POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Sound Source Directivity Interpolation With Physics-informed Neural Networks

Tesi di Laurea Magistrale in
Music and Acoustic Engineering

Author: **Edoardo Morena**

Student ID: 996003
Advisor: Mirco Pezzoli
Co-advisors: Raffaele Malvermi, Marco Olivieri
Academic Year: 2022-23

Abstract

In the context of augmented/virtual reality, reproducing the directivity of sound sources is an important procedure to ensure a high-quality immersive experience. However, accurately measuring the directivity of a source with a large number of microphones proves to be challenging which is why these measurements are often sparsely sampled. Furthermore, applications in the VR/AR domain require access to directivity values in any arbitrary direction, which are not practically measurable in real-world scenarios. In these scenarios, the need of spatial interpolation schemes becomes evident, in order to reconstruct the sound field with optimal fidelity. However, current mathematical interpolation methods exhibit limitations, as they are designed for generic interpolation problems and lack an understanding of the underlying physics. In our work, we aim to exploit an approach based on deep learning. The demonstrated capabilities of deep learning methods in identifying complex patterns and models from processed data have made it a widely employed approach in various fields, including the acoustic domain. Currently, Physics-informed Neural Networks (PINN) are receiving particular attention from the scientific research within the acoustic engineering area, owing to their ability to capture and represent fundamental physical phenomena and thus better predict the evolution of dynamical systems. In this study, we propose a spatial interpolation scheme based on PINN for the reconstruction of sound source directivity. PINN are used to solve the Helmholtz equation starting from acquisitions of the directivity over a limited set of sparsely distributed points. Results are compared with the outcomes of well-established interpolation techniques: Spherical Harmonics Decomposition (SH), Thin Plate Pseudo-Spline Interpolation (SPLINE) and Piece-wise Linear Spherical Triangular Interpolation (TRI). The results highlight the promising capabilities of the PINN method in reconstructing the acoustic field

of the sound sources taken into consideration, even when dealing with sparsely sampled grids. Notably, the PINN method demonstrated performance superiority in several instances, surpassing the effectiveness of the considered mathematical methods, which lack integrated physical information.

Keywords: Physics-informed Neural Networks, Directivity, Interpolation

Abstract in lingua italiana

Nel campo della realtà aumentata/virtuale, la riproduzione della direttività delle sorgenti sonore è fondamentale per un'esperienza immersiva di qualità. Tuttavia, misurare accuratamente la direttività di una sorgente con un gran numero di microfoni non sempre risulta fattibile, motivo per cui queste misurazioni vengono spesso campionate con una distribuzione sparsa. Inoltre, le applicazioni nel dominio della VR/AR necessitano di valori di direttività in tutte le direzione arbitrariamente, cosa impossibile da misurare nel pratico. In questi casi, diventa evidente la necessità di schemi di interpolazione spaziale per ricostruire fedelmente il campo sonoro.

Tuttavia, gli attuali metodi matematici di interpolazione presentano dei limiti, poiché sono ideati per problemi di interpolazione generici e non sfruttano in nessun modo le leggi fisiche che governano i sistemi presi in considerazione. Nel nostro lavoro, miriamo a sfruttare un approccio basato sul deep learning. Le capacità, dimostrate dai metodi di deep learning, nell'estrapoalre modelli complessi dai dati elaborati li hanno resi un approccio ampiamente impiegato in vari campi, incluso quello acustico. Attualmente, le Physics-informed Neural Networks (PINN) stanno ricevendo particolare attenzione da parte della ricerca scientifica nell'ambito dell'ingegneria acustica, grazie alla loro capacità di modellare e rappresentare i fenomeni fisici fondamentali e quindi prevedere meglio l'evoluzione dei sistemi dinamici.

In questo studio, proponiamo uno schema di interpolazione spaziale basato sulle PINN per la ricostruzione della direttività delle sorgenti sonore. Le PINN vengono utilizzate per risolvere l'equazione di Helmholtz partendo dalle acquisizioni della direttività su un set limitato di punti distribuiti in modo sparso. I risultati vengono confrontati con quelli di tecniche di interpolazione ben consolidate: Spherical Har-

monics Decomposition (SH), Thin Plate Pseudo-Spline Interpolation (SPLINE) e Piece-wise Linear Spherical Triangular Interpolation (TRI). I risultati evidenziano le promettenti capacità delle PINN nella ricostruzione del campo acustico delle sorgenti sonore prese in esame, anche quando si lavora con griglie campionate in modo sparso. In particolare, il metodo basato sulle PINN ha dimostrato una superiorità nelle prestazioni in diverse istanze, superando l'efficacia dei metodi matematici considerati, che mancano di informazioni fisiche integrate.

Parole chiave: Physics-informed Neural Networks, Direttività, Interpolazione

Contents

Introduction

The directivity of a sound source refers to the way in which sound is radiated along different directions. As the scientific research towards the Virtual/Augmented Reality realm makes progress, it has emerged how an accurate representation of sound source directivity plays a pivotal role in enhancing the immersive experience for users [10, 17, 26, 27, 33, 50]. Room-acoustics simulations usually provide source-to-receiver room impulse responses (RIRs), thus having a faithful and continuous representation of radiation patterns in room acoustical simulations and auralization is mandatory to obtain a spatial excitation of the room closer to what occurs in reality [50].

As individuals engage with auralized virtual spaces, the realistic rendering of sound source directionality becomes fundamental in shaping a convincing and immersive auditory landscape. It has been demonstrated how the angular resolution of the directivity affects the simulated room impulse response (RIR) and several other room acoustic parameters, up to a spherical harmonics (SH) order of $N = 10$ [26]. Otondo *et al.* [33] successfully showed a clear influence of the changes in the representation of directivities on the distribution of acoustical parameters in a room auralization context. Specifically, the listeners perceived loudness as the parameter mostly influenced by directivity, followed by reverberance and clarity. Moreover, modifying the directivity of an audio source has the potential to influence many human perceptual aspects, such has localization [10]. As an example, Politis *et al.* [27] demonstrated how to effectively control the perceived distance of an auditory object, by changing its directivity pattern and consequently the direct-to-reverberant signal ratio. The study conducted a comparative analysis between two different distance rendering methods: one solely controlling the sound power emitted by the source through level gain, and another also controlling the

directivity pattern of the sound source. The results demonstrated how impactful the manipulation of the directivity pattern is when altering the distance perception of a sound source. Directivity also plays a big role when considering speech sound sources, as shown by Ehret *et al.* [17]. Their work evaluated the effect that the directivity of a speech sound source has on the social presence of a virtual agent.

Directivity measurements are frequently conducted using microphone arrays [9] and for the sake of obtaining optimal results, a large-aperture array with closely spaced microphones is preferred. Nevertheless, practical constraints such as cost, microphone arrangement or the high number of required channels may hinder the realization of such an ideal configuration. Therefore, different strategies for the interpolation of the directivities have been adopted to compute the sound in arbitrary directions [2, 4]. In [2], different state-of-the-art strategies for the interpolation of directivities have been analyzed, including Spherical Harmonics (SH) decomposition, thin-plate pseudo-spline interpolation (SPLINE) and the piece-wise linear spherical triangular interpolation (TRI). The interpolation results on different musical instruments have been compared, highlighting the superior accuracy of SPLINE and TRI, with respect to SH, for the interpolation of sparse directivity data. In this thesis, we aim at presenting a learning-based directivity reconstruction method and comparing it with the aforementioned conventional interpolation methods. Moreover, unlike other learning-based methods that are purely data-driven, our method is based on a Physics-informed Neural Network (PINN). PINNs enable the regularization of the estimations to conform to specified partial differential equations (PDEs) that govern the analyzed system, thereby providing solutions informed by the underlying physics [44]. This feature allowed us to model the directivity by training the network to adhere not only to the training measurements data, but also to be a solution of the underlying wave equation, namely the Helmholtz equation, offering an efficient approach for solving the directivity upsampling task with limited data. Taking the system domain (e.g., the spatial coordinates) as input, the PINN outputs an estimate of the directivity in any arbitrary direction. As a result, the proposed PINN model presents what is referred to as an *implicit neural representation* of the directivity, which could potentially

be reconstructed in a continuous fashion. The evaluation of the proposed method and the baselines has been conducted on the directivity measurements of different sound sources [3, 19]. Two specific evaluation metrics served the purpose of assessing the performance of the analyzed methods: average Normalized Mean Square Error ($\overline{\text{NMSE}}$) and frequency-dependent Normalized Cross Correlation (NCC). In this work, leveraging on the specified evaluation metrics and utilizing 3D directivity patterns for visual feedback, we denoted the superiority of the PINN method in the directivity interpolation task.

1 | Theoretical background

In this section, we examine the basis concepts essential for grasping the two main topics addressed in this thesis: sound source directivity and Physics-informed Neural Networks (PINNs).

As far as the directivity of sound sources is concerned, our focus is on the far-field sound radiation. Therefore, we aim to outline the mathematical model that characterizes this physical phenomenon. Moreover, an overview will be presented on the employed free-field measurement techniques, explaining how sound source directivity is usually quantified.

We then provide the theoretical foundations of Physics-informed Neural Networks (PINNs). Specifically, this section aims to explain how PINNs integrate physical principles into their architecture, enabling them to learn and represent complex physical models. By addressing these theoretical foundations, we intend to ease a comprehensive understanding of the subsequent discussions on sound source directivity interpolation using PINNs.

1.1. Sound source directivity

The directivity is a frequency-dependent complex function representing the distribution of sound when emitted by an acoustic source, i.e., indicating how the sound pressure level varies with respect to the angle from the source reference axis. Let us assume a 3D reference system with Cartesian coordinates $[x, y, z]^T$. The system has its origin in the acoustic centre of the source. Figure 1.1 shows an example of directivity pattern in the adopted coordinate system. The dependency of the directivity with respect to the direction of propagation can be expressed in terms of the spherical coordinates $[r, \theta, \phi]^T$, where r is the radial distance, θ is the incli-

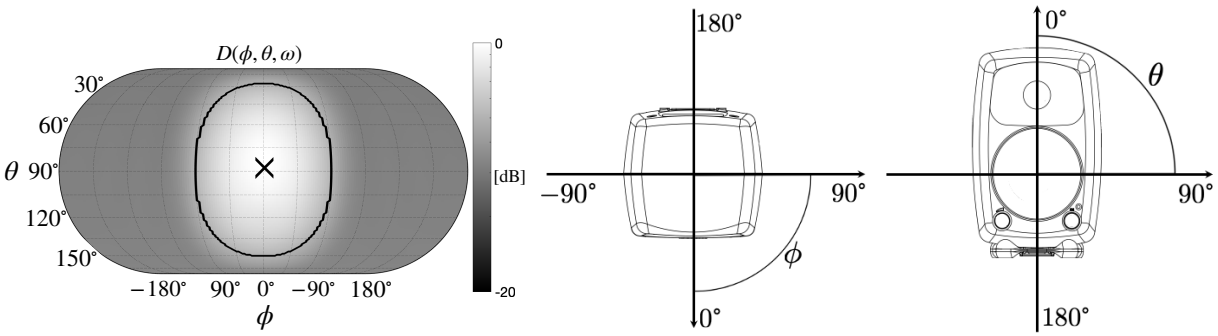


Figure 1.1: Example of directivity pattern $D(\phi, \theta, \omega)$ of a Genelec 8030A at 1.4 kHz, taken from [56]. A solid black line delimits the region with maximum acoustic energy emission, while a black cross identifies the direction with maximum acoustic energy emission. The reference system is reported from top, i.e., xy plane, and frontal, i.e., yz plane, views.

nation coordinate and ϕ is the azimuth coordinate. The spherical coordinates can be written as a function of the Cartesian coordinates.

Accordingly to [13], a general definition of the directivity can be derived starting from the far-field radiation model. Indeed, the Fourier transform of the far-field sound field P_∞ can be written, for the free-field case, as

$$P_\infty(\omega, x, y, z) = \frac{e^{-i\frac{\omega}{c}r}}{r} S(\omega)\, D(\omega, \theta, \phi), \quad \text{with} \quad \begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \phi = \arctan\left(\frac{y}{x}\right) \\ \theta = \arccos\left(\frac{z}{r}\right) \end{cases}, \tag{1.1}$$

where c is the speed of sound in air, ω is the angular frequency, $S(\omega)$ is the source signal in the frequency domain, when evaluated at its acoustic center, and $D(\omega, \theta, \phi)$ is the source directivity. In this thesis the magnitude of the directivity $|D(\omega, \theta, \phi)|$ is designated as the *directivity pattern*, although in the literature, it may also be referred to as the *polar pattern*.

