applied sciences

MDPI

Review

A Review on Head-Related Transfer Function Generation for Spatial Audio

Valeria Bruschi [1,*], Loris Grossi [2], Nefeli A. Dourou [1], Andrea Quattrini [2], Alberto Vancheri [2], Tiziano Leidi [2,*] and Stefania Cecchi [1]

[1] Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60121 Ancona, Italy; n.a.dourou@pm.univpm.it (N.A.D.); s.cecchi@staff.univpm.it (S.C.)
[2] Institute of Information Systems and Networking, University of Applied Sciences and Arts of Southern Switzerland, 6928 Manno, Switzerland; loris.grossi@supsi.ch (L.G.); andrea.quattrini@supsi.ch (A.Q.); alberto.vancheri@supsi.ch (A.V.)
* Correspondence: v.bruschi@staff.univpm.it (V.B.); tiziano.leidi@supsi.ch (T.L.); Tel.: +39-071-2204-486 (V.B.); +41-58-6666-618 (T.L.)

Abstract: A head-related transfer function (HRTF) is a mathematical model that describes the acoustic path between a sound source and a listener's ear. Using binaural synthesis techniques, HRTFs play a crucial role in creating immersive audio experiences through headphones or loudspeakers, using binaural synthesis techniques. HRTF measurements can be conducted either with standardised mannequins or with in-ear microphones on real subjects. However, various challenges arise in, for example, individual differences in head shape, pinnae geometry, and torso dimensions, as well as in the extensive number of measurements required for optimal audio immersion. To address these issues, numerous methods have been developed to generate new HRTFs from existing data or through computer simulations. This review paper provides an overview of the current approaches and technologies for generating, adapting, and optimising HRTFs, with a focus on physical modelling, anthropometric techniques, machine learning methods, interpolation strategies, and their practical applications.

Keywords: spatial audio processing; head related transfer function (HRTF); HRTF measurement; HRTF numerical modelling; HRTF personalisation; HRTF anthropometric and morphological approaches; HRTF interpolation

Citation: Bruschi, V.; Grossi, L.; Dourou, N.A.; Quattrini, A.; Vancheri, A.; Leidi, T.; Cecchi. S. A Review on Head-Related Transfer Function Generation for Spatial Audio. *Appl. Sci.* **2024**, *14*, 11242. https://doi.org/10.3390/app142311242

Academic Editors: Nikolaos Vryzas and Lazaros Vrysis

Received: 17 October 2024
Revised: 11 November 2024
Accepted: 25 November 2024
Published: 2 December 2024

1. Introduction

The mathematical model describing the acoustic path between a sound source and a listener's ear is represented by the head-related transfer function (HRTF) in the frequency domain and by the head-related impulse response (HRIR) in the time domain. Therefore, for each position of the sound source, two HRTFs are defined, one for the left ear and another for the right ear. HRTFs are unique to each individual, as they are heavily influenced by the reflections of sound waves on the upper body; as such, they are also influenced by the shape of the head, pinnae geometry, and torso dimensions. HRTFs encapsulate all the auditory cues that the human sound perception system (ears and brain) uses to discern the direction and distance of a sound source, such as the perceived sound differences between the ears (i.e., interaural time difference (ITD) and interaural level difference (ILD)) and the spectral cues [1].

The properties of these functions enable their application in the development of spatial audio systems. These systems can be derived, for example, by filtering an audio signal with specific HRTFs [2]. When these signals are reproduced through headphones, the listener perceives the sound source as originating from the original position encoded in the HRTFs. This technique is known as binaural synthesis. HRTFs are also exploited in the reproduction of binaural signals through loudspeakers. In this scenario, the sound

emitted from each loudspeaker is perceived by both ears, and a network of filters, known as a crosstalk canceller, is typically employed to mitigate this issue. The crosstalk canceller is obtained by the inversion of the 2×2 HRTFs matrix related to the position of the two loudspeakers with respect to the two ears [3]. In recent years, considerable research efforts have been focused on developing advanced crosstalk cancellation techniques that utilise HRTFs to achieve optimal sound immersion as described, for instance, in [4–12]. These techniques aim to eliminate the undesired effects caused by sound reaching both ears from multiple loudspeakers, thereby enhancing spatial audio accuracy. By leveraging HRTFs, crosstalk cancellation methods can deliver a more precise and immersive listening experience, making them crucial for applications in virtual and augmented reality.

