,

$$\mathcal{F}_{\boldsymbol{\theta}}(\mathbf{z}) = (\phi_R \circ \phi_{R-1} \circ \cdots \circ \phi_1)(\mathbf{z}). \tag{31}$$

drawback: ripple effects, tuning hyper-parameters

c) Sinusoidal feature (SF):

$$\gamma(\mathbf{z}) = [\sin(2\pi\mathbf{W}_1\mathbf{z} + \mathbf{b}_1)], \tag{32}$$

In both the approaches, the hyper-parameter offers a trade-off between reconstruction fidelity and overfitting. It has been shown that it is related to the bandwidth of the target function and it must be tuned accordingly.

Given the set $\{h_n\}_{n=1}^N$, with $h_n \in \mathbb{C}$, measurements of the sound field at angular frequency and location $\{(\omega_n, \mathbf{q}_n)\}_{n=1}^N \subset \mathbb{R} \times \mathbb{R}^3$, a NF, $\mathcal{F}_{\boldsymbol{\theta}_{\text{NF}}}$, can be used for regression problem. The loss function use to train the network reads

$$\mathcal{L}_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{n=1}^{N}\left|h_n - \hat{h}(\omega_n, \mathbf{q}_n; \boldsymbol{\theta})\right|^2. \tag{33}$$

In general, at test time, the networks can always evaluate continuous any new coordinate $(\omega_\star, \mathbf{q}_\star) \in \mathbb{R} \times \mathbb{R}^3$.

d) NF regularization techniques:

- Dataset: number of frequencies
- Architecture: number of neurons and layers
- Regularization loss
- Weight decay

e) Multiplication by geometric steering vectors: : time-aligned spectrum is composed of the minimum-phase and nonlinear phase all-pass components of the HRTF. [143]

E. Physics-informed Neural Networks

As any acoustic field, this field must comply with the Helmholtz (wave) equation, and the natural basis for representing it is the set of the elementary solutions of that equation. These elementary solutions are a product of spherical harmonics (theta,phi) and of spherical Hankel functions in range, which become constant if the range is fixed. The remaining spherical harmonics basis is well-suited for HRTF representation; an advantage of such representation is that the interpolation results are guaranteed to be physically valid, which can not be said about other ad hoc (e.g., spline-based) HRTF interpolation methods

A common PINNs consider a feed-forward MLP network for modeling a dynamical function u of a physical system in a space $\mathbf{z} \in \Omega$, with networks parameter $\boldsymbol{\theta}$ to be optimized. In general, u mathematically obeys known priors, such as PDE of the general form

$$\mathcal{M}_z[u(z)] = 0, \quad z \in \Omega \tag{34}$$

$$\mathcal{B}[u(z)] = d(z), \quad z \in \partial\Omega \tag{35}$$

where $\mathcal{M}_z[\cdot]$ is a general combination of nonlinear differential operator, which can include any combination of derivative with respect the input variable z, such as the first- and second-order derivative $\frac{\partial u}{\partial z}$ and $\frac{\partial^2 u}{\partial z^2}$, respectively. The boundary operator $\mathcal{B}[\cdot]$ enforces the desired condition $d(z)$ at the boundary $\delta\Omega$. In case of the wave equation in the frequency domain $\mathcal{M}_{\mathbf{q},\omega} = \nabla^2_{\mathbf{q}} - \frac{\omega^2}{c^2}$.

The training loss function of a PINN extends the regular loss function in (28) with PDE-based regularization

$$\mathcal{R}(\boldsymbol{\theta}) = \lambda_{\text{PDE}}\mathcal{L}_{\text{PDE}} + \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} \tag{36}$$

$$\mathcal{L}_{\text{PDE}} = \|\mathcal{M}_z[\hat{u}(z;\boldsymbol{\theta})]\|^2 \qquad z \in \Omega \tag{37}$$

$$\mathcal{L}_{\text{IC}} = \|\mathcal{B}_z[\mathcal{F}_{\boldsymbol{\theta}}(z)] - d(z)\|^2 \qquad z \in \partial\Omega. \tag{38}$$

The relative weight, $\lambda_{(\cdot)}$ control the trade-off between different components and need to be scaled depending on the problem at hand. The physical loss components $\mathcal{L}_{\text{PDE}}$ and $\mathcal{L}_{\text{IC}}$ are defined over the continuous domain Ω, but in practice, they are computed over a finite set of collocation points that must be sampled, for example, on a uniform grid. The computation of differential operators can be conveniently computed via automatic differentiation.

In case of sound field reconstruction regression, the loss function use do optimize the internal parameters $\boldsymbol{\theta}_{\text{PINN}}$ of the PINN reads

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\theta}} = \frac{1}{N}\sum_{n=1}^{N}\Big(h_n - \hat{h}(\omega_n, \mathbf{q}_n; \boldsymbol{\theta})\Big)^2 \\ + \lambda_{\text{PDE}}\frac{1}{M}\left(\nabla^2\hat{h}(\omega_n, \mathbf{q}_n; \boldsymbol{\theta}) + \frac{\omega_n^2 R_n^2}{c^2}\hat{h}(\omega_n, \mathbf{q}_n; \boldsymbol{\theta})\right)^2.\end{aligned} \tag{39}$$

Similarly to the NFs, a PINNs can evaluate any continuous coordinate at test time.