The definition provided in (1.1) can be further detailed in terms of the so-called *finite-distance signature* of the directivity, first introduced in [5]. This representation proves to be useful in the formalization of interpolation problems because it employs spherical harmonics, which enable the interpolation along angular di-

rections. Given the far-field scenario and a SH decomposition of the sound field, (1.1) can be rewritten as [5]:

$$P_\infty(\omega, x, y, z) = \frac{e^{-i\omega\frac{r}{c}}}{r}\frac{c}{i\omega}\sum_{n=0}^{\infty}\sum_{m=-n}^{n}(-i)^n \check{c}_{n,m}(\omega)\, Y_{n,m}(\theta, \phi), \tag{1.2}$$

where $Y_{n,m}(\theta, \phi)$ and $\check{c}_{n,m}(\omega)$, for $n \geq 0$ and $-n \leq m \leq n$, are the SH basis functions and the frequency-dependent coefficients of the corresponding linear combination, respectively. The SH basis functions have the following form:

$$Y_n^m(\theta, \phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)e^{im\theta}, \tag{1.3}$$

where $P_n^m(\cdot)$ are the associated Legendre functions, $(\cdot)!$ represents the factorial function, $m \in \mathbb{Z}$ specifies the function degree, and $n \in \mathbb{N}$ the order of the function. It follows that the coefficients $\check{c}_{n,m}(\omega)$ contain information about the directivity. The term P_∞ in (1.2) is referred to as the *far-field signature* of the directivity and indicates that, at a sufficient distance r, a source of finite spatial extent radiates spherical wave fronts, represented by the term $\frac{e^{-i\omega\frac{r}{c}}}{r}$. The complex amplitudes of the wave fronts depend on the angles θ, ϕ and the coefficients $\check{c}_{n,m}(\omega)$.

Directional characteristics are strongly influenced by factors such as the size, shape, and design of the sound source [31]. For example, a highly directional sound source, like a directional loudspeaker, will exhibit a narrower directivity pattern, compared to other sources. Furthermore, as demonstrates the frequency dependence of the function $D(\omega, \theta, \phi)$, the same sound source exhibits different directivity patterns according to the frequency it is emitting, as showed in Fig. 1.2.

Measurements of common sound sources (e.g., musical instruments, speech etc.) have been conducted to extract their directivity patterns. These measurements have been carried out in various ways throughout the history of acoustic research. Indeed, the directivity of musical instruments has been object of many studies [14, 18, 34, 45], with one of the earliest attempts being made by Meyer in the 1970s [32]. Nevertheless, Meyer's collected data mainly offer averages of directivity across the entire frequency range of the instruments. In his work, limited details

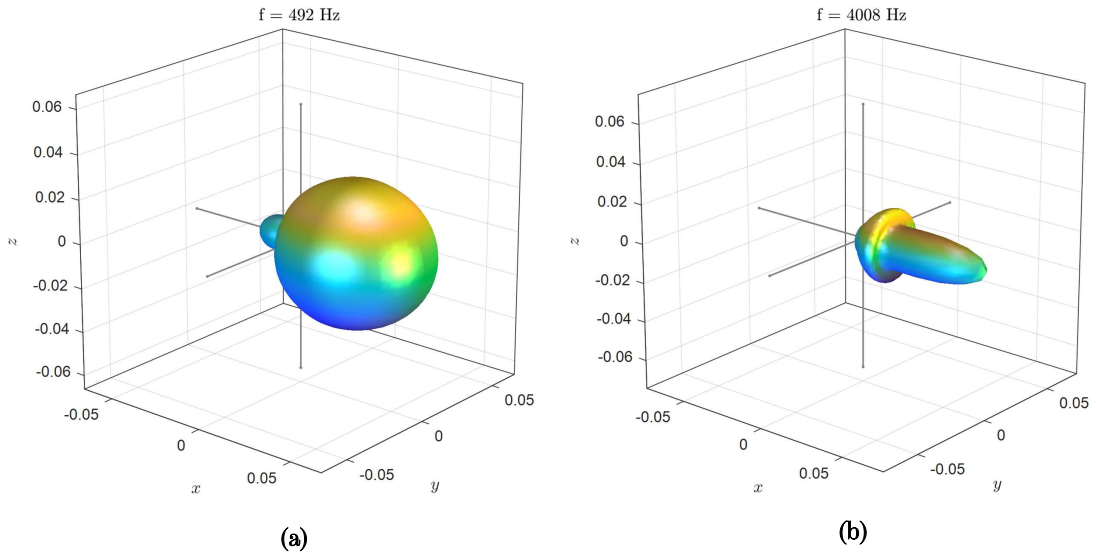


Figure 1.2: An example of two directivity polar patterns of a *Avantgarde Acoustic Solo* loudspeaker. Pattern **(a)** at $f_k = 492\,\text{Hz}$ displays a broader shape, whereas pattern **(b)** at $f_k = 4008\,\text{Hz}$ exhibits a narrower one.

are provided regarding the directivity of instruments for specific tones, even though it can undergo significant changes depending on the frequency that is emitted [32].

Many research works followed Meyer's experiments, exploiting different microphone array configurations, including a spherical microphone array of 12 microphones [16] and 13 microphones displaced in two surrounding circular arrays [33]. One of the most documented and comprehensive gathering of directivity data has been accomplished at the Technical University of Berlin for 41 modern and authentic musical instruments, i.e., of different historical periods of instrument making, by Pollow *et al.* [38]. To achieve their goal, the team constructed a lightweight spherical array with 32 microphones, which were regularly distributed on a sphere with a diameter of approximately 4.2 m surrounding the musician. In order to meet some budget restrictions, reasonably priced electret microphones with a flat frequency response were used (Sennheiser KE4-211-2 [48]). Recording in an anechoic cham-

ber is an important aspect when collecting directivity measurements from natural sound sources, as it can get rid of unwanted reflections in the measured impulse response. Therefore in [38], the measurement campaign was hold in the full-anechoic chamber located at the Department for Audio Communication of the TU Berlin.

An inherent challenge in modeling the radiation pattern of musical instruments lies in the definition and estimation of the acoustical center. Acoustic source centering is arguably considered the most mathematically demanding aspect in collecting directivity data [49]. As musical instruments exhibit spatially extended and note-dependent vibration patterns, the radiated sound center often does not coincide with the center of the surrounding microphone array. Additionally, variations in the time of arrival at different microphone positions introduce phase shifts and unwanted artefacts. If multiple contributing sources are present, interference could make the radiation patterns even more complex. For these reasons, representing the radiation pattern in such cases needs the definition of a distinct acoustic center. This type of task has been approached through various methods, resulting in the development of several noteworthy algorithms that serve this purpose [20, 49]. It is also important to notice that, because of spatial aliasing, caused by an insufficient spatial resolution of the microphone distribution [43], acoustic source centering cannot be applied to the radiation pattern at higher frequencies. The more microphones are present in the array, the higher the frequencies that can benefit from acoustic source centering. For example, in [50], with a 32 microphones spherical array, acoustic source centering has not been applied to directional patterns at frequencies higher than 1 kHz. Furthermore, rigid spherical microphone arrays have provided an innovative alternative to traditional arrays for sound field measurements [41]. Unlike typical arrays, rigid spherical configurations minimize the effects of reflections and scattering in acoustic environments. If unaddressed, these phenomena can significantly impact the sound field estimation accuracy.

1.2. Physics-informed neural networks

Physics-informed Neural Networks (PINNs) are a class of neural networks that are designed to incorporate prior knowledge about physical systems into their architecture, first introduced in the seminal work [44]. They are mostly used in the context of solving partial differential equations (PDEs) and other physical problems, as they can embed the knowledge of physical laws governing a given dataset and that can be described by PDEs. Their intrinsic relationship with PDEs comes from automatic differentiation (AD), the fundamental mathematical tool underlying the functioning of neural networks, that greatly simplifies the computing of the derivatives of any order of a NN output with respect to any of its input. Indeed, leveraging on the known neural network capability as universal function approximator [22], PINNs aim to represent a function that conforms to a physical law, having as input variable the domain of interest (e.g., space and time). The derivatives of such a function with respect to a point within the domain, crucial for PDE computation, are then efficiently approximated numerically exploiting the PINN automatic differentiation (AD). The physical law that the PINN output, after the training phase, aims to obey, is embedded in the PINN through its loss function. A PINN loss function usually consists of two parts, namely the data loss ($\mathcal{L}_{data}$) and the PDE loss ($\mathcal{L}_{PDE}$):

$$\mathcal{L}_{PINN} = \mathcal{L}_{data} + \mathcal{L}_{PDE}. \tag{1.4}$$

The data loss makes the PINN output fit the training dataset, while the PDE loss regularizes the PINN output to conform with an underlying physical law. Typical examples of data losses employed in a PINN context are the Mean Square Error (MSE) Loss, namely

$$\mathcal{L}_{data} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2, \tag{1.5}$$

or the the Mean Absolute Error (MAE) Loss

$$\mathcal{L}_{data} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right|, \tag{1.6}$$

where $\hat{y}_i$ is the predicted value, y_i is the actual training data point and N is the total number of training points. In a direct problem scenario, where PINNs are exploited to provide physical solutions at input points in the domain of interest, without intending to fit any training sample point [54], the loss function only consists in the PDE loss, i.e.,

$$\mathcal{L}_{PINN} = \mathcal{L}_{PDE}. \tag{1.7}$$

In the event that the PINN output must respect additional physical constraints, for example matching boundary conditions at specific points in the input domain, adding a boundary condition loss term ($\mathcal{L}_{bc}$) in the total loss function could be mandatory:

$$\mathcal{L}_{PINN} = \mathcal{L}_{PDE} + \mathcal{L}_{bc}. \tag{1.8}$$

PINN's connection with PDEs and physical laws has made them a fundamental tool applied in many fields of engineering that concern dynamic systems. Fluid mechanics [12], biomechanics [47], geophysical [23] and acoustics [30, 59] are amongst them. In the context of acoustics, PINNs demonstrate remarkable efficacy through the integration of the Helmholtz equation [54] into their loss function. We illustrate here the Helmholtz equation in its homogeneous form and in the frequency domain:

$$P\left(\omega, x, y, z\right) + \frac{1}{(\omega/c)^2}\nabla^2 P\left(\omega, x, y, z\right) = 0 \tag{1.9}$$

where $c = 340m/s$ is the speed of sound, and ∇^2 denotes the Laplacian operator. By embedding this fundamental acoustic wave equation within the training process, PINNs combine observed data with the governing principles of sound propagation, resulting in an efficient learning, especially in scenarios where limited data is available [54].

2 | State-of-the-art directivity interpolation techniques

In acoustics, sound field interpolation or sound field reconstruction (SFR) refers to the problem of estimating values within a sound field at locations where measurements are not available. SFR is part of a larger family of problems known as function interpolation in the spatial domain. It involves predicting sound pressure level values in between existing measured points, to create a continuous representation of the sound field. SFR can be employed to estimate directivity patterns at unmeasured locations over a given sampling grid.

This section offers an overview of the current state-of-the-art in the study of the interpolation of audio source directivity, a crucial aspect within the realm of acoustic research. In doing so, we focus on three prominent conventional methods, namely *Spherical Harmonics Decomposition* (SH), *Thin Plate Pseudo-Spline Interpolation* (SPLINE), and *Piece-wise Linear Spherical Triangular Interpolation* (TRI). We examine the SH technique given its predominant use in the field of SFR. The TRI and SPLINE techniques, generally used in spatial interpolation problems, have also been analyzed in this case for the reconstruction of directivity patterns. These methods are thoroughly examined for their efficacy in interpolating the directional characteristics of sound sources, highlighting their strengths and their limits.

2.1. Model-based interpolation techniques

In the realm of SFR, there exists a multitude of techniques that have played a pivotal role in the acoustic research over time. The use of analytical models, instead of data-driven algorithms, is common to these conventional methodologies. Many

of them are based on basis function decomposition. According to this paradigm, sound field measurements are decomposed into a set of basis functions and the weights of their linear combination, such that the best approximation of the original data is achieved. The chosen basis functions are solutions of the partial differential equation (PDE) modeling acoustic wave propagation, namely the Helmholtz equation, which can be evaluated at arbitrary positions in space. The decomposition allows for a compact representation of the sound field and facilitates the estimation of values at unobserved positions. Examples of suitable basis functions are spherical harmonics (SH) [1, 57], cylindrical harmonics (CH) [60] and plane waves [35, 40]. In the realm of spatial domain function interpolation, approaches as the discontinuous nearest-neighbor interpolation, as well as the continuous linear and natural neighbor interpolation [2], relies on simpler models that compute the interpolated values harnessing the samples close to the extimation point. Alternatively, other methods may leverage on more mathematically complex models, capable of utilizing an extensive set of sample points to achieve a higher precision in interpolations. The Thin Plate Pseudo-Spline Interpolation [58] serves as an example within this category.

Harnessing the work conducted by Ackermann *et. al* in [2], we focus on three traditional, model-based interpolation techniques: SH, SPLINE, and TRI, unravelling their theoretical foundations. We included SH interpolation due to its prevalent usage in musical acoustics. SPLINE interpolation was selected as it demonstrated to outperform other interpolation techniques when a limited number of sample points are at disposal [21, 51]. Ultimately, the spherical triangular interpolation, is mainly employed in three-dimensional vector based amplitude panning (VBAP) [39], an algorithm commonly used for robust virtual sound source positioning [62]. Additionally, the aforementioned upsampling methods are utilized as baselines for our comparative analysis, as detailed in the subsequent chapters.