Numerous studies have investigated the effectiveness of individual HRTFs in spatial audio systems. Tailor-made HRTFs are considered to provide the most accurate 3D audio experience. Furthermore, research, such as that conducted by Majdak et al. [13], has demonstrated that using generic HRTFs often leads to suboptimal perception of sound source localisation. The use of non-personalised HRTFs can introduce unnatural sound colouration and other artefacts as highlighted, for example, by Brinkmann [14]. In [15,16], it has been demonstrated that source localisation is enhanced when individual HRTFs are employed. However, in studies such as [17], it has been observed that personalised HRTFs are not always the preferred subjective choice; this may be attributed to issues arising during the measurement process. In [18], the impact of head movements during HRTF measurements has been assessed, revealing that pitch movements (i.e., the upward or downward motions of the head) have a more significant influence than roll, and yaw. These movements can introduce spectral variations in the HRTFs of up to 6 dB. To mitigate this issue, a neck support can be used for the listener during the measurement process [19], alternatively, databases of individual HRTFs can be developed [20,21], incorporating information about the azimuth and elevation of the sound source, as well as the listener's head orientation as recorded by a tracking system.

In this context, it is evident that HRTFs are essential for creating immersive 3D audio experiences, with several factors requiring careful consideration. This extended review covers a wide range of approaches and technologies aimed at generating HRTFs. Starting by introducing measurement techniques and physical modelling, this paper will focus on anthropometric and morphological approaches for HRTF personalisation and interpolation methods for HRTF generation. An overview of the methodologies for HRTF generation discussed in this paper can be found in Figure 1.

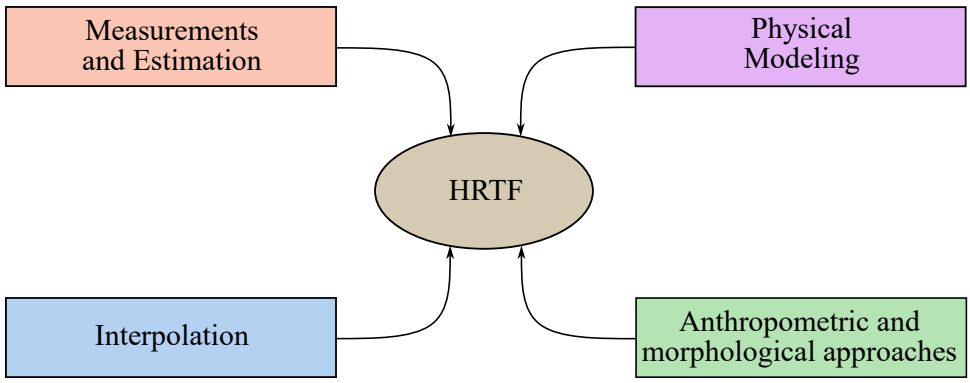


Figure 1. Scheme of the methodologies for HRTF generation presented in this paper.

The rest of this paper is organised as follows. Section 2 provides a definition of ans HRTF and the reference system used to localise the sound sources relative to the head, along with the mathematical framework underlying this definition. Section 3 reports a short overview of the methods used for measuring and estimating HRTFs and other complementary data such as 3D meshes. Section 4 reports the process of HRTF physical modelling. Section 5 describes data-driven anthropometric and morphological approaches

focusing on HRTFs personalisation. Section 6 analyses HRTFs interpolation algorithms. Its integration into consumer products and a real-world evaluation are reported in Section 7. Finally, conclusions are drawn in Section 8.

2. General Setting and Definitions

In this section, a clear mathematical definition of an HRTF along with the reference system used for measurements is provided. In more detail, Sections 2.1 and 2.2 show the HRTF definition and some metrics to evaluate the distance between a measured and a generated HRTF, respectively. Finally, the mathematical framework behind the definition of the HRTF is described in Section 2.3.

2.1. Mathematical Definition of HRTF

As we will see in Section 2.3 in more detail, the general setting for the definition of the HRTF includes a point source S and a scatterer, usually the head and torso of a human individual, which are both placed in an anechoic environment. A receiver R is placed at a point near the entrance of the ear channel of the person as shown in Figure 2. In order to give a precise definition of the HRTF, we need to define a fixed reference system $Oxyz$ with respect to the head, where the origin O is a reference point conventionally identified with the centre of the head (for instance, the midpoint between the ears). Usually, the xy plane is the horizontal plane of the head, with x pointing in the frontal direction and y running through the ear channel from the right to the left ear, and z is such that $Oxyz$ is right handed. Spherical coordinates (r, θ, ϕ) associated with $Oxyz$, with θ the azimuth angle and ϕ the elevation angle, are also frequently used.