1) PINNs configurations:

- Spherical coordinates vs. Cartesian configuration: In the proposed PINN method, the input coordinates are expressed in the Cartesian system. Spherical coordinates are numerical unstable due to the $\sin\theta$ term in the denominator **[?]** and the due no variation of the radial direction (distance) for the computation of the first- and second-order radial gradient.
- PINNs size:
- Frequency-wise upsampling: as input we consider a subset of frequency
- Loss balancing: re-arranging the terms in the PDE so the regularization terms have the same physical unit as the data-loss hence, leading to a more simple scaling. Sampling strategy (highest scales, batch size, how often a new batch)

V. Proposed Models

A. The inwards model

- We want to upsample the sound field on the sphere around the head (assuming that the mics are on a sphere of radius $r \approx 0.1$ m)
- the spherical harmonics depends on $\mathbf{m}$
- if frequency 24kHz, then $\kappa r = \omega r/c = 44$ order $\rightarrow$ 1980 coeffs
- if frequency 8kHz, then $\kappa r = \omega r/c = 14$ order $\rightarrow$ 210 coeffs

The first model regards the sound pressure h as a sum of incident and scattering fields h^{inc} and h^{scat} **[?]**, as

$$h_{ij}(f) = h^{inc}(x_{ij}, f) + h^{scat}(x_{ij}, f) + \epsilon \tag{40}$$

where

$$h^{inc}(x_{ij}, f) = \sum_{lm} c_{lm}^{inc}(x_{ij,f})\psi^{inc}(\mathbf{m}_i, f) \tag{41}$$

$$= svect_{ij}(f) \tag{42}$$

$$h^{scat}(x_{ij}, f) = \sum_{lm} c_{lm}^{scat}(x_{ij}, f)\psi^{scat}(\mathbf{m}_i, f) + \epsilon \tag{43}$$

where $\sum_{l,m} = \sum_{l=1}^{\infty}\sum_{m=-l}^{l}$ and x_n is a shorthand for $(\mathbf{m}_i, \mathbf{s}_j)$. In practice, this summation can be computed only up to a certain order $L < \infty$, resulting in a truncation of the expansion of the sound field. The summation order L depends on the sampling scheme of the measurements

[], and the wavenumber κ_f. First, generally, a minimum of $N = (L+1)^2$ observation per frequency is needed to resolve the expansion of order L. Second, a band-limited expansion leads to negligible aliasing if $\kappa_f r \leq L$ [].

We dubbed this model "inwards" to suggest the presence of the incident sound wave. Note that the spherical harmonics basis functions for the scattering field are computed with respect to the microphone positions.

The basis functions are

$$\psi^{inc}(\mathbf{m}_i, f) = j_l^1(\kappa_f r) Y_l^m(\bar{\mathbf{m}}_i) \tag{44}$$

$$\psi^{scat}(\mathbf{m}_i, f) = h_l^1(\kappa_f r) Y_l^m(\bar{\mathbf{m}}_i) \tag{45}$$

where j_l^1 and h_l^1 are the n-th order spherical Bessel and Hankle functions of the first kind, respectively. $\kappa_f = \omega_f / c$ is the wave number at the angular frequency ω_f and speed of sound c. For a point on the unit 2D sphere $\mathbf{p} \in \mathbb{S}^2$, the spherical harmonic $Y_l^m(\bar{\mathbf{p}})$ is defined as

$$Y_l^m(\bar{\mathbf{p}}) = Y_l^m(\theta, \varphi) = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{lm}(\cos\varphi) e^{\jmath l \theta}. \tag{46}$$

Comments:

- microphone position in the basis function
  $\rightarrow$ no modulation / naive interpolation at test time when querying new source position.
- svect is the mean of the GP process
  $\rightarrow$ elegant way to include the svect algebraic model
- fewer coefficients are needed for resolving the function on the sphere.
  $\rightarrow$ Exp decay $\implies$ smoothness

This can be translated in the following GP

$$h_{ij}(f) = GP(svect_{ij}(f), k_{inw}((x_n, f), (x_{n'}, f')) \tag{47}$$

We assume the following form for the covariance

$$k_{inw}((x_n, f), (x_{n'}, f)) = k^{RBF}(f, f') k^{SPH}(x_n, x_{n'}) + \sigma_\epsilon^2 \tag{48}$$

where

$$k^{RBF}(f, f') = \exp(-|f - f|/2\ell_f^2) \tag{49}$$

$$k^{SPH}(x_n, x_{n'}) = g(f, x_n) g(f, x_{n'}) \tag{50}$$

$$g(f, x_n) = \sum_{lm} c_{lm}(x_n, f) \psi_{lm}^{scat}(\mathbf{m}_i) \tag{51}$$

where $\cdot^*$ denotes complex conjugation.

B. Outwards model

- we want to upsample the pressure on the sphere of the source position with $R = 1.5$ meter
- Assuming reciprocity, we model the entire field as an expanding sound wave
- the spharm basis functions depends on $\mathbf{s}$ only
- if frequency 24k Hz, then $\kappa * R = \omega / c \cdot R = 659$ order $\rightarrow$ 434940 coeffs.
  if frequency 8k Hz, then $\kappa * R = \omega / c \cdot R = 220$ order $\rightarrow$ 48620 coeffs.
  $R = 1.5$ meter is the distance of the sources from the center.