2.1.1. Spherical Harmonics Decomposition

Spherical Harmonics Decomposition emerges as a powerful technique when dealing with data that can be expressed in a polar coordinate reference system. The

decomposition involves expressing a given function as a weighted sum of spherical harmonics, i.e., a set of orthogonal functions, defined on the surface of a sphere, each characterized by an angular frequency and an amplitude coefficient. The order of the spherical harmonics determines the number of lobes or nodes in their shape. For instance, zero-order spherical harmonics represent a omnidirectional spherical shape, while higher orders introduce more complexity with additional nodes and lobes on the sphere.

The Spherical Harmonics Decomposition approach has proven to be highly valuable across various research domains, including signal processing [46] and acoustics [6, 15, 37]. When a finite number of microphones is employed to sample the sound pressure of a source over the surface of a sphere, it is possible to compute spherical harmonics coefficients based on the acquired data. These coefficients, derived from the measured values, can subsequently be employed to approximate the sound pressure function across the entire measuring surface [42]. The finite number of measurements allows for reconstructing the sound field up to a certain level of detail, corresponding to a maximum order of spherical harmonics. Consequently, we obtain a reconstruction of the sound field $f(\theta, \phi)$ (where θ denotes inclination, and ϕ denotes azimuth) through a weighted summation of a finite set of orthogonal base functions:

$$f(\theta, \phi) \sim \sum_{n=0}^{N} \sum_{m=-n}^{n} \check{c}_{n,m} Y_n^m(\theta, \phi), \tag{2.1}$$

where $N \in \mathbb{N}$ indicates the spherical harmonics order and $\check{c}_{n,m}$ are the considered weights of the corresponding spherical harmonics.

The spherical coefficients $\check{c}_{n,m}$ completely describe and could perfectly represent an order-constrained sound field $f(\theta, \phi)$ on the entire sphere. However, it is important to notice that the sound field function $f(\theta, \phi)$ does not always happen to be order-constrained, which means that it exists at least one $f_{nm} \neq 0$ for a $n > N$, this is the typical case for a sound field composed of plane waves. In this scenario, errors due to spatial aliasing are unavoidable when reconstructing the sound pressure starting from its known samples. Nevertheless, these errors can be made negligible if the magnitude of the high-order coefficients is sufficiently small. This condition

holds true for all $n \gg kr$. Hence, assuming that the choice of sampling method, frequency and sphere radius satisfy the condition $kr < N$, spatial aliasing errors are sufficiently small [42].

By the means of a spherical microphone array comprising Q channels, the sound pressure field $f(\theta, \phi)$ is sampled. The resulting samples, denoted as $p_q = f(\theta_q, \phi_q)$, are acquired at the specific positions (θ_q, ϕ_q) corresponding to each microphone in the array, for $q \in \{1, 2, ..., Q\}$. When expressed in matrix form, and considering an order-constrained sound field, Eq. (2.1) can be articulated as follows:

$$\mathbf{f} = \mathbf{Y}\check{\mathbf{c}}_{nm}, \tag{2.2}$$

where the matrix $\mathbf{Y} \in \mathbb{C}^{Q\times(N+1)^2}$ is defined as follows:

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\theta_1, \phi_1) & Y_1^{-1}(\theta_1, \phi_1) & \dots & Y_N^N(\theta_1, \phi_1) \\ Y_0^0(\theta_2, \phi_2) & Y_1^{-1}(\theta_2, \phi_2) & \dots & Y_N^N(\theta_2, \phi_2) \\ \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\theta_Q, \phi_Q) & Y_1^{-1}(\theta_Q, \phi_Q) & \dots & Y_N^N(\theta_Q, \phi_Q), \end{bmatrix}, \tag{2.3}$$

and the vector $\mathbf{f}$ contains the Q sound pressure measurements. The value of Q with respect to the number of spherical harmonics, i.e., $(N+1)^2$, outlines three different scenarios. When $Q = (N+1)^2$, we fall into the rarest scenario and, in this case, if the measuring points are perfectly distributed [53] and thus $\mathbf{Y}$ is a well-conditioned full-rank matrix, Eq. (2.2) can be solved by means of inverting matrix $\mathbf{Y}$:

$$\check{\mathbf{c}}_{nm} = \mathbf{Y}^{-1}\mathbf{f}. \tag{2.4}$$

When $Q > (N+1)^2$, the system becomes an over-determined system of linear equations and can be solved through the least-square best fit, that is by taking the Moore-Penrose inverse of $\mathbf{Y}$ and thus seeking a solution $\check{\mathbf{c}}_{nm}$ that minimizes the energy of the error:

$$\min_{\check{\mathbf{c}}_{nm}} \|\mathbf{f} - \mathbf{Y}\check{\mathbf{c}}_{nm}\|^2 \Rightarrow \check{\mathbf{c}}_{nm} = \mathbf{Y}^{\dagger}\mathbf{f}. \tag{2.5}$$

If $Q < (N+1)^2$, the system of equations is under-determined and Eq. (2.2) pro-

vides an infinite number of solutions. In this case, the Moore-Penrose inverse of the matrix $\mathbf{Y}$ seeks a solution $\check{\mathbf{c}}_{nm}$ with minimum Euclidean norm, i.e., with minimal wave-spectral power $\|\check{\mathbf{c}}_{nm}\|^2$ ([61], p.79):

$$\min_{\check{\mathbf{c}}_{nm}} \|\mathbf{f} - \mathbf{Y}\check{\mathbf{c}}_{nm}\|^2 \quad s.t. \quad \mathbf{f} = \mathbf{Y}\check{\mathbf{c}}_{nm} \Rightarrow \check{\mathbf{c}}_{nm} = \mathbf{Y}^{\dagger}\mathbf{f}. \tag{2.6}$$

2.1.2. Thin Plate Pseudo-Spline Interpolation

Thin Plate Pseudo-Spline (SPLINE) Interpolation [58] is another traditional interpolation technique particularly employed in the field of spatial interpolation, which aims to estimate values at unmeasured locations over a grid, providing a smooth and continuous representation of the spatial field. The term "Thin Plate" refers to the mathematical model underlying the interpolation scheme. Indeed, this technique involves fitting a thin-plate spline, a particular mathematical surface, to the given data points [25]. SPLINE intends to find a continuous and smooth function $\check{f}(\theta, \phi)$, such that the values of $\check{f}(\theta_q, \phi_q)$ closely approximate the measured values p_q, while minimizing the bending energy of the sphere surface. The SPLINE method is highly regarded among interpolation techniques, as it provides a closed-form expression for interpolating sparsely distributed measurements while effectively addressing computational complexity.

An optimal interpolating Thin Plate Pseudo-Spline can be obtained by seeking the solution of the following optimization problem:

$$\min_{\check{f}} J_M(\check{f}) \quad s.t. \quad \check{f}(\theta_q, \phi_q) = p_q, \tag{2.7}$$

where $J_M(\check{f})$ is the bending energy function, defined by:

$$J_M\left(\check{f}\right) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \frac{(\hat{f}_{nm})^2}{\xi_{nm}}, \tag{2.8}$$

where

$$\hat{f}_{nm} = \int_S \check{f}\left(\theta, \phi\right) Y_n^m\left(\theta, \phi\right) d\theta d\phi \tag{2.9}$$

and

$$\xi_{nm} = \left[\left(n + \frac{1}{2}\right)(n+1)(n+2)\cdots(n+M+1) \right]^{-1}. \tag{2.10}$$

The bending energy function is parametrized according to the spline order $M \in \mathbb{N}$ which determines the derivability of the spline and ensures its continuity up to the $(M-1)\,th$ derivative [58], so that it is C^{M-1} smooth.

In order to find a smoothing interpolation function, Eq (2.7) can be reformulated as follows:

$$\min_{\check{f}} \frac{1}{Q} \sum_{q=1}^{Q} \left[p_q - \check{f}\left(\theta_q, \phi_q\right) \right]^2 + \lambda J_M\left(\check{f}\right), \tag{2.11}$$

where $\lambda \geq 0$ denotes the smoothing parameter.

A solution of the two problems showed in Eq. (2.7) and Eq. (2.11) has the following form:

$$\check{f}_{M,\lambda}\left(\theta, \phi\right) = \sum_{q=1}^{Q} c_q R\left(\theta, \phi; \theta_q, \phi_q\right) + d, \tag{2.12}$$

where $R\left(\theta, \phi; \theta_q, \phi_q\right)$ is the reproducing kernel for the Hilbert space $H_M^0\left(S\right)$ with norm $J_M^{\frac{1}{2}}\left(\cdot\right)$, namely

$$R\left(\theta, \phi; \theta_q, \phi_q\right) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \xi_{nm} Y_n^m\left(\theta, \phi\right) Y_n^m\left(\theta_q, \phi_q\right) = \tag{2.13}$$

$$\frac{1}{2\pi} \sum_{n=1}^{\infty} \frac{1}{(n+1)(n+2)\cdots(n+M+1)} P_n\left(z\right), \tag{2.14}$$

with P_n referring to the associated Legendre polynomials and z denoting the cosine of the spherical angle α between the two arguments of the kernel function, i.e.,

$$z = \cos\alpha = \sin\left(\phi\right)\sin\left(\phi_q\right) + \cos\left(\phi\right)\cos\left(\phi_q\right)\cos\left(\theta - \theta_q\right). \tag{2.15}$$

A closed-form expression for the reproducing kernel $R\left(\theta, \phi; \theta_q, \phi_q\right)$, is

$$R\left(\theta, \phi; \theta_q, \phi_q\right) = \frac{1}{2\pi} \left[\frac{1}{M!} q_M\left(z\right) - \frac{1}{(M+1)!} \right] \tag{2.16}$$

with

$$q_M(z) = \int_0^1 (1-h)^M \left(1 - 2hz + h^2\right)^{-\frac{1}{2}} dh. \tag{2.17}$$

In order to compute $R(\theta, \phi; \theta_q, \phi_q)$, the $q_M(z)$ term must be evaluated recursively and can be found in ([58], Tab.1) for $M = \{1, 2, 3, \cdots, 10\}$), together with the coefficient vector $\mathbf{c} = [c_1, \cdots, c_Q]^T \in \mathbb{R}^{Q\times 1}$ and $d \in \mathbb{R}$ from Eq. (2.12) in matrix form:

$$\begin{bmatrix} \mathbf{c} \\ d \end{bmatrix} = \begin{bmatrix} \mathbf{R}_Q + Q\lambda\mathbf{I} & \mathbf{T} \\ \mathbf{T}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ 0 \end{bmatrix}, \tag{2.18}$$

where $\mathbf{R}_Q \in \mathbb{C}^{Q\times Q}$ is the $Q\times Q$ matrix with the element i, j defined as

$$(\mathbf{R}_Q)_{i,j} = R(\theta_i, \phi_i; \theta_j, \phi_j), \tag{2.19}$$

$\mathbf{I}$ is the $Q\times Q$ identity matrix, the vector $\mathbf{f} = [p_1, \cdots, p_Q]^T$ contains the Q sound pressure measurements and $\mathbf{T} = [1, \cdots, 1]^T$.

In some cases, further smoothing of the function $\check{f}$ might be useful. For example, if the sampled measurements are mixed with noise, the parameter λ can be tuned to apply more regularization and suppress unwanted variations in the field.

2.1.3. Piece-wise Linear, Spherical Triangular Interpolation

The following technique belongs to the class of polynomial interpolation [2]. In practice, the method is based on a piece-wise degree-1 barycentric spherical Bernstein-Bézier polynomial interpolation [2, 7]. We express the whole set of Q microphone positions (θ_q, ϕ_q) as a $3 \times Q$ matrix containing its three-dimensional unit direction vectors, namely

$$\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_Q], \quad \mathbf{u}_q = \begin{bmatrix} \cos\phi_q \sin\theta_q \\ \sin\phi_q \sin\theta_q \\ \cos\theta_q \end{bmatrix}. \tag{2.20}$$

By the means of the Quickhull algorithm [8], we obtain a set of triangular facets that span the convex hull of the vertices stored in $\mathbf{U}$ and that are described by

the vertex index triplets $\mathbf{v}_l = [v_{1l}, v_{2l}, v_{3l}]$, where l is the index representing the l_{th} triangle. Any arbitrary unit direction vector $\mathbf{u}$ can be expressed using the non-negative spherical barycentric or area coordinates $\mathbf{g} = [g_1, g_2, g_3]^T$ associated to the vertices $\mathbf{U}_l$ of the given l_{th} triangle, namely

$$\mathbf{u} = \mathbf{U}_l \mathbf{g}, \quad \mathbf{U}_l = [\mathbf{u}_{v_{1l}}, \mathbf{u}_{v_{2l}}, \mathbf{u}_{v_{3l}}], \tag{2.21}$$

which leads to

$$\mathbf{g} = \mathbf{U}_l^{-1} \mathbf{u}, \tag{2.22}$$

where $g_i \geq 0$ and $\sum_i g_i \geq 1$. The necessary all-positive spherical barycentric coordinates can only be obtained by selecting a suitable spherical triangle l from the convex hull, which will encompass the direction $\mathbf{u}$. As the spherical barycentric coordinates $\mathbf{g}$ reproduce the direction $\mathbf{u}$, the spherical triangular interpolation algorithm exploits the corresponding planar barycentric coordinates $\tilde{g}_i = \frac{g_i}{\sum_j g_j}$ [7] to linearly interpolate the values measured at the microphones positioned at the vertices $\mathbf{v}_l$ of the triangle l, denoted as $[p_{v_{1l}}, p_{v_{2l}}, p_{v_{3l}}]$, with their weighted average, namely

$$f(\mathbf{u}) = \tilde{g}_1 p_{v_{1l}} + \tilde{g}_2 p_{v_{2l}} + \tilde{g}_3 p_{v_{3l}}. \tag{2.23}$$

At the boundaries, this interpolation is able to perfectly reconstruct the values at the vertices and linearly interpolate the value pairs along any edge of the lth triangle. Because adjacent triangles share edges and vertices, interpolation across triangles is continuous.