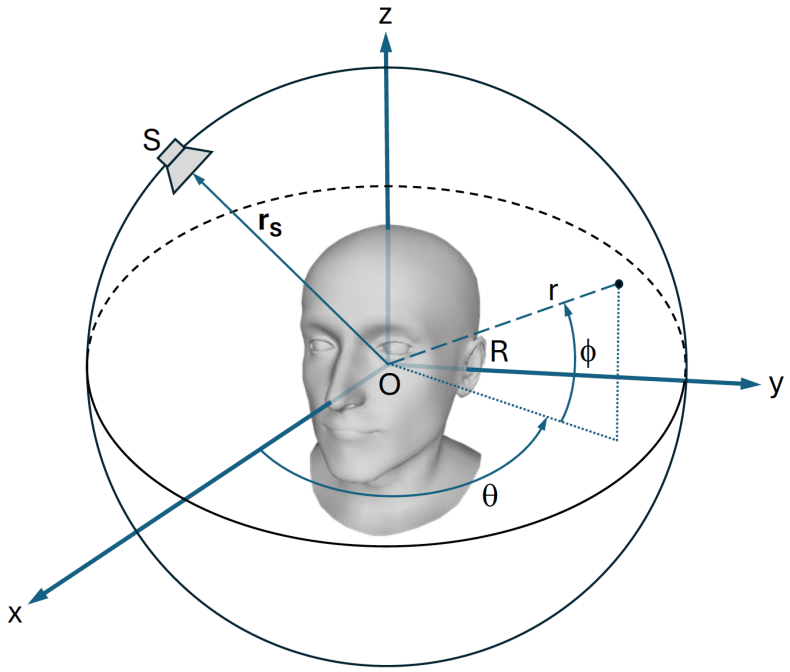


Figure 2. Reference system $Oxyz$ associated with the head and related spherical coordinates (r, θ, ϕ). The position of a sound source S is described by the position vector $\mathbf{r_S}$. The point R is the position of the receiver.

When a Dirac pulse $\delta(t)$ is sent by the source S, a signal $h(t)$, i.e., the HRIR, is received at the receiver R. Clearly, the HRIR is a function of the position of the point source. It is well known from the theory of linear, time-invariant (LTI) systems that, in a time-independent setting (for instance, the person must be in a static position), the HRIR contains all the information needed to reconstruct the signal received at R for every possible signal emitted by S. As mentioned above, the information encoded in the HRIR is expressed in the

frequency domain by the HRTF. The HRTF is usually normalised with sound pressure, which would be measured at the centre of the head O if the head were not present:

$$H_R(\omega, r, \theta, \phi) = \frac{p_R(\omega, r, \theta, \phi)}{p_O(\omega, r, \theta, \phi)}, \tag{1}$$

where $p_R(\omega, r, \theta, \phi)$ and $p_O(\omega, r, \theta, \phi)$ are the values of the pressure field, in the frequency domain, at the receiver R and at the centre of the head O (if the head were not present), respectively. The dependency of the HRTF from the position (r, θ, ϕ) of the source is explicitly indicated. The definition (1) also explicitly shows that the HRTF $H_R(\omega, r, \theta, \phi)$ depends on the position of the receiver R. Thus, one usually has two HRTF : one for each ear. When the source is on the opposite side of the head from the ear, it is called the contralateral HRTF. When the source is on the same side of the ear, it is called ipsilateral HRTF. In [22], it is shown that it is possible to transform an ipsilateral HRTF into a contralateral HRTF using a simple model of a spherical head and introducing an error of approximately 5° in sound localisation.

In experimental measurements, the normalisation of Equation (1) allows us to obtain an HRTF independent of the transfer functions of the devices used to reproduce and record the sound. In addition, normalisation allows the contribution of propagation in space from the source to the centre of the head to be separated from the HRTF.