Assuming reciprocity, we model the entire field as the exterior sound field satisfying the homogeneous Helmholtz equation, which the spherical wave-function expansion can approximately represent as

$$h(f, x_n) = \sum_{lm} c_{lm}(f, x_n) \psi_{lm}^{scat}(f, \mathbf{s}_j) + \varepsilon \tag{52}$$

$$= GP\left(0, k(x_n, x_n')\right) \tag{53}$$

or alternatively

$$h(f, x_n) = \sum_{lm} c_{lm}(f, x_n) \psi_{lm}^{scat}(\mathbf{s}_j) + \varepsilon \tag{54}$$

$$= svect(f, x_n) + \sum_{l \geq 2, m} c_{lm}(f, x_n) \psi_{lm}^{scat}(f, \mathbf{s}_j) + \varepsilon \tag{55}$$

$$= GP\left(svect(f, x_n), k(x_n, x_n')\right) \tag{56}$$

Note that the basis functions are computed with respect to the sound source position $\mathbf{s}_j$, which are assumed to be placed on the sphere of radius R meter. Since this model model is an expanding wave, it is dubbed "outwards".

Within this model, we assume the covariance with the following form:

$$k(x_n, x_n') = k_{freq}^{RBF}(f_l, f_l') k^{SPH}(x_n, x_n') \tag{57}$$

with

$$k^{SPH}(x_n, x_n') = g(x_n) g^*(x_{n'}) \tag{58}$$

$$g(x_n) = \left(\sum_{lm} c_{lm}(x_n) \psi_{lm}(\mathbf{s}_i) \right) \tag{59}$$

1) Model with convolution: When modeling measured steering vectors across different microphones, some natural equivalence arises. To account for symmetries due to rotations of the microphones and source position, we propose to model the sound field $\bar{h}$ at a reference point $\mathbf{r} \in \mathbb{R}^3$, which is spatialized via the theoretical steering vectors.

$$h(f, x_n) = svect(f, x_n) \bar{h}(\mathbf{r}, f) \tag{60}$$

$$= svect(f, x_n) \sum_{lm} c_{lm}(f, x_n) \psi_{lm}^{scat}(\mathbf{s}_i) + \epsilon \tag{61}$$

$$= \sum_{lm} svect(f, x_n) c_{lm}(f, x_n) \psi_{lm}^{scat}(\mathbf{s}_i) + \epsilon \tag{62}$$

$$= svect(f, x_n) + \sum_{\geq 2, m} \frac{c_{lm}(f, x_n)}{svect(x_n)} \psi_{lm}^{scat}(\mathbf{s}_i) + \epsilon \tag{63}$$

$$= svect(f, x_n) + \sum_{l \geq 2, m} c_{lm}'(f, x_n) \psi_{lm}^{scat}(\mathbf{s}_i) + \epsilon \tag{64}$$

This model led to the following GP formulation

$$h(f, x_n) = GP(svect(f, x_n), k((f, x_n), (f', x_{n'})). \tag{65}$$

Model name	Regressor	Basis functions	Mean function	Covariance function	Soft constraint	Hard constraint	Require training	Independen
LRR	Linear Regression	Spherical harmonics	N.A.	N.A.	Tickonov	Basis functions	No	freqs, mics
SP	Linear Regression	Spherical splines	N.A.	N.A.	Tickonov	Basis functions	No	freqs, mics
GP	GP	Spherical harmonics	0	k	Smoothness	Basis functions	Yes*	
	GP	Spherical harmonics	svect	k	Smoothness	Basis functions	Yes*	
NF	Neural Field	N.A.	N.A.	N.A.	Smoothness	N.A.	Yes	
PINN	Neural Field	N.A.	N.A.	N.A.	PDE	N.A.	Yes	
Inwards	Neural Field + GP	Spherical harmonics	0	k	Smoothness	Basis functions	Yes	
	Neural Field + GP	Spherical harmonics	svect	k	Smoothness	Basis functions	Yes	
Outwards	Neural Field + GP	Spherical harmonics	0	k	Smoothness	Basis functions	Yes	
			svect	k	Smoothness	Basis functions	Yes	

VI. Continuous processing with Neural Fields

A. Inwards model

We proposed to use a coordinate NN (neural field) to estimate the expansion coefficients, that is

$$c_{lm}(x_j, f) \leftarrow DNN(\mathbf{s}_j, f). \tag{66}$$

Ideally, the expansion coefficients depend only on the source and the frequency, but we also explore the following extended formulation:

$$c_{lm}(x_j, f) \leftarrow DNN(\mathbf{m}_i, \mathbf{s}_j, f). \tag{67}$$

We can add a prior of the coefficients c_{lm}.

- For now, just a simple ℓ_1 norm
- Later, an exponential decay for $c_l = \sum_m c_{lm}$ with respect to l. This implies smoothness [].

The spherical harmonic spectrum SHS! is defined in Pollack et al. 1993! as A(m(Umn 2 1Vmn 2)/(2n11) for each degree n."[Evans et al., 1998, p. 2403]

Number of harmonics = 10 (should be 16) [Evans et al., 1998, p. 2403]

B. Neural Field-based Gaussian Process

C. Implementation

D. Architecture and training

E. Model selection

Choice of the validation set

- considering the validation set as a random subset of a training set in the form `nObs` $\times 3$ does not work due to the unbalanced proportion between $F = 512$ and $D = 8$. By simpling doing this, the model is biased towards a smooth interpolation over the frequency axis, leading to an always-decreasing validation loss function.
- Alternatively, one can use one full directional observation, but it prevents the model from using the knowledge about that data point during training.
- We opt for an intermediate way, where we frame the observation along the frequency axis in N frames, and for each DOA, we consider a percentage P for validation. This method for splitting the training and validation dataset is illustrated in Figure **??**.
- the optimal values for N and P are chosen with cross-validation considering the downstream task.