3 | PINN-based directivity interpolation

By embedding the fundamental acoustic wave equation within the training process, PINNs combine observed data with the governing principles of sound propagation, resulting in an efficient learning, especially in scenarios where limited data are available. An application of the PINN tool in the acoustics research field has been described by Ma *et al.* in [29] when applied to the Head-related transfer functions (HRTFs) upsampling task. HRTF denote the transfer function between a point source and a position inside of a person's ear, characterizing a person's individual filtering effect of a sound field. The PINN architecture proposed, based on a fully connected NN and leveraging on the embedding of the Helmholtz equation, improved the HRTF upsampling accuracy, successfully outperforming the SH method in both extrapolation and interpolation scenarios.

Another remarkable case study involving PINNs has been examined by Pezzoli *et al.* [36]. Their work proposed a PINN-based approach for the reconstruction of early part of RIRs, specifically using a Physics-informed SIREN (PI-SIREN) architecture. SIREN is a particular multilayer perceptron (MLP) NN architecture that features sinusoidal activation functions. By doing so, SIREN becomes particularly suitable for applications in which the NN must implicitly represent a signal that undergoes constraints involving its derivatives [52], as it is for any acoustic signal modelling problem. The research work showed competitive results with state-of-the-art techniques based on compressed sensing and deep learning.

Given the favorable results demonstrated, the application of PINNs emerges as a compelling approach for modeling the directivity of acoustic sources. This thesis presents a system that exploits the capabilities of PINNs to provide representa-

tions of free-field directivities starting from sparse sampled sound fields, effectively performing a sound field interpolation.

3.1. Problem formulation

The upsampling problem we address is based on a layout as the one shown in Fig 1.1, where a spherical and a Cartesian coordinate system are set up with respect to the acoustic center of a specific sound source, namely point O. A spherical array of microphones is centered in O, measuring the free-field sound pressure of the source. The array is able to capture the pressure values at the microphone locations $\{x_q, y_q, z_q\}_{q=1}^{Q} = \mathcal{Q}$ in Cartesian coordinates, or $\{r_q, \theta_q, \phi_q\}_{q=1}^{Q}$, in spherical ones. Pressure functions at sampling locations will be denoted as $p\left(t, x_q, y_q, z_q\right)$ or $p\left(t, r_q, \theta_q, \phi_q\right)$. We are then interested in addressing the problem in the frequency domain, so the time-domain pressure measurements are transformed according to Fourier to obtain the transfer functions, i.e., $P\left(\omega, x_q, y_q, z_q\right)$ or $P\left(\omega, r_q, \theta_q, \phi_q\right)$, where $\omega = 2\pi f$ is the angular frequency and f is the frequency. According to the data model in Sec. 1.1, $P\left(\omega, x_q, y_q, z_q\right)$ effectively represents a discrete approximation of the sound source directivity. On a single sphere the radial coordinate does not change, so the r coordinate is actually negligible. Starting from a subset of the measurements $\mathcal{D} = \{x_d, y_d, z_d\}_{d=1}^{D} \subset \mathcal{Q}$, with $D < Q$, we aim at properly reconstructing the pressure values in the missing measurement points in a data-driven fashion. Let us express the measurement point locations in matrix from, by means of the matrices $\mathbf{Q} \in \mathbb{R}^{Q\times 3}$ and $\mathbf{D} \in \mathbb{R}^{D\times 3}$, and let us collect the measured pressure values into two matrices $\mathbf{P_Q} \in \mathbb{C}^{K\times Q}$ and $\mathbf{P_D} \in \mathbb{C}^{K\times D}$, with elements

$$[\mathbf{P_Q}]_{k,q} = P\left(\omega_k, x_q, y_q, z_q\right), \quad \text{and} \quad [\mathbf{P_D}]_{k,d} = P\left(\omega_k, x_d, y_d, z_d\right), \tag{3.1}$$

where K is the number of considered frequency bins. In practice, we seek a parametric interpolation scheme $g\left(\boldsymbol{\gamma}^*, \mathbf{D}, \mathbf{P_D}, \mathbf{Q}\right)$, whose parameters $\boldsymbol{\gamma}^*$ solve the following minimization problem:

$$\boldsymbol{\gamma}^* = \underset{\gamma}{\operatorname{argmin}} \;\; E\left(g\left(\boldsymbol{\gamma}, \mathbf{D}, \mathbf{P_D}, \mathbf{Q}\right) - \mathbf{P_Q}\right), \tag{3.2}$$

where $E(\cdot)$ is the data-fidelity term, e.g., the mean square error, between the predicted pressure values and the target measurements. The function that we exploited to generate estimations is a neural network informed by the Helmholtz equation, a second-order partial differential equation employed across a moltitude of applications in physics and engineering to model wave-like phenomena. As we are not considering any external sound source but the one we are measuring in the free-field, the Helmholtz equation is used in its homogeneous form, i.e. (in Cartesian coordinates),

$$P(\omega, x, y, z) + \frac{1}{(\omega/c)^2}\nabla^2 P(\omega, x, y, z) = 0, \tag{3.3}$$

where $c \simeq 340\,\mathrm{m\,s^{-1}}$ is the speed of sound, and ∇^2 denotes the Laplacian operator. In Cartesian coordinates, the Laplacian operator is given by:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \tag{3.4}$$

and in spherical coordinates it is given by:

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2\sin^2(\theta)}\frac{\partial^2}{\partial x^2}. \tag{3.5}$$

3.2. Network architecture

In this section we are going to explore the details of the PINN architecture we propose to solve the interpolation task. We opted for a real-valued network to model frequency-domain complex values, instead of using a complex-valued network [28], which can lead to challenging training procedures [30]. The design we present models both the real and the imaginary part of the directivity, therefore exploiting a single network and leading to a single loss function. This has been made to achieve better performances by harnessing both the real and the imaginary parts of the directivity. We expressed the directivity measurements in Cartesian coordinates, simplifying the calculation of the Laplacian by the PINN and leading to better reconstruction performances. Therefore, the inputs for our network are the Carte-

sian coordinates (x, y, z) and the outputs are the real and imaginary parts of the reconstructed complex directivity transfer function. The real and imaginary parts are then used to compose the actual complex-valued estimations of the directivity, denoted as $\hat{\mathbf{P}}(x, y, z) \in \mathbb{C}^{K\times 1}$, where K is the number of frequency bins in the target directivity transfer function. Note how the $\hat{\mathbf{P}}$ vector contains one estimated complex value for each angular frequency ω in the target directivity transfer function $\mathbf{P}(x, y, z) \in \mathbb{C}^{K\times 1}$. A PINN is usually built as a multiple layer fully connected feed-forward neural network [24]. The output of one layer is

$$\mathfrak{O} = \sigma(\mathbf{x}^T\mathbf{w} + b), \tag{3.6}$$

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

$$\Phi(\mathbf{x}, \boldsymbol{\gamma}) = (\mathfrak{O}_L \circ \circ \circ \mathfrak{O}_2 \circ \mathfrak{O}_1)(x), \tag{3.7}$$

where $\boldsymbol{\gamma}$ represents the set of all trainable parameters. We adjust the parameters $\boldsymbol{\gamma}$ by minimizing a physics-informed loss function, in a typical PINN approach as explained in Sec. 1.2. The loss function is:

$$\mathfrak{L} = \underbrace{\frac{1}{D}\sum_{d=1}^{D}\left\|\mathbf{P}(x_d, y_d, z_d) - \hat{\mathbf{P}}(x_d, y_d, z_d)\right\|_2^2}_{\mathrm{L}_{data}} + \lambda\underbrace{(\frac{1}{W}\sum_{w=1}^{W}\nabla^2\hat{\mathbf{P}}(x_w, y_w, z_w) + \left(\frac{\boldsymbol{\omega}}{c}\right)^2\hat{\mathbf{P}}(x_w, y_w, z_w))}_{\mathrm{L}_{PDE}}, \tag{3.8}$$

where $\|\cdot\|_2^2$ is the squared ℓ_2 norm, ∇^2 is the Laplacian operator, $\{x_d, y_d, z_d\}_{d=1}^{D} = \mathcal{D}$ are the Cartesian coordinates of the measured directivities that the system has at its disposal during training, $\{x_w, y_w, z_w\}_{w=1}^{W} = \mathcal{W}$ is a superset of $\mathcal{D}$ containing the points where the PDE is calculated during training, so that $\mathcal{D} \subseteq \mathcal{W} \subseteq \mathcal{Q}$, λ is the regularization term and $\boldsymbol{\omega} \in \mathbb{R}^{K\times 1}$ is the vector that comprises all the angular frequencies under consideration: $\boldsymbol{\omega} = [\omega_1, \omega_2, \cdots, \omega_K]$. L_{data} and L_{PDE} denote the data loss and the PDE loss, respectively. Following the PINN loss function

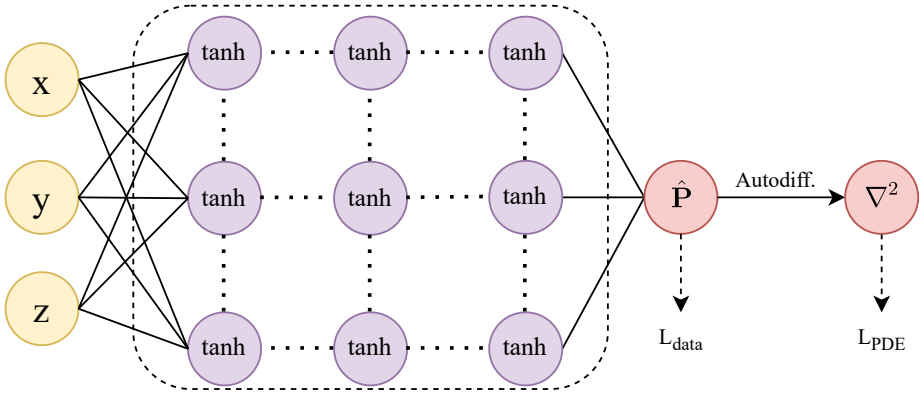


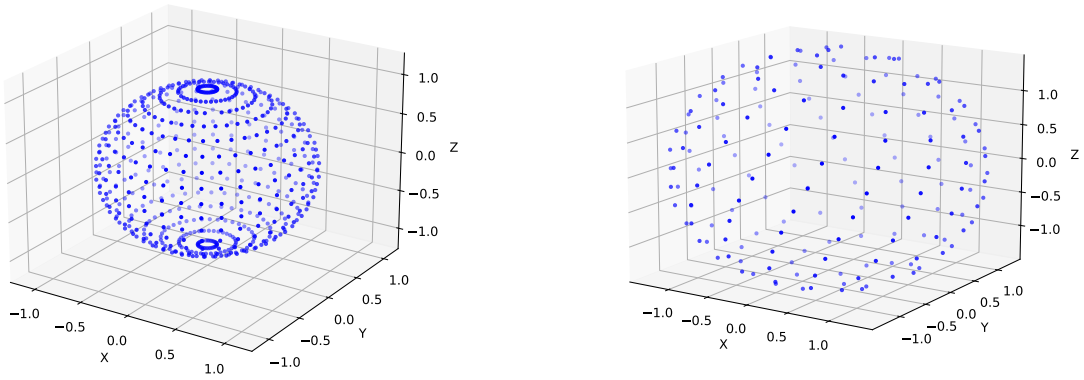
Figure 3.1: PINN structure scheme, where the input is a Cartesian coordinate triplet, tanh is the activation function and the output is a directivity transfer function approximation. The output is then used in the computing of the L_{data} and L_{PDE}, the latter by the means of the Laplacian operator

paradigm, the first term of the loss function computes the error between the network output and the directivity target values, inducing the network to adhere to the target training directivity measurements, while the second term acts as a regularization term, leading the PINN output to comply with the Helmholtz equation. Once trained, the PINN can reconstruct the directivity of the sound source for any arbitrary coordinate (x, y, z). We reached the conclusion that utilizing tanh as the activation function was the optimal choice [29, 30]. Being a smooth, second-order differentiable function capable of producing both positive and negative values, tanh demonstrated superior performance compared to other alternatives. A scheme of the network architecture is shown in Fig 3.1.