Choice of the validation metrics

- Using the MSE as a stopping criterion was found not beneficial. The relative improvement of this metric does not reflect an improvement in the relevant metrics of the downstream task.
- As shown in Figure **??**, the only metric whose behavior is correlated to the downstream task is the SI-SDR.

VII. Experiments

A. Baselines

- **Nearest Neighbour**
- **Spherical Spline**
  - Implementation adapted from MNE [81]
  - parameters to tune: smoothness, number of Legendre terms, stiffness.
- **Regularized Linear Regression** as in [**?**] with **Spherical Harmonics**
  - parameters to tune: spherical harmonics order (automatically tuned, aka balanced), smoothing coefficient
  - parameters not to tune: spherical harmonics order (automatically tuned for having a determined system to solve)
- **Neural Field**
- **PINN**
  - parameter to tune: optimizer (learning rate), PDE regularization term,
  - parameter not to tune: heuristics as network dimension (as in [**?**])
- **Gaussian Processes** regression with **Spherical Harmonics kernel**
  - parameter to tune: the scale of RBF along freqs, spherical harmonics order along sources, scale of RBF along mics
  - manually tuned or optimized with NLL

B. Implementation

complex MLP

- compute posterior on collocation points. Heuristic to select frequencies
- backpropagable complex Cartesian spherical harmonics. To avoid convention ambiguities.
- asymptotic Hankel function

C. Ablation Study

D. Interpolation task

a) Performances across angles:

b) Performances across frequencies:

E. Speech enhancement downstream tasks

Let I be the number of microphones attending to J sound sources. In the frequency domains, the mixture model writes

$$x_i = \sum_{j=0}^{J-1} h_{ij} s_j \tag{68}$$

where h_{ij} encode the acoustic transfer function from the j-th source to the i-th microphones.

F. Discussion

G. Conclusion

H. Future work

RTF vs ATF, frequency as latent variables

VIII. Conclusion

The conclusion goes here.

Acknowledgments

This should be a simple paragraph before the References to thank those individuals and institutions who have supported your work on this article.

Appendix
Proof of the Zonklar Equations

Use `\appendix` if you have a single appendix: Do not use `\section` anymore after `\appendix`, only `\section*`. If you have multiple appendixes use `\appendices` then use `\section` to start each appendix. You must declare a `\section` before using any `\subsection` or using `\label` (`\appendices` by itself starts a section numbered zero.)

References

References

[1] R. M. Corey, "Microphone array processing for augmented listening," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2019.

[2] J. Engel, K. Somasundaram, M. Goesele, A. Sun, A. Gamino, A. Turner, A. Talattof, A. Yuan, B. Souti, B. Meredith *et al.*, "Project aria: A new tool for egocentric multi-modal ai research," *arXiv preprint arXiv:2308.13561*, 2023.

[3] G. Kailas and N. Tiwari, "Design for immersive experience: Role of spatial audio in extended reality applications," in *Design for Tomorrow—Volume 2: Proceedings of ICoRD 2021*. Springer, 2021, pp. 853–863.

[4] E. H. A. De Haas and L.-H. Lee, "Deceiving audio design in augmented environments: a systematic review of audio effects in augmented reality," in *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2022, pp. 36–43.

[5] J. Herskovitz, J. Wu, S. White, A. Pavel, G. Reyes, A. Guo, and J. P. Bigham, "Making mobile augmented reality applications accessible," in *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, 2020, pp. 1–14.

[6] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 81–91, 2015.

[7] M. Cobos, J. Ahrens, K. Kowalczyk, and A. Politis, "An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2022, no. 1, p. 10, 2022.

[8] V. Tourbabin, P. Guiraud, S. Hafezi, P. A. Naylor, A. H. Moore, J. Donley, and T. Lunner, "The spear challenge-review of results," in *Proc Forum Acusticum*, 2023.

[9] S. Hafezi, A. H. Moore, P. Guiraud, P. A. Naylor, J. Donley, V. Tourbabin, and T. Lunner, "Subspace hybrid beamforming for head-worn microphone arrays," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[10] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2002.

[11] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.

[12] G. Chardon, "Theoretical analysis of beamforming steering vector formulations for acoustic source localization," *Journal of Sound and Vibration*, vol. 517, p. 116544, 2022.

[13] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Applied Sciences*, vol. 7, no. 5, p. 532, 2017.

[14] X. Karakonstantis and E. Fernandez-Grande, "Generative adversarial networks with physical sound field priors," *The Journal of the Acoustical Society of America*, vol. 154, no. 2, pp. 1226–1238, 2023.

[15] A. Levi and H. F. Silverman, "An alternate approach to adaptive beamforming using srp-phat," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 2726–2729.

[16] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, 2012.

[17] M. Jälmby, F. Elvander, and T. Van Waterschoot, "Low-rank room impulse response estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 957–969, 2023.

[18] A. Ratnarajah, I. Ananthabhotla, V. K. Ithapu, P. Hoffmann, D. Manocha, and P. Calamia, "Towards improved room impulse response estimation for speech recognition," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[19] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on speech and audio processing*, vol. 13, no. 1, pp. 135–143, 2004.

[20] Y. Zhang, Y. Wang, and Z. Duan, "Hrtf field: Unifying measured hrtf magnitude representation with neural fields," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[21] V. Bruschi, L. Grossi, N. A. Dourou, A. Quattrini, A. Vancheri, T. Leidi, and S. Cecchi, "A review on head-related transfer function generation for spatial audio," *Applied Sciences*, vol. 14, no. 23, p. 11242, 2024.

[22] C. Pörschmann, J. M. Arend, D. Bau, and T. Lübeck, "Comparison of spherical harmonics and nearest-neighbor based interpolation of head-related transfer functions," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2020.

[23] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in *Audio Engineering Society Conference: 16th International Conference: Spatial Sound Reproduction*. Audio Engineering Society, 1999.

[24] H. Gamper, "Head-related transfer function interpolation in azimuth, elevation, and distance," *The Journal of the Acoustical Society of America*, vol. 134, no. 6, pp. EL547–EL553, 2013.

[25] 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," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[26] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, "Measurement of a head-related transfer function database with high spatial resolution," in *7th forum acusticum (EAA)*, 2014.

[27] S. Li and J. Peissig, "Measurement of Head-Related Transfer Functions: A Review," *Applied Sciences*, vol. 10, no. 14, p. 5014, Jan. 2020.

[28] D. R. Begault and L. J. Trejo, *3-D sound for virtual reality and multimedia*. San Diego, CA, USA: Academic Press Professional, Inc., 2000.

[29] F. P. Freeland, L. W. Biscainho, and P. S. Diniz, "Efficient hrtf interpolation in 3d moving sound," in *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.

[30] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, and A. Reyes-Lecuona, "3d tune-in toolkit: An open-source library for real-time binaural spatialisation," *PloS one*, vol. 14, no. 3, p. e0211899, 2019.

[31] A. Acosta, F. Grijalva, R. Álvarez, and B. Acuña, "Bilinear and triangular spherical head-related transfer functions interpolation on non-uniform meshes," in *2020 IEEE ANDESCON*. IEEE, 2020, pp. 1–6.

[32] Z. Ben-Hur, D. Alon, P. W. Robinson, and R. Mehra, "Localization of virtual sounds in dynamic listening using sparse hrtfs," in *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2020.

[33] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[34] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.

[35] J. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 439–452, 1995.

[36] V. Larcher, O. Warusfel, J.-M. Jot, and J. Guyard, "Study and comparison of efficient methods for 3-D audio spatialization based on linear decomposition of hrtf data," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

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

[38] B.-S. Xie, "Recovery of individual head-related transfer functions from a small set of measurements," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 282–294, 2012.

[39] M. Zhang, Z. Ge, T. Liu, X. Wu, and T. Qu, "Modeling of individual hrtfs based on spatial principal component analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 785–797, 2020.

[40] J. C. Torres and M. R. Petraglia, "Hrtf interpolation in the wavelet transform domain," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2009, pp. 293–296.

[41] Z. Jiang, J. Sang, C. Zheng, A. Li, and X. Li, "Modeling individual head-related transfer functions from sparse measurements using a convolutional neural network," *The Journal of the Acoustical Society of America*, vol. 153, no. 1, pp. 248–259, 2023.

[42] I. D. Gebru, D. Marković, A. Richard, S. Krenn, G. A. Butler, F. De la Torre, and Y. Sheikh, "Implicit hrtf modeling using temporal convolutional networks," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3385–3389.

[43] A. Hogg, H. Liu, M. Jenkins, and L. Picinali, "Exploring the impact of transfer learning on gan-based hrtf upsampling," in *Proc. EAA Forum Acusticum, Eur. Congress on Acoust*, 2023.

[44] T.-Y. Chen, T.-H. Kuo, and T.-S. Chi, "Autoencoding hrtfs for dnn based hrtf personalization using anthropometric features," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 271–275.

[45] J. W. Lee, S. Lee, and K. Lee, "Global hrtf interpolation via learned affine transformation of hyper-conditioned features," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

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

[47] B. Zhi, D. N. Zotkin, and R. Duraiswami, "Towards fast and convenient end-to-end hrtf personalization," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 441–445.

[48] 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, p. 904398, 2022.

[49] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in neural information processing systems*, vol. 33, pp. 7537–7547, 2020.

[50] Y. Xie, T. Takikawa, S. Saito, O. Litany, S. Yan, N. Khan, F. Tombari, J. Tompkin, V. Sitzmann, and S. Sridhar, "Neural fields in visual computing and beyond," in *Computer Graphics Forum*, vol. 41, no. 2. Wiley Online Library, 2022, pp. 641–676.

[51] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of the audio engineering society*, vol. 45, no. 6, pp. 456–466, 1997.

[52] A. Franck, W. Wang, and F. M. Fazi, "Sparse ℓ_1-optimal multiloudspeaker panning and its relation to vector base amplitude panning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 5, pp. 996–1010, 2017.

[53] J. Chen, B. D. Van Veen, and K. E. Hecox, "Synthesis of 3d virtual auditory space via a spatial feature extraction and regularization model," in *Proceedings of IEEE Virtual Reality Annual International Symposium*. IEEE, 1993, pp. 188–193.

[54] T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolating head related transfer functions in the median plane," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452)*. IEEE, 1999, pp. 167–170.