4 | Evaluation

4.1. Directivity data collection and preparation

For the training and validation phases of our study, the dataset containing directivity measurements was sourced from the GitHub repository [3] and from [19]. In addition to transforming the sound pressure measurementes into the frequency domain, we performed two additional operations on the data. The first involved the conversion of all the coordinates from spherical to Cartesian, a step taken to enhance the training process. The second manipulation entailed the resampling of the entire dataset over a new grid by means of spherical harmonics interpolation, which allows such operations as described in 2.1.1. The resampling transitioned from a spherical grid design constructed from concentric circles, to a spiral-like one, thereby achieving a more evenly distributed sampling on the sphere surface. In Fig. 4.1 the different grid structures are illustrated. The dataset, after the aforementioned manipulations, undergoes a splitting process into a training set $\mathcal{D}$ and



(a) "Concentric circles" sampling scheme.

(b) Spiral sampling scheme.

Figure 4.1: A comparison between the two sampling schemes.

a validation set $\mathcal{V}$. This division is executed ensuring that the training dataset, although containing a small percentage of the entries from the original dataset, achieves an evenly distributed sampling across the sphere's surface. A training dataset obtained with an evenly distributed sampling serves the purpose of providing the neural network with a representative set of data points, supporting its ability to generalize and predict directivities across the entire spherical space.

4.2. Performance evaluation metrics

In assessing the performance of our neural network, we employ two specific evaluation metrics. One of the employed metrics is the Normalized Mean Square Error (NMSE). NMSE is an evaluation metric often used to measure the error between predicted values from a model and actual target values, normalized with respect to the energy of the target data. We calculated the average $\overline{\text{NMSE}}$ in the following way:

$$\overline{NMSE} = \frac{1}{V}\left(\sum_{v=1}^{V} NMSE_v\right), \tag{4.1}$$

being V the cardinality of the validation dataset $\mathcal{V}$ and

$$NMSE_v = \frac{\sum_{k=1}^{K}\left|\hat{P}(\omega_k, x_v, y_v, z_v) - P\left(\omega_k, x_v, y_v, z_v\right)\right|^2}{\sum_{k=1}^{K}\left|P\left(\omega_k, x_v, y_v, z_v\right)\right|^2}. \tag{4.2}$$

Therefore, the NMSE values are computed as a function of the validation set sample locations and they are subsequently averaged with respect to the validation dataset. Finally, the $\overline{\text{NMSE}}$ is assessed in dB scale. The explained $\overline{\text{NMSE}}$ computation is derived by gradually expanding the training dataset, allowing us to evaluate the network performance as the training dataset size increases. Another employed metric is the Normalized Cross Correlation (NCC). NCC is a metric commonly used in signal processing and image analysis to quantify the similarity between two signals or images. It is a normalized version of the cross-correlation, which measures the similarity between two signals as they are shifted relative to each

other. The frequency-dependent NCC is defined as

$$NCC(\omega_k) = \frac{\sum_{q=1}^{Q}\left(\left|\hat{P}(\omega_k,x_q,y_q,z_q)\right|-\overline{\left|\hat{P}(\omega_k,x,y,z)\right|}\right)\left(|P(\omega_k,x_q,y_q,z_q)|-\overline{|P(\omega_k,x,y,z)|}\right)}{\sqrt{\sum_{q=1}^{Q}\left(\left|\hat{P}(\omega_k,x_q,y_q,z_q)\right|-\overline{\left|\hat{P}(\omega_k,x,y,z)\right|}\right)^2\sum_{q=1}^{Q}\left(|P(\omega_k,x_q,y_q,z_q)|-\overline{|P(\omega_k,x,y,z)|}\right)^2}}, \quad (4.3)$$

where ω_k is the selected angular frequency bin, the $\overline{(\cdot)}$ symbol represents the average with respect to the sample coordinates and Q is the cardinality of the whole dataset $\mathcal{Q}$. This metric serves the purpose of facilitating a comparative analysis of how well the network reconstructs the shape of the directivity magnitude pattern in comparison to the baselines. NCC produces a value between -1 and 1, where 1 indicates perfect similarity, 0 indicates no correlation, and -1 indicates perfect non-correlation. The values calculated with respect to the frequencies could be averaged, but instead we wanted to show a frequency-dependent behaviour of this metric, in order to see which frequencies are better reconstructed by the PINN and the baselines. Both the evaluation metrics are derived for increasing size of the training dataset, allowing us to evaluate the network performance as a function of the number of known samples.

4.3. Baselines

We established three baselines for a comparative analysis with our neural network. These baselines comprise traditional techniques that are both well-established and extensively utilized in the SFR context. Specifically, we consider SH, SPLINE, and TRI. Details about these three methods are unraveled in Sec. 2.1. By setting the same training dataset employed in the PINN training, these baselines aim to interpolate missing directivity values. The SH method is implemented by the means of a set of functions available in the GitHub repository containing the dataset [3]. The order of spherical harmonics used is adapted to the size of the training dataset, ensuring an appropriate reconstruction. SPLINE and TRI are implemented using the AKtools toolbox [11]. We choose to set the spline order for the SPLINE interpolation to $M = 1$. It's noteworthy that the toolbox does not let the TRI method interpolate complex values. Consequently, TRI is employed to interpolate the magnitudes of the directivity, enabling a comparison with the magnitudes pre-

dicted by our neural network. Through the evaluation, harnessing average NMSE and frequency-dependent NCC, we gain a comprehensive understanding of how our PINN performance in comparison to the three benchmark methods.

4.4. Numerical experiments

We examined four distinct audio sources, namely three loudspeakers (*Fisher STV-720*, *Avantgarde Acoustic Solo*, *KRK RP6 Rokit G2*) and one musical instrument (*Bassoon*). Besides the bassoon data, sourced from [19], the measurement data for these sources were organized in the GitHub repository [3]. Specifically, the loudspeaker data originally came from the work done by the 3-D Audio and Applied Acoustics Lab at Princeton University [55].

For every sound source, a resampling of the original measurements was conducted using a spherical harmonics-based interpolation at 366 different positions with a radial distance of 1.7 meters. For the training of our PINN, an ADAM optimizer was employed, with a learning rate set to 0.0001. The λ hyperparameter, crucial for tuning the regularization loss (L_{PDE}) with respect to the data loss (L_{data}), was set to ensure that both losses were of the same order of magnitude during training, at a value of 10^{-10}. The training process was structured to endure for 3000 epochs. Our experiments were conducted while gradually increasing the training dataset size. The training dataset was composed by 10, 30, 50, and eventually 60 samples, and has been employed to both train the PINN and to undergo interpolation by the means of the baseline methods.

A distinctive case, denoted as the "*Non-Uniform Sampling Scenario*", was introduced through the 50-sample training dataset. In this scenario, the training dataset was created with a less uniform sampling of the sphere surface. As shown in Fig. 4.2, part of the sampling points that were originally belonging to the upper spherical cap are missing. This variation allowed us to examinate how our PINN performs compared to traditional interpolation methods in a non-ideal sampling context. The optimal neural network dimensionality, namely the number of layers and neuron quantity, remains an open research area. In our study, we empiri-

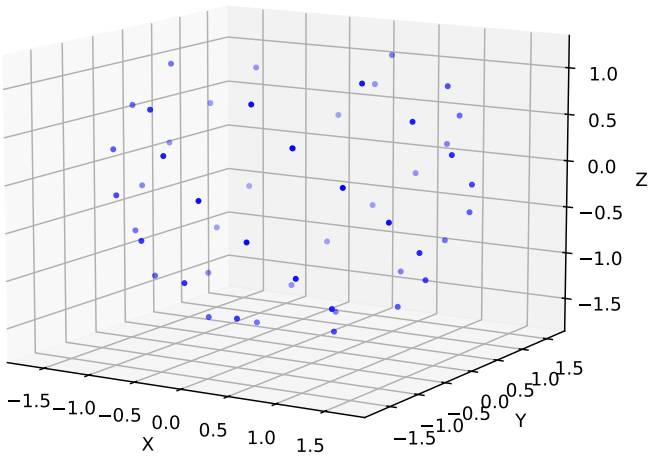


Figure 4.2: The "*Non-Uniform Scenario*", where the 50 training points are not uniformly sampling the sphere

cally selected a dimensionality configuration, through a process of hyperparameter tuning. Specifically, the PINN consisted of 13 layers, each with 160 neurons.

4.5. Results

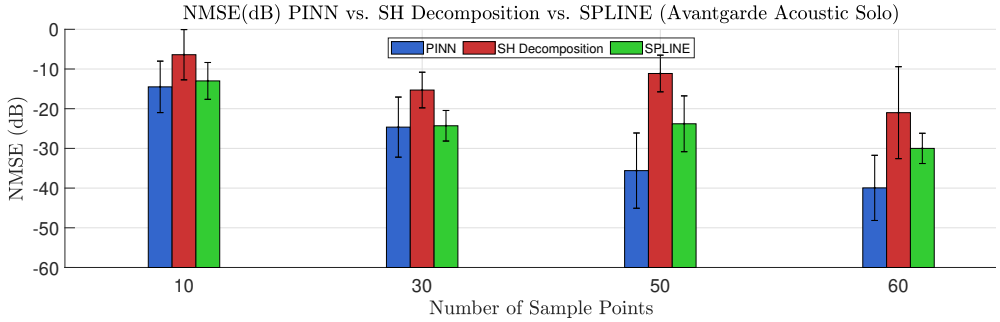
In the following section we present the outcomes, showcasing graphical representations along with corresponding insights and comments. We aim to offer an overview of the PINN method performance in sound field reconstruction with respect to the baselines. We begin by displaying results for the average NMSE and frequency-dependent NCC, evaluating the performance for the audio sources. Following that, we show various 3D balloon plots illustrating directivity reconstructions at low, medium, and high frequencies. These plots demonstrate how reconstruction quality varies with the number of points in the training dataset.

4.5.1. Average NMSE

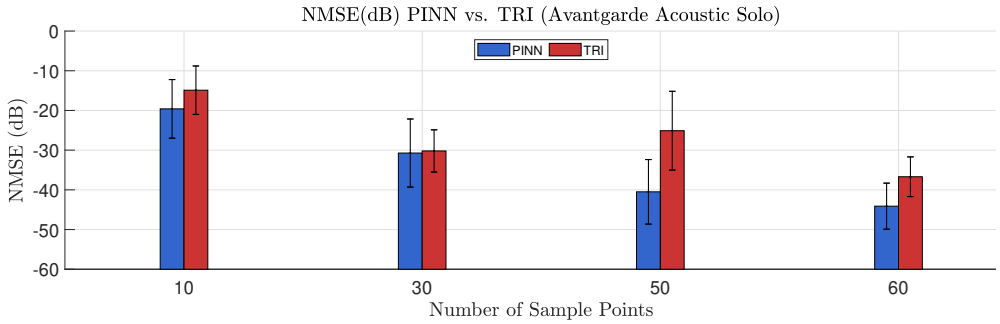
We exploit average NMSE as a metric to quantify the accuracy of the resulting sound field estimations. Details regarding how the metric was computed are ex-

plained in Sec. 4.2. In this section, a series of bar plots are presented in order to have a visual comparison of the resulting $\overline{\text{NMSE}}$ values, expressed in dB. For each sound source, we present two different bar plots. The first plot compares the $\overline{\text{NMSE}}$ of the PINN with respect to the SH and the SPLINE methods. The second plot shows a comparison between the resulting $\overline{\text{NMSE}}$ values of the PINN method with respect to the ones from the TRI technique. This separation is mandatory as the TRI method could only provide magnitude estimations of the directivity, as explained in Sec. 4.3, therefore inducing us to relate it to the magnitudes of the directivity predictions estimated by our PINN.

In each bar plot, the methods are compared at different training dataset dimensions, expressed as number of samples along the x-axis. Specifically, the interpolations have been executed with an increasing training dataset of 10, 30, 50 and 60 sample points. Each bar represent the average NMSE value of the corresponding interpolation technique, while the standard deviation is shown by the means of an error bar.