[55] S. Carlile, C. Jin, and J. Leung, "Performance measures of the spatial fidelity of virtual auditory space: Effects of filter compression and spatial sampling," in *Proceedings of the International Conference on Auditory Display (ICAD 2002)*, 2002.

[56] R. L. Jenison and K. Fissell, “A spherical basis function neural network for modeling auditory space,” *Neural computation*, vol. 8, no. 1, pp. 115–128, 1996.

[57] Y. Luo, D. N. Zotkin, H. Daume, and R. Duraiswami, “Kernel regression for head-related transfer function interpolation and spectral extrema extraction,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 256–260.

[58] X. Chen, F. Ma, Y. Zhang, A. Bastine, and P. N. Samarasinghe, “Head-related transfer function interpolation with a spherical cnn,” *arXiv preprint arXiv:2309.08290*, 2023.

[59] E. Thuillier, C. Jin, and V. Välimäki, “Hrtf interpolation using a spherical neural process meta-learner,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[60] P. Runkle, M. Blommer, and G. Wakefield, “A comparison of head related transfer function interpolation methods,” in *Proceedings of 1995 workshop on applications of signal processing to audio and accoustics*. IEEE, 1995, pp. 88–91.

[61] K. Watanabe, S. Takane, and Y. Suzuki, “Interpolation of head-related transfer functions based on the common-acoustical-pole and residue model,” *Acoustical science and technology*, vol. 24, no. 5, pp. 335–337, 2003.

[62] G. Ramos and M. Cobos, “Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications,” *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 1735–1738, 2013.

[63] B. Al-Sheikh, M. A. Matin, and D. J. Tollin, “Head related transfer function interpolation based on finite impulse response models,” in *2019 Seventh International Conference on Digital Information Processing and Communications (ICDIPC)*. IEEE, 2019, pp. 8–11.

[64] M. gebru2021implicitroz and G. H. De Sousa, “Structured iir models for hrtf interpolation.” in *ICMC*, 2010.

[65] P. Nowak and U. Zölzer, “Spatial interpolation of hrtfs approximated by parametric iir filters,” in *Proc. DAGA*, 2022.

[66] J. W. Lee and K. Lee, “Neural fourier shift for binaural speech rendering,” in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[67] D. Di Carlo, A. A. Nugraha, M. Fontaine, Y. Bando, and K. Yoshii, “Neural steerer: Novel steering vector synthesis with a causal neural field over frequency and direction,” in *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, 2024, pp. 740–744.

[68] Y. Masuyama, G. Wichern, F. G. Germain, Z. Pan, S. Khurana, C. Hori, and J. Le Roux, “Niirf: Neural iir filter field for hrtf upsampling and personalization,” in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 1016–1020.

[69] F. Ma, T. D. Abhayapala, P. N. Samarasinghe, and X. Chen, “Physics informed neural network for head-related transfer function upsampling,” *arXiv preprint arXiv:2307.14650*, 2023.

[70] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, “Sound field analysis along a circle and its application to hrtf interpolation,” *Journal of the Audio Engineering Society*, vol. 56, no. 3, pp. 156–175, 2008.

[71] M. J. Evans, J. A. Angus, and A. I. Tew, “Analyzing head-related transfer function measurements using surface spherical harmonics,” *The Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2400–2411, 1998.

[72] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, “Interpolation and range extrapolation of hrtfs [head related transfer functions],” in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. IEEE, 2004, pp. iv–iv.

[73] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, “Regularized hrtf fitting using spherical harmonics,” in *2009 IEEE workshop on applications of signal processing to audio and acoustics*. IEEE, 2009, pp. 257–260.

[74] J. Ahrens, M. R. Thomas, and I. Tashev, “Hrtf magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data,” in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, 2012, pp. 1–5.

[75] C. Pörschmann, J. M. Arend, and F. Brinkmann, “Directional equalization of sparse head-related transfer function sets for spatial upsampling,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 6, pp. 1060–1071, 2019.

[76] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, “Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint,” *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[77] 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 Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2249–2262, 2019.

[78] J. M. Arend, F. Brinkmann, and C. Pörschmann, “Assessing spherical harmonics interpolation of time-aligned head-related transfer functions,” *Journal of the Audio Engineering Society*, vol. 69, no. 1/2, pp. 104–117, 2021.

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

[80] G. D. Romigh, R. M. Stern, D. S. Brungart, and B. D. Simpson, “A bayesian framework for the estimation of head-related transfer functions,” *Journal of the Acoustical Society of America*, vol. 137, no. 4_Supplement, pp. 2323–2323, 2015.

[81] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier, “Spherical splines for scalp potential and current density mapping,” *Electroencephalography and clinical neurophysiology*, vol. 72, no. 2, pp. 184–187, 1989.

[82] Y. Luo, D. N. Zotkin, and R. Duraiswami, “Gaussian process data fusion for heterogeneous hrtf datasets,” in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2013, pp. 1–4.

[83] J. Engel, L. Hantrakul, C. Gu, and A. Roberts, “Ddsp: Differentiable digital signal processing,” *arXiv preprint arXiv:2001.04643*, 2020.

[84] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend, “Binaural reproduction of plane waves with reduced modal order,” *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 972–983, 2014.

[85] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, “Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, pp. 1–14, 2019.

[86] J. M. Arend and C. Pörschmann, *Spatial upsampling of sparse head-related transfer function sets by directional equalization-influence of the spherical sampling scheme*. Universitätsbibliothek der RWTH Aachen Aachen, Germany, 2019.

[87] F. Brinkmann and S. Weinzierl, “Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition,” in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

[88] J. G. Ribeiro, S. Koyama, R. Horiuchi, and H. Saruwatari, “Sound field estimation based on physics-constrained kernel interpolation adapted to environment,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[89] S. Koyama, J. G. Ribeiro, T. Nakamura, N. Ueno, and M. Pezzoli, “Physics-informed machine learning for sound field estimation,” *arXiv preprint arXiv:2408.14731*, 2024.

[90] F. Lluis, P. Martinez-Nuevo, M. Bo Møller, and S. Ewan Shepstone, “Sound field reconstruction in rooms: Inpainting meets super-resolution,” *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. 649–659, 2020.

[91] M. S. Kristoffersen, M. B. Møller, P. Martínez-Nuevo, and J. Østergaard, “Deep sound field reconstruction in real rooms: introducing the isobel sound field dataset,” *arXiv preprint arXiv:2102.06455*, 2021.

[92] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, “Generative models for sound field reconstruction,” *The Journal of the Acoustical Society of America*, vol. 153, no. 2, pp. 1179–1190, 2023.

[93] M. Pezzoli, D. Perini, A. Bernardini, F. Borra, F. Antonacci, and A. Sarti, “Deep prior approach for room impulse response reconstruction,” *Sensors*, vol. 22, no. 7, p. 2710, 2022.

[94] F. Miotello, L. Comanducci, M. Pezzoli, A. Bernardini, F. Antonacci, and A. Sarti, “Reconstruction of sound field through diffusion models,” in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 1476–1480.

[95] M. Olivieri, X. Karakonstantis, M. Pezzoli, F. Antonacci, A. Sarti, and E. Fernandez-Grande, “Physics-informed neural network for volumetric sound field reconstruction of speech signals,” *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, p. 42, 2024.

[96] T. Lübeck, J. M. Arend, and C. Pörschmann, “Spatial upsampling of sparse spherical microphone array signals,” *IEEE/ACM Transactions*

on Audio, Speech, and Language Processing, vol. 31, pp. 1163–1174, 2023.

[97] P. Samarasinghe, T. Abhayapala, M. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2217–2227, 2015.

[98] M. Pezzoli, M. Cobos, F. Antonacci, and A. Sarti, "Sparsity-based sound field separation in the spherical harmonics domain," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 1051–1055.

[99] S. Koyama and L. Daudet, "Sparse representation of a spatial sound field in a reverberant environment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 172–184, 2019.

[100] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of helmholtz equation for sound field interpolation," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018, pp. 1–440.

[101] D. Caviedes-Nozal, N. A. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, "Gaussian processes for sound field reconstruction," *The Journal of the Acoustical Society of America*, vol. 149, no. 2, pp. 1107–1119, 2021.

[102] Y. Haneda, S. Makino, Y. Kaneda, and N. Koizumi, "Arma modeling of a room transfer function at low frequencies," *Journal of the Acoustical Society of Japan (E)*, vol. 15, no. 5, pp. 353–355, 1994.

[103] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205–216, 2013.

[104] N. Bertin, L. Daudet, V. Emiya, and R. Gribonval, "Compressive sensing in acoustic imaging," in *Compressed Sensing and its Applications: MATHEON Workshop 2013*. Springer, 2015, pp. 169–192.

[105] N. Murata, S. Koyama, H. Kameoka, N. Takamune, and H. Saruwatari, "Sparse sound field decomposition with multichannel extension of complex nmf," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 345–349.

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

[107] E. Zea, "Compressed sensing of impulse responses in rooms of unknown properties and contents," *Journal of Sound and Vibration*, vol. 459, p. 114871, 2019.

[108] E. F. Grande, "Sound field reconstruction in a room from spatially distributed measurements," in *23rd International Congress on Acoustics*. German Acoustical Society (DEGA), 2019, pp. 4961–68.

[109] O. Das, P. Calamia, and S. V. A. Gari, "Room impulse response interpolation from a sparse set of measurements using a modal architecture," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 960–964.

[110] M. Hahmann and E. Fernandez-Grande, "A convolutional plane wave model for sound field reconstruction," *The Journal of the Acoustical Society of America*, vol. 152, no. 5, pp. 3059–3068, 2022.

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

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

[113] X. Chen, F. Ma, A. Bastine, P. Samarasinghe, and H. Sun, "Sound field estimation around a rigid sphere with physics-informed neural network," in *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2023, pp. 1984–1989.

[114] F. Ma, S. Zhao, and I. S. Burnett, "Sound field reconstruction using a compact acoustics-informed neural network," *The Journal of the Acoustical Society of America*, vol. 156, no. 3, pp. 2009–2021, 2024.

[115] F. Miotello, F. Terminiello, M. Pezzoli, A. Bernardini, F. Antonacci, and A. Sarti, "A physics-informed neural network-based approach for the spatial upsampling of spherical microphone arrays," in *2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2024, pp. 215–219.

[116] X. Karakonstantis, D. Caviedes-Nozal, A. Richard, and E. Fernandez-Grande, "Room impulse response reconstruction with physics-informed deep learning," *The Journal of the Acoustical Society of America*, vol. 155, no. 2, pp. 1048–1059, 2024.