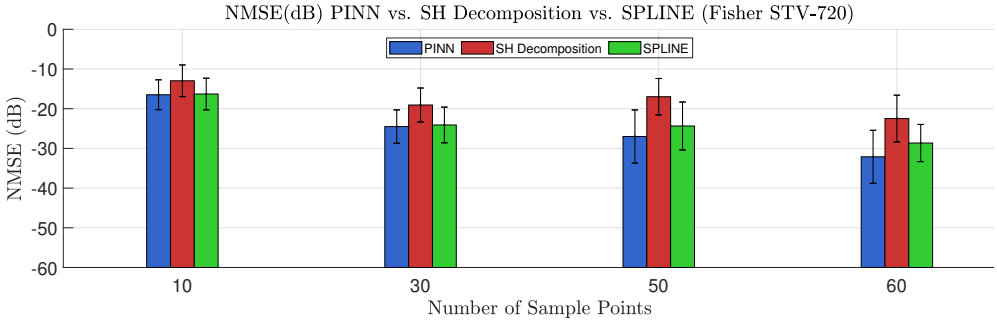


(a)

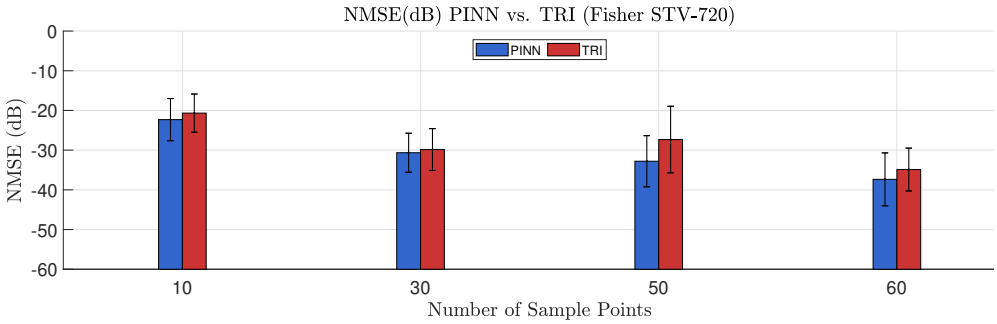


(b)

Figure 4.3: Average NMSE values for the reconstructed directivity of the *Avantgarde Acoustic Solo* loudspeaker. **(a)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the complex-valued reconstructions by means of PINN (blue bar), SH Decomposition (red bar) and SPLINE (green bar). **(b)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the magnitude reconstructions by means of PINN (blue bar) and TRI (red bar). The black error bars denote the corresponding standard deviations.

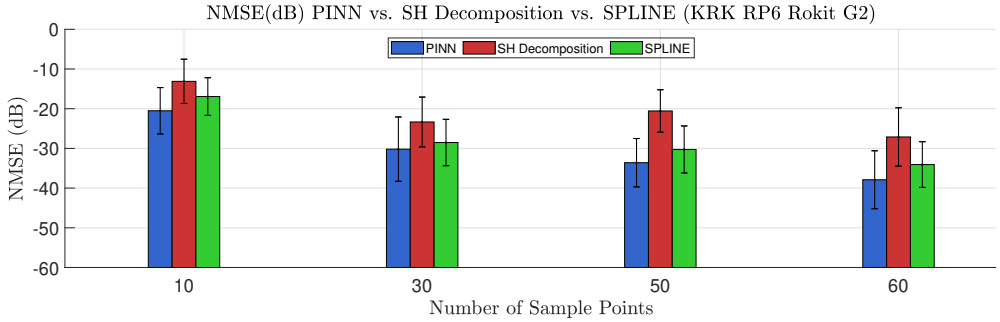


(a)

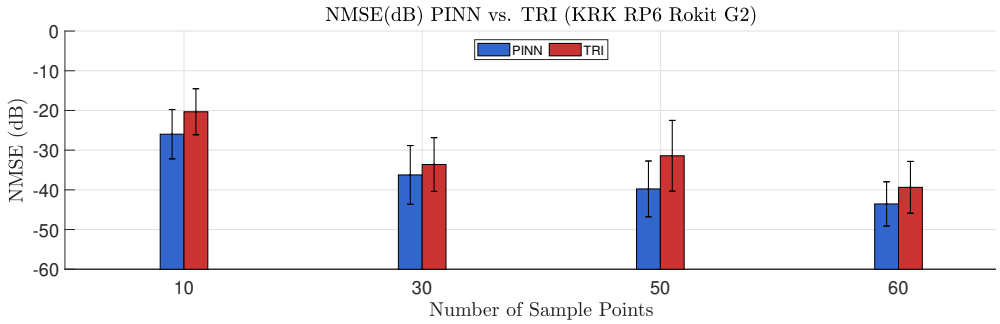


(b)

Figure 4.4: Average NMSE values for the reconstructed directivity of the *Fisher STV-720* loudspeaker. **(a)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the complex-valued reconstructions by means of PINN (blue bar), SH Decomposition (red bar) and SPLINE (green bar). **(b)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the magnitude reconstructions by means of PINN (blue bar) and TRI (red bar). The black error bars denote the corresponding standard deviations.

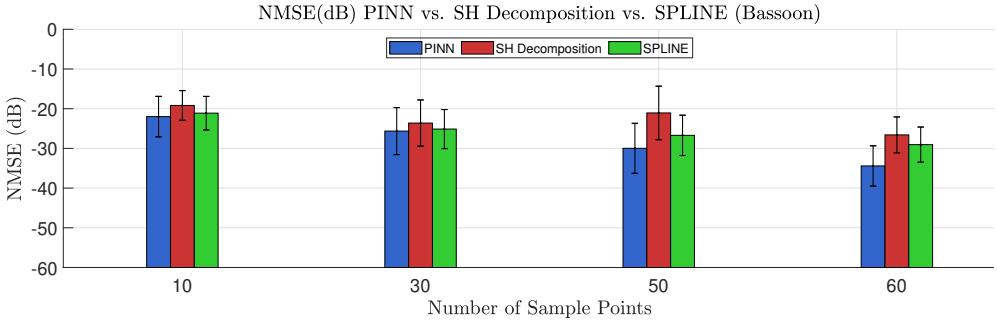


(a)

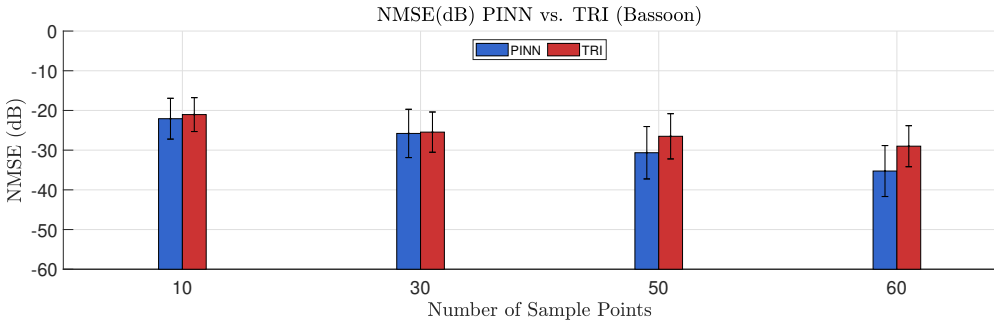


(b)

Figure 4.5: Average NMSE values for the reconstructed directivity of the *KRK RP6 Rokit G2* loudspeaker. **(a)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the complex-valued reconstructions by means of PINN (blue bar), SH Decomposition (red bar) and SPLINE (green bar). **(b)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the magnitude reconstructions by means of PINN (blue bar) and TRI (red bar). The black error bars denote the corresponding standard deviations.



(a)



(b)

Figure 4.6: Average NMSE values for the reconstructed directivity of the *Bassoon*. **(a)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the complex-valued reconstructions by means of PINN (blue bar), SH Decomposition (red bar) and SPLINE (green bar). **(b)** The $\overline{\text{NMSE}}$ values are reported in dB scale for the magnitude reconstructions by means of PINN (blue bar) and TRI (red bar). The black error bars denote the corresponding standard deviations.

By inspecting Fig. 4.3, it can be clearly noticed that the methods present better $\overline{\text{NMSE}}$ values as the number of sample points in the training dataset increases and an evenly distributed sampling on the sphere surface is maintained. The SPLINE interpolation resulted as a better traditional upsampling technique with respect to the SH method. However, The PINN method exhibits better performances, in terms of average NMSE values, in every case. As the training dataset expands, the differences between the PINN method and the baselines become more pronounced,

providing a clearer demonstration of the PINN's superiority in addressing the interpolation task. Examining the $\overline{\text{NMSE}}$ value obtained by the PINN for a training dataset of 50 sample points, it can be noticed how the proposed method stands out as the sole approach displaying a consistently decreasing trend in values, persistently decreasing the average NMSE even in the "*Non-Uniform Scenario*" (refer to Sec. 4.4). In contrast, traditional techniques appear to encounter greater challenges in interpolating the sound field when provided with a less uniformly sampled grid of data points. The same considerations can be applied to the other audio sources examined, as can be seen from Fig. 4.4, Fig. 4.5 and Fig. 4.6

4.5.2. Frequency-dependent NCC

We opted to use the NCC metric to provide a clearer representation of how well the shape of the polar directivity patterns conform to the ground truth (GT), depending on the frequency. Insights into our frequency-dependent computation of this metric are presented in Sec. 4.2. We chose to illustrate a specific case, namely the interpolation of the directivity for the *Fisher STV-720* sound source, in a scenario where the training dataset consists of 60 sample points. These graphs serve as an illustrative example of the NCC trends across frequencies, but they exhibit trajectories that we consistently observed in other plots depicting the NCC metric for different sources and varying sizes of the training dataset. In the following graphs, the NCC values are shown depending on the frequency. As the sampling frequency for all the measurements was $Fs = 48\,\text{kHz}$, we show the frequencies on the x-axis ranging from $0\,\text{kHz}$ to $24\,\text{kHz}$.

To ease a comprehensive comparison of NCC trajectories between the PINN method and the baseline, we present three distinct plots (Fig. 4.7, Fig. 4.8, Fig. 4.9). These illustrate the comparisons between the PINN method and SH, SPLINE, and TRI, respectively.

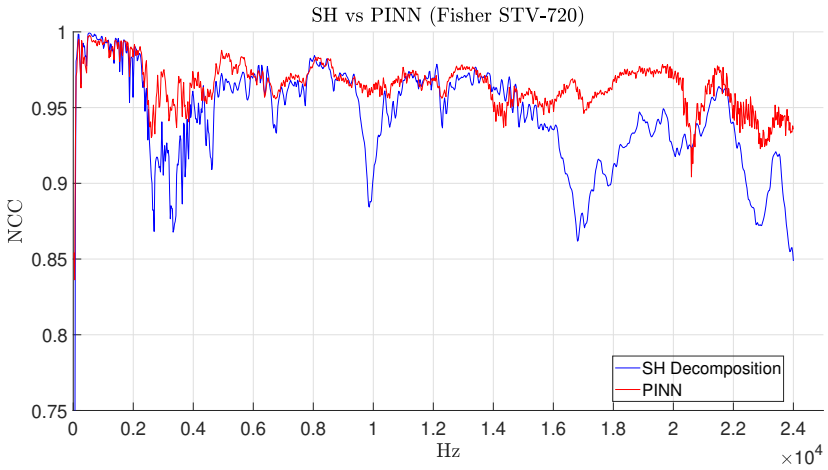


Figure 4.7: Frequency-dependent NCC values for the reconstructed directivity of the *Fisher STV-720* loudspeaker, employing a training dataset of 60 sample points. The NCC plots are reported for frequency values ranging from 0 kHz to 24 kHz for the PINN method (red line) and the SH method (blue line)

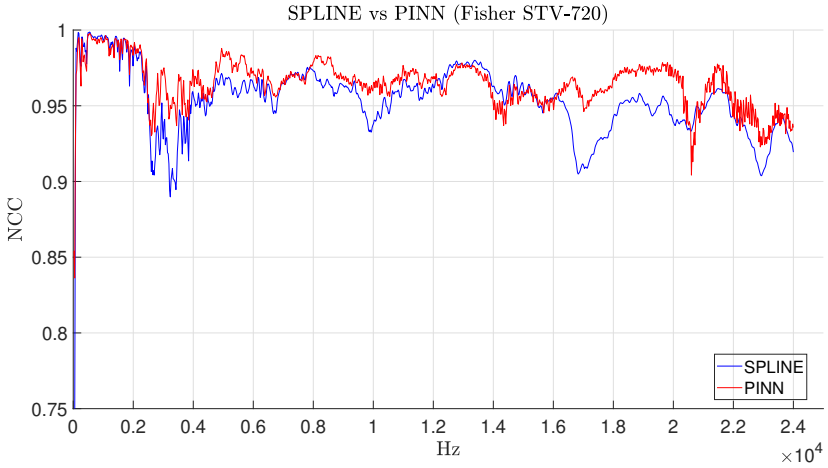


Figure 4.8: Frequency-dependent NCC values for the reconstructed directivity of the *Fisher STV-720* loudspeaker, employing a training dataset of 60 sample points. The NCC plots are reported for frequency values ranging from 0 kHz to 24 kHz for the PINN method (red line) and the SPLINE method (blue line)

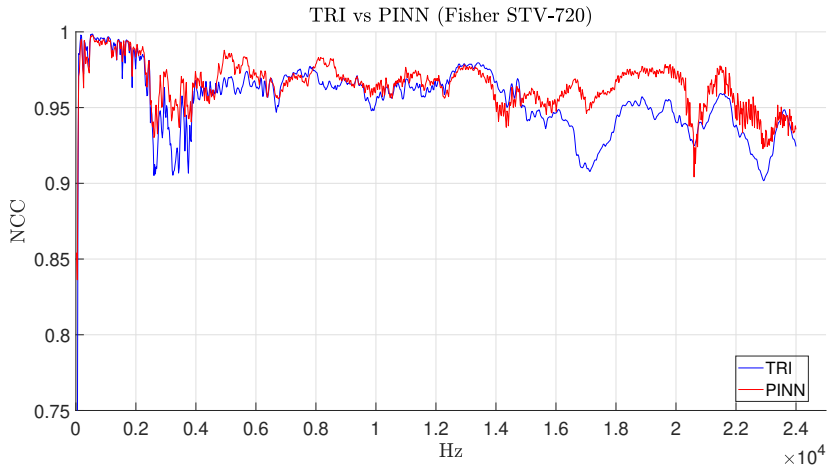


Figure 4.9: Frequency-dependent NCC values for the reconstructed directivity of the *Fisher STV-720* loudspeaker, employing a training dataset of 60 sample points. The NCC plots are reported for frequency values ranging from 0 kHz to 24 kHz for the PINN method (red line) and the TRI method (blue line)

While examining various directivity polar patterns of different sound sources, it became evident that sound sources exhibit more intricate polar patterns in the medium frequency range, approximately between 2 kHz and 4 kHz. For these reasons, the plots demonstrate a reduced NCC resulting value within that range. A similar trend is illustrated in a higher frequencies range between 16 kHz and 18 kHz. However, it is important to emphasize that, within this frequency range, the complexity of the directivity pattern shapes may be attributed to the inability of certain sources to accurately reproduce sound fields at such high frequencies, due to physical limitations. For analogous reasons, values beyond the 18 kHz frequency range are essentially negligible, as the sound sources are not designed to exhibit a meaningful directivity polar pattern within that spectrum. The PINN method demonstrates superior results in the range where the complexity of directivity pattern shapes increases compared to traditional baseline methods. This behavior is notably evident when compared to the SH method, as illustrated in Fig. 4.7. Overall, the PINN method demonstrates better outcomes in the majority of frequency ranges, even in instances where the ground truth directivity shapes are less intricate, such as in the lower frequency range. However, in these cases,

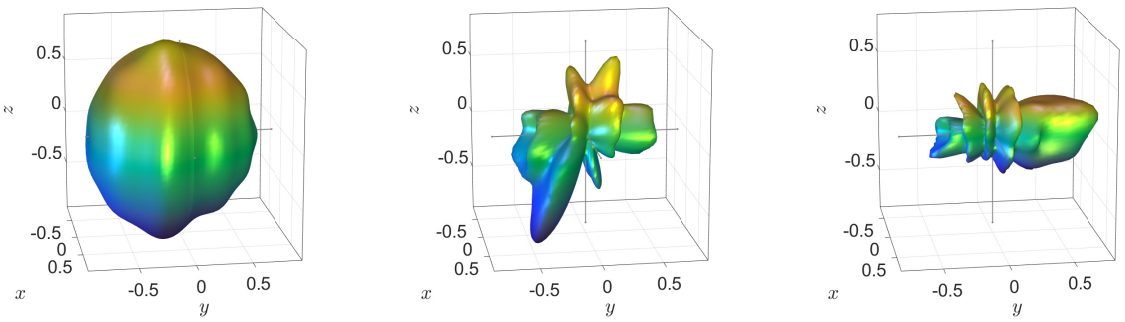
	PINN	SH	SPLINE	TRI
$\overline{\text{NCC}}$ ($\pm$ std)	**0.965** (**$\pm$0.015**)	0.936 ($\pm$0.035)	0.954 ($\pm$0.023)	0.955 ($\pm$0.024)

Table 4.1: Average $\overline{\text{NCC}}$ along with the relative standard deviations for the reconstructed directivity of the *Fisher STV-720*, computed across all the frequencies, for the PINN and the comparison methods

disparities are not significantly appreciable. To better confirm this interpretation of the graphs, Table 4.1 shows the average $\overline{\text{NCC}}$ computed across all the frequencies for the *Fisher STV-720* sound source, along with their respective standard deviations.

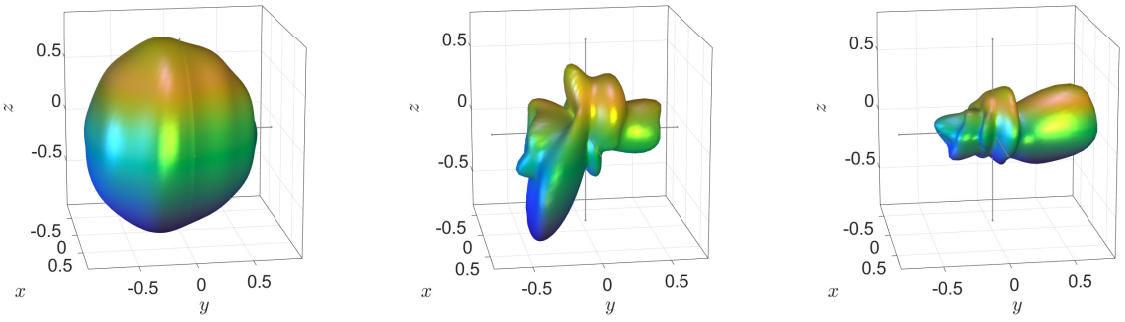
4.5.3. 3D directivity plots

The following figures illustrate a series of examples of 3D directivity polar patterns reconstructed using the evaluated methods, compared with the ground truth. We selected patterns from the *Bassoon* source due to its complex directivity characteristics. Three distinct polar patterns are presented for the ground truth (Fig.4.10) and for the corresponding interpolated patterns obtained with the methods under examination: PINN (Fig.4.11), SH (Fig.4.12), SPLINE (Fig.4.13), and TRI (Fig. 4.14). Each figure displays three polar patterns representing the directivity characteristics of the *Bassoon* source at low frequencies ($f = 176\,\text{Hz}$), mid frequencies ($f = 872\,\text{Hz}$), and high frequencies ($f = 1920\,\text{Hz}$). The interpolated results are derived using a training dataset grid of 60 sample points. The plots are depicted in a Cartesian space to facilitate the interpretation of the outgoing directions relative to the source.



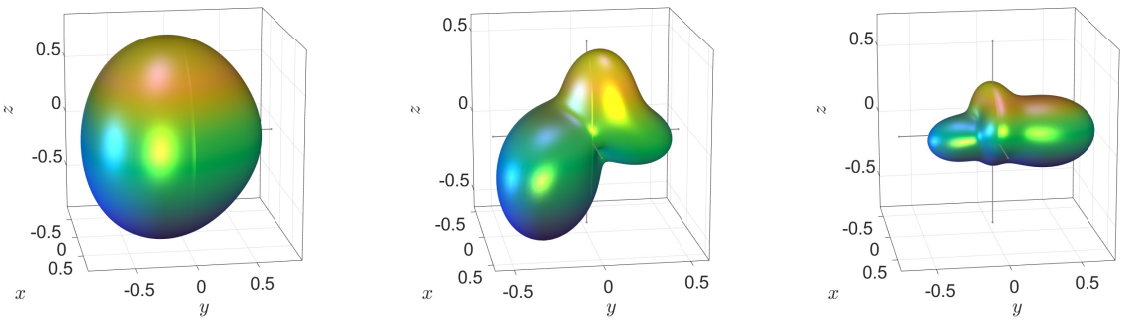
(a) GT LF. (b) GT MF. (c) GT HF.

Figure 4.10: 3D ground truth bassoon directivity polar patterns at: **(a)** $f = 176\,\text{Hz}$, **(b)** $f = 872\,\text{Hz}$ and **(c)** $f = 1920\,\text{Hz}$.



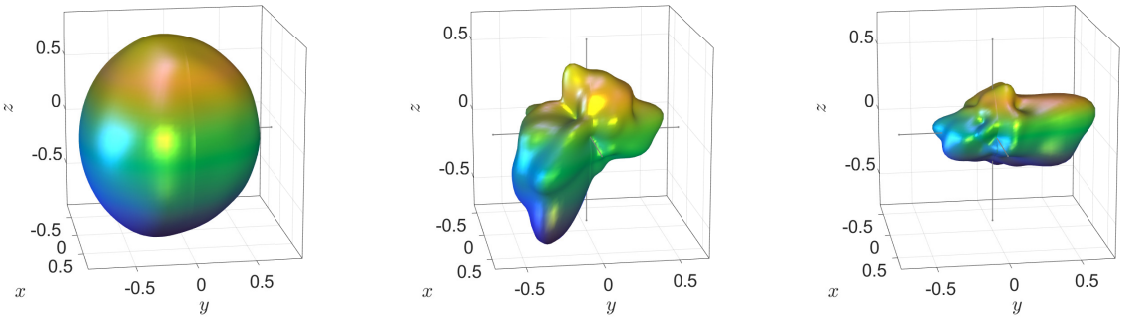
(a) PINN LF. (b) PINN MF. (c) PINN HF.

Figure 4.11: 3D bassoon directivity polar patterns interpolated by the means of the PINN method from 60 measurements at: **(a)** $f = 176\,\text{Hz}$, **(b)** $f = 872\,\text{Hz}$ and **(c)** $f = 1920\,\text{Hz}$.



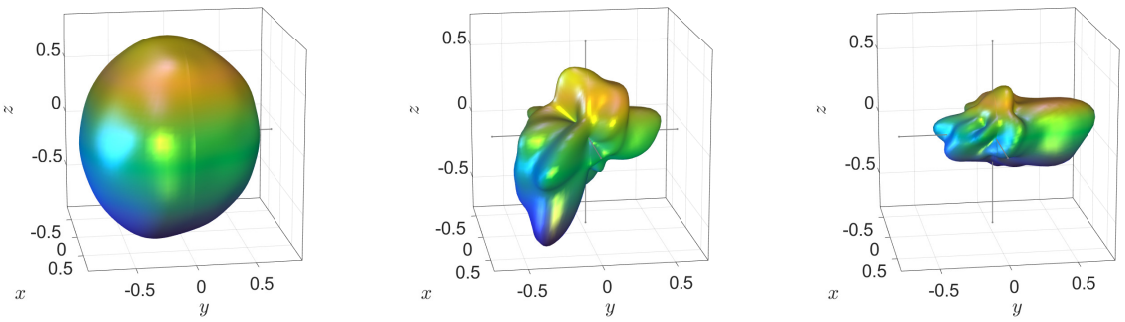
(a) SH LF. (b) SH MF. (c) SH HF.

Figure 4.12: 3D bassoon directivity polar patterns interpolated by the means of the SH method from 60 measurements at: **(a)** $f = 176\,\text{Hz}$, **(b)** $f = 872\,\text{Hz}$ and **(c)** $f = 1920\,\text{Hz}$.



(a) SPLINE LF. (b) SPLINE MF. (c) SPLINE HF.

Figure 4.13: 3D bassoon directivity polar patterns interpolated by the means of the SPLINE method from 60 measurements at: **(a)** $f = 176\,\text{Hz}$, **(b)** $f = 872\,\text{Hz}$ and **(c)** $f = 1920\,\text{Hz}$.



(a) TRI LF. (b) TRI MF. (c) TRI HF.

Figure 4.14: 3D bassoon directivity polar patterns interpolated by the means of the TRI method from 60 measurements at: **(a)** $f = 176\,\text{Hz}$, **(b)** $f = 872\,\text{Hz}$ and **(c)** $f = 1920\,\text{Hz}$.

By observing the figures, it can be noticed how the PINN method interpolates the initial directivity measurement grid in a more detailed way than the baseline methods, thus resembling more accurately the ground truth directivity plot shape. The superior performance of the PINN is particularly noticeable when addressing more complex directivity polar patterns, specifically in the mid-high frequency range. It is apparent that the SH method, when utilizing a small grid of sample points for interpolation, fails to capture the finer details of the plots. In contrast, the SPLINE and TRI methods perform better in this regard, producing interpolated results that closely resemble each other.

5 | Conclusions

In this thesis, we propose a PINN-based method that tackles the problem of sound source directivity interpolation. Sparsely-sampled directivity measurements are often unavoidable because of practical limitations. Therefore, we presented and evaluate a physics-informed neural network-based interpolation method that addresses this problem setup, by exploiting the physical law underlying the directivity data, namely the Helmholtz partial differential equation. The PINN is able to provide as output an estimate of the directivity in any arbitrary direction, potentially reconstructing it in a continuous fashion. The proposed method has been evaluated on a dataset of directivity measurements, captured from different audio sources. The evaluation of the proposed method relies on a comparison with a series of state-of-the-art interpolation approaches of [2], namely SH interpolation, SPLINE, and TRI. The assessment of the performance of the proposed method and the baselines have been accomplished in terms of two different metrics, i.e., average $\overline{\text{NMSE}}$ and frequency-dependent NCC. The metrics demonstrated a superior capability of the PINN method, with respect to the comparison methods, in the reconstruction of the directivity of the examined sources starting from a sparsely-sampled grid. To confirm what the numerical metrics indicate, a visual feedback is displayed by the means of three dimensional polar directivity patterns. While SH, SPLINE, and TRI also exhibit remarkable results when the initial measurements are sufficiently dense, they do not perform as well in cases of sparsely-sampled measurements, such as those we have analyzed in this research work. It is reasonable to think that the achievement of the PINN may be attributed to the integration, within its architecture, of the physical laws governing the signal to be interpolated. Different improvements can be applied to the proposed method. Currently, a different PINN model has to be trained individually for each audio source. Drawing inspiration

from the methodology presented in [59] for global HRTF personalization, there is potential to upgrade this process. By extrapolating and harnessing the physical features of the audio sources, such as their dimensionality, it may be possible to conditionally train a single PINN model, capable of predicting the directivity measurements for different audio sources, enhancing the model's versatility.