[117] M. Middleton, D. T. Murphy, and L. Savioja, "The application of fourier neural operator networks for solving the 2d linear acoustic wave equation," in *Proceedings of Forum Acusticum, European Acoustics Association, Turin, Italy*, 2023, pp. 1–8.

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

[119] H. Bi and T. D. Abhayapala, "Point neuron learning: a new physics-informed neural network architecture," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2024, no. 1, p. 56, 2024.

[120] D. Caviedes-Nozal and E. Fernandez-Grande, "Spatio-temporal bayesian regression for room impulse response reconstruction with spherical waves," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023.

[121] J. M. Schmid, E. Fernandez-Grande, M. Hahmann, C. Gurbuz, M. Eser, and S. Marburg, "Spatial reconstruction of the sound field in a room in the modal frequency range using bayesian inference," *The Journal of the Acoustical Society of America*, vol. 150, no. 6, pp. 4385–4394, 2021.

[122] X. Feng, J. Cheng, S. Chen, and Y. Shen, "Room impulse response reconstruction using pattern-coupled sparse bayesian learning with spherical waves," *IEEE Signal Processing Letters*, 2024.

[123] S. Lee, H.-S. Choi, and K. Lee, "Yet another generative model for room impulse response estimation," in *2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2023, pp. 1–5.

[124] D. Sundström, F. Elvander, and A. Jakobsson, "Estimation of impulse responses for a moving source using optimal transport regularization," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 921–925.

[125] M. Pezzoli, F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Reconstruction of the virtual microphone signal based on the distributed ray space transform," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 1537–1541.

[126] C. Pörschmann and J. M. Arend, "A method for spatial upsampling of directivity patterns of human speakers by directional equalization," *Proceedings of the 45th DAGA*, pp. 1458–1461, 2019.

[127] M. Lemke, A. Hölter, and S. Weinzierl, "Physics-informed identification and interpolation of the directivity of sound sources," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[128] T. Quatieri and A. Oppenheim, "Iterative techniques for minimum phase signal reconstruction from phase or magnitude," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1187–1193, 1981.

[129] R. Nishimura, H. Kato, and N. Inoue, "Interpolation of head-related transfer functions by spatial linear prediction," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 1901–1904.

[130] A. Srivastava, G. Routray, and R. M. Hegde, "Spatial hrtf interpolation using spectral phase constraints," in *2020 International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2020, pp. 1–5.

[131] G. Kestler, S. Yadegari, and D. Nahamoo, "Head Related Impulse Response Interpolation and Extrapolation Using Deep Belief Networks," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 266–270.

[132] R. Duraiswami and V. C. Raykar, "The manifolds of spatial hearing," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 3. IEEE, 2005, pp. iii–285.

[133] A. Deleforge, F. Forbes, and R. Horaud, "Acoustic space learning for sound-source separation and localization on binaural manifolds," *International journal of neural systems*, vol. 25, no. 01, p. 1440003, 2015.

[134] F. Grijalva, L. C. Martini, D. Florencio, and S. Goldenstein, "Interpolation of head-related transfer functions using manifold learning," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 221–225, 2017.

[135] P. Guillon, R. Nicol, and L. Simon, "Head-related transfer functions reconstruction from sparse measurements considering a priori knowledge from database analysis: A pattern recognition approach," in *Audio Engineering Society Convention 125*. Audio Engineering Society, 2008.

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

[137] 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, 2021.

[138] W. Zhang, M. Zhang, R. A. Kennedy, and T. D. Abhayapala, "On High-Resolution Head-Related Transfer Function Measurements: An Efficient Sampling Scheme," vol. 20, no. 2, pp. 575–584. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5957268?casa_token=nItYlRra8xIAAAAA:5sf2TEl4VjjbGyV6LKNuNkUTYM3RVN-OLWolHxdlMCRecFZVk8T_4ZuRIX5Ss57sxSdGvt_y

[139] S. Damiano, F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, "Soundfield reconstruction in reverberant rooms based on compressive sensing and image-source models of early reflections," in *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2021, pp. 366–370.

[140] G. Del Galdo, O. Thiergart, T. Weller, and E. A. Habets, "Generating virtual microphone signals using geometrical information gathered by distributed arrays," in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*. IEEE, 2011, pp. 185–190.

[141] R. K. Cook, R. Waterhouse, R. Berendt, S. Edelman, and M. Thompson Jr, "Measurement of correlation coefficients in reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 27, no. 6, pp. 1072–1077, 1955.

[142] N. A. Gumerov and R. Duraiswami, *Fast multipole methods for the Helmholtz equation in three dimensions*. Elsevier, 2005.

[143] J. Nam, J. S. Abel, and J. O. Smith III, "A method for estimating interaural time difference for binaural synthesis," in *Audio Engineering Society Convention 125*. Audio Engineering Society, 2008.

Biography Section

If you have an EPS/PDF photo (graphicx package needed), extra braces are needed around the contents of the optional argument to biography to prevent the LaTeX parser from getting confused when it sees the complicated `\includegraphics` command within an optional argument. (You can create your own custom macro containing the `\includegraphics` command to make things simpler here.)

If you include a photo:

Michael Shell Use `\begin{IEEEbiography}` and then for the 1st argument use `\includegraphics` to declare and link the author photo. Use the author name as the 3rd argument followed by the biography text.

If you will not include a photo:

John Doe Use `\begin{IEEEbiographynophoto}` and the author name as the argument followed by the biography text.