Bibliography

[1] T. D. Abhayapala and D. B. Ward. Theory and design of high order sound field microphones using spherical microphone array. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II–1949–II–1952, 2002. doi: 10.1109/ICASSP.2002.5745011.

[2] D. Ackermann, F. Brinkmann, and F. Zotter. Comparative evaluation of interpolation methods for the directivity of musical instruments. *J AUDIO SPEECH MUSIC PROC.*, 2021.

[3] J. Ahrens. Database of spherical harmonic representations of sound source directivities, 2023. URL `https://github.com/AppliedAcousticsChalmers/sound-source-directivities`.

[4] J. Ahrens and S. Bilbao. Interpolation and range extrapolation of sound source directivity based on a spherical wave propagation model. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4662–4666, 2020. doi: 10.1109/ICASSP40776.2020.9053099.

[5] J. Ahrens and S. Bilbao. Computation of spherical harmonic representations of source directivity based on the finite-distance signature. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:83–92, 2021. ISSN 2329-9290, 2329-9304. doi: 10.1109/TASLP.2020.3037471. URL `https://ieeexplore.ieee.org/document/9257177/`.

[6] J. Ahrens, H. Helmholz, D. L. Alon, and S. V. A. Garí. Spherical harmonic decomposition of a sound field based on microphones around the circumference of a human head. In *2021 IEEE Workshop on Applications of Signal*

Processing to Audio and Acoustics (WASPAA), pages 231–235, 2021. doi: 10.1109/WASPAA52581.2021.9632751.

[7] P. Alfeld, M. Neamtu, and L. L. Schumaker. Bernstein-bézier polynomials on spheres and sphere-like surfaces. *Computer Aided Geometric Design*, 13(4): 333–349, 1996. ISSN 0167-8396. doi: https://doi.org/10.1016/0167-8396(95)00030-5. URL `https://www.sciencedirect.com/science/article/pii/0167839695000305`.

[8] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa. The quickhull algorithm for convex hulls. 22(4):469–483, dec 1996. ISSN 0098-3500. doi: 10.1145/235815.235821. URL `https://doi.org/10.1145/235815.235821`.

[9] J. Benesty, J. Chen, and Y. Huang. *Microphone Array Signal Processing*. 2008. ISBN 978-3-540-78611-5.

[10] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1996. ISBN 9780262268684.

[11] F. Brinkmann and S. Weinzierl. Aktools - the open software toolbox for signal acquisition, processing, and inspection in acoustics. 02 2017.

[12] S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis. Physics-informed neural networks (pinns) for fluid mechanics: A review, 2021.

[13] A. Canclini, F. Antonacci, S. Tubaro, and A. Sarti. A methodology for the robust estimation of the radiation pattern of acoustic sources. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:211–224, 2020.

[14] R. Caussé and P. Dérogis. Radiation of musical instruments and improvement of the sound diffusion techniques for synthesized, recorded, or amplified sounds (revisited). In *Proceedings of the 1995 International Computer Music Conference, ICMC 1995, Banff, AB, Canada, September 3-7, 1995*, pages 359—-360. Michigan Publishing, 1995. URL `https://hdl.handle.net/2027/spo.bbp2372.1995.105`.

[15] M. Cobos, M. Pezzoli, F. Antonacci, and A. Sarti. Acoustic source lo-

calization in the spherical harmonics domain exploiting low-rank approximations. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, June 2023. doi: 10.1109/icassp49357.2023.10095324. URL http://dx.doi.org/10.1109/ICASSP49357.2023.10095324.

[16] P. R. Cook and D. Trueman. A database of measured musical instrument body radiation impulse responses , and computer applications for exploring and utilizing the measured filter functions. 1998. URL https://api.semanticscholar.org/CorpusID:17971082.

[17] J. Ehret, B. Weyers, A. Bönsch, and J. Stienen. Does the directivity of a virtual agent's speech influence the perceived social presence? *IEEE VR Workshop on Virtual Humans and Crowds in Immersive Environments (VHCIE)*, 2018.

[18] N. Fletcher and T. Rossing. *The physics of musical instruments*. Springer-Verlag, 1997. ISBN 3540983759.

[19] T. Grothe and M. Kob. Bassoon directivity data, 2020.

[20] I. Hagai, M. Pollow, M. Vorländer, and B. Rafaely. Acoustic centering of sources measured by surrounding spherical microphone arrays. *J. Acoust. Soc. Am.*, pages 2003–2015, 2011.

[21] 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*, Mar 1999. URL http://www.aes.org/e-lib/browse.cfm?elib=8026.

[22] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks929 are universal approximators. *Neural networks 2*, pages 2149–2157, 1989.

[23] S. Karimpouli and P. Tahmasebi. Physics informed machine learning: Seismic wave equation. *Geoscience Frontiers*, 11(6):1993–2001, 2020. ISSN 1674-9871. doi: https://doi.org/10.1016/j.gsf.2020.07.007. URL https://www.sciencedirect.com/science/article/pii/S1674987120301717.

[24] G. Karniadakis, Y. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. pages 1–19, 05 2021. doi: 10.1038/s42254-021-00314-5.

[25] W. Keller and A. Borkowski. Thin plate spline interpolation. *Journal of Geodesy*, 93, 02 2019. doi: 10.1007/s00190-019-01240-2.

[26] J. Klein and M. Vorländer. Simulative investigation of required spatial source resolution in directional room impulse response measurements. *EAA Spatial Audio Signal Processing Symposium*, pages 37–42, 2019.

[27] M. Laitinen, A. Politis, I. Huhtakallio, and V. Pulkki. Controlling the perceived distance of an auditory object by manipulation of loudspeaker directivity. *The Journal of the Acoustical Society of America*, 137:EL462, 2015.

[28] C. Lee, H. Hasegawa, and S. Gao. Complex-valued neural networks: A comprehensive survey. *IEEE/CAA Journal of Automatica Sinica*, 9(8):1406–1426, 2022. doi: 10.1109/JAS.2022.105743.

[29] F. Ma, T. D. Abhayapala, P. N. Samarasinghe, and X. Chen. Spatial upsampling of head-related transfer functions using a physics-informed neural network, 2023.

[30] F. Ma, S. Zhao, and I. S. Burnett. Sound field reconstruction using a compact acoustics-informed neural network, 2024.

[31] R. Mehra and D. Manocha. Wave-based sound propagation for vr applications. *2014 IEEE VR Workshop: Sonic Interaction in Virtual Environments, SIVE 2014*, pages 41–46, 01 2015. doi: 10.1109/SIVE.2014.7006289.

[32] J. Meyer. *Acoustics and the Performance of Music*. Springer, 5 edition, 2009. ISBN 978-0-387-09516-5.

[33] F. Otondo and J. Rindler. The influence of the directivity of musical instruments in a room. *Acta Acustica united with Acustica*, pages 1178–1184, 2004.

[34] M. Pezzoli, A. Canclini, F. Antonacci, and A. Sarti. A comparative analysis

of the directional sound radiation of historical violins. *The Journal of the Acoustical Society of America*, 152(1):354–367, 2022.

[35] 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)*, pages 1051–1055, 2022. doi: 10.1109/ICASSP43922.2022.9746391.

[36] M. Pezzoli, F. Antonacci, and A. Sarti. Implicit neural representation with physics-informed neural networks for the reconstruction of the early part of room impulse responses. In *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, FA2023. European Acoustics Association, Jan. 2024. doi: 10.61782/fa.2023.1182. URL `http://dx.doi.org/10.61782/fa.2023.1182`.

[37] M. Pezzoli, J. Carabias-Orti, P. Vera-Candeas, F. Antonacci, and A. Sarti. Spherical-harmonics-based sound field decomposition and multichannel nmf for sound source separation. *Applied Acoustics*, 218:109888, 2024.

[38] M. Pollow, G. Behler, and B. Masiero. Measuring directivities of natural sound sources with a spherical microphone array. page 6 S., 01 2009.

[39] V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc*, 45(6):456–466, 1997. URL `http://www.aes.org/e-lib/browse.cfm?elib=7853`.

[40] B. Rafaely. Plane-wave decomposition of the sound field on a sphere by spherical convolution. *J. Acoust. Soc. Am.*, pages 2149–2157, 2004.

[41] B. Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 13(1):135–143, 2005. doi: 10.1109/TSA.2004.839244.

[42] B. Rafaely. *Fundamentals of Spherical Array Processing*. Springer Cham, 2 edition, 2018. ISBN 978-3-319-99560-1.

[43] B. Rafaely, B. Weiss, and E. Bachmat. Spatial aliasing in spherical microphone

arrays. *IEEE Transactions on Signal Processing*, 55(3):1003–1010, 2007. doi: 10.1109/TSP.2006.888896.

[44] M. Raissi, P. Perdikaris, and G. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL `https://www.sciencedirect.com/science/article/pii/S0021999118307125`.

[45] T. Rossing, R. Moore, P. Wheeler, and F. Moore. *The science of sound*. Addison-Wesley, 3 edition, 2002. ISBN 0805385657.

[46] P. N. Samarasinghe, T. D. Abhayapala, Y. Lu, H. Chen, and G. Dickins. Spherical harmonics based generalized image source method for simulating room acoustics. *The Journal of the Acoustical Society of America*, 144(3): 1381–1391, 09 2018. ISSN 0001-4966. doi: 10.1121/1.5053579. URL `https://doi.org/10.1121/1.5053579`.

[47] K. Sel and R. A. Mohammadi, Pettigrew. Physics-informed neural networks for modeling physiological time series for cuffless blood pressure estimation. *npj Digit. Med.*, page 110, 2023. doi: https://doi.org/10.1038/s41746-023-00853-4.

[48] Sennheiser. Ke 4-211-2. *002280*, 2013.

[49] N. Shabtai and M. Vorländer. Acoustic centering of sources with high-order radiation patterns. *J. Acoust. Soc. Am.*, pages 1947–1961, 2015.

[50] N. Shabtai, G. Behler, M. Vorländer, and S. Weinzierl. Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments. *The Journal of the Acoustical Society of America*, pages 1246–1256, 2017.

[51] G. Simpson and Y. H. Wu. Accuracy and effort of interpolation and sampling: Can gis help lower field costs? *ISPRS International Journal of Geo-Information*, 3(4):1317–1333, 2014. ISSN 2220-9964. doi: 10.3390/ijgi3041317. URL `https://www.mdpi.com/2220-9964/3/4/1317`.

[52] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions, 2020.

[53] I. Sloan and R. Womersley. Extremal systems of points and numerical integration on the sphere. *Advances in Computational Mathematics*, pages 107–125, 2004.

[54] C. Song, T. Alkhalifah, and U. Bin Waheed. A versatile framework to solve the helmholtz equation using physics-informed neural networks. *Geophysical Journal International*, 228, 10 2021. doi: 10.1093/gji/ggab434.

[55] J. Tylka, R. Sridhar, and E. Choueiri. A database of loudspeaker polar radiation measurements. *Audio Engineering Society Convention*, page 139, 2015.

[56] J. G. Tylka, R. Sridhar, and E. Choueiri. A database of loudspeaker polar radiation measurements. In *Audio Engineering Society Convention 139*. Audio Engineering Society, 2015.

[57] S. Verburg and E. Grande. Reconstruction of the sound field in a room using compressive sensing. *J. Acoust. Soc. Am.*, pages 3770–3779, 2018.

[58] G. Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981. doi: 10.1137/0902002. URL `https://doi.org/10.1137/0902002`.

[59] Y. Wang, Y. Zhang, Z. Duan, and M. Bocko. Global hrtf personalization using anthropometric measures. In *Audio Engineering Society Convention 150*, May 2021. URL `https://www.aes.org/e-lib/browse.cfm?elib=21095`.

[60] E. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999. ISBN 978-0-12-753960-7.

[61] F. Zotter. *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*. PhD thesis, 01 2009.

[62] F. Zotter and M. Frank. *A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer Cham, 1 edition, 2019. ISBN 978-3-030-17206-0.

List of Figures

List of Tables

Acknowledgements

Questa tesi rappresenta il raggiungimento di un traguardo molto significativo per la mia personale crescita. La sua stesura ha richiesto molto impegno, e non sarebbe stata possibile senza la supervisione e i preziosi consigli del Prof. Pezzoli, il Dr. Malvermi e il Dr. Olivieri, ai quali voglio dedicare un sincero grazie. Altrettanto sincero e altrettanto dovuto è il ringraziamento che voglio porgere alla mia famiglia, e in particolare ai miei genitori, per il loro incommensurabile supporto e per aver sempre creduto in me. Grazie Silvia, Federico, Arianna e Stefano. Per merito vostro questi mesi, non privi di difficoltà, sono stati fondamentali per la mia maturazione. Spero di potervi continuare a dimostrare la mia gratitudine negli anni a venire, regalandovi sempre molte soddisfazioni. Un sincero grazie va anche alle persone, amici e parenti, che direttamente o indirettamente hanno contribuito alla realizzazione di questo importante passo della mia vita.

Con amore, Edoardo.