2.2. Metrics for HRTF Evaluation

For a generated HRTF, quality evaluation is usually performed by comparing with a measured one by means of objective metrics. The difference between two responses can be estimated by the mean square error (MSE), which could be calculated both in the time [23] and frequency domains [24]. However, one of the frequently used metrics is the spectral distortion (SD), or log-spectral distortion (LSD) [25–27], defined as

$$\mathrm{SD} = \sqrt{\frac{1}{k_2 - k_1 + 1} \sum_{k=k_1}^{k_2} \left(20 \log_{10} \frac{|H(k)|}{|\hat{H}(k)|} \right)^2}, \tag{2}$$

where k_1 and k_2 define the lower and the higher limits of the evaluation band $B = [k_1 f_s/K, k_2 f_s/K]$, with K as the FFT length and f_s as the sampling frequency. $H(k)$ is the measured HRTF, and $\hat{H}(k)$ is the generated one.

To take into account the binaural nature of the HRTF, the error can also be determined by evaluating the difference in the localisation cues as the ILD [28,29] or the ITD [30]. In this regard, the just noticeable difference (JND), i.e., the limit below which no difference is perceived, can be evaluated [31]. A JND limit of 1 dB is considered for the ILD error to guarantee a significant perceptual reproduction [29], while the JND is estimated at around 8 µs/° for ITD, defining a maximal acceptable error of 16 µs [30].

However, the objective evaluation does not accurately estimate the realistic effect of the generation process on binaural reproduction. For this reason, subjective evaluation is essential for understanding the performance of the binaural audio, which is obtained by convolving the input signal with the proper HRTFs. Subjective tests aim to evaluate the localisation error, which quantifies the spatial distortion, or the sound quality, assessed through pair-wise comparisons [32]. In the first case, fixed sound source positions [27,28,33] or moving sound sources [32,34] can be considered. In the second case, the comparison can be performed among different measured HRTFs, in order to create an optimised HRTF subset [35], or between the measured HRTF and the simulated one [28]. In [27,33], fixed sound source positions are tested, and the timbre is evaluated by comparing the interpolated and measured HRTFs following the MUSHRA methodology. In [27], spatial impression and transparency, defined in ITU-R BS.1534-3 [36], are also included in the analysis. In [34], the listening tests are performed using both fixed and moving sound sources to also evaluate the effect of the continuous sound movement. Finally, in [37],

a perceptual evaluation of speech quality is carried out for an accurate analysis of speech enhancement and synthesis applications.

Subjective perception can also be simulated through models of sound directionality localisation. In [38], the authors propose a sagittal-plane localisation model (SPLM) based on the reproduction of some aspects of auditory system processes. This auditory model uses directional cues to create an internal model of the incoming sound and, in a second stage, compares this model with a template model learned by the individual. The difference between these two models is mapped to a probability that the subject localises the sound source in a given direction.

2.3. Mathematical Setting

In this section, we will formulate the mathematical setting underlying the definition of an HRTF given by Equation (1). In general terms, the problem can be formulated as a scattering problem, where a sound wave generated by a point source S is scattered by a body (the head and torso of an individual) in an anechoic environment. The scattering surface is indicated with Σ and the boundless simulation domain outside to Σ with Ω. The normal vector to the surface Σ is conventionally directed toward the external of the simulation domain Ω (that is, toward the internal of the body). The scheme is depicted in Figure 3.

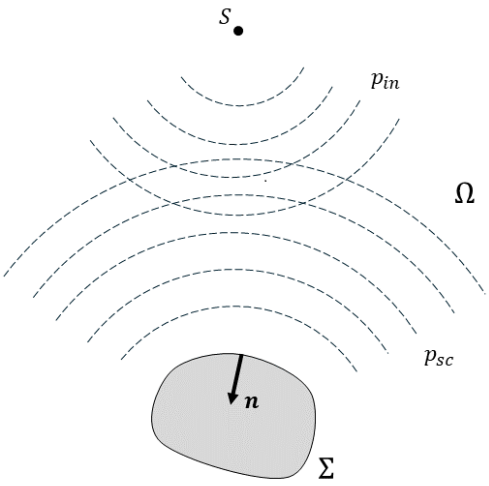


Figure 3. Mathematical definition of the problem: the sound wave p_{in} generated by a point source S is scattered by a surface Σ. The scattered wave is indicated with p_{sc}. The simulation domain is the boundless space Ω external to Σ. By convention, the normal vector $\mathbf{n}$ is directed outward from the simulation domain.

The pressure field $p(\mathbf{r}, t)$ generated at a point $\mathbf{r}$ at time t by a point source S placed at $\mathbf{r_S}$ can be subdivided into an incoming wave $p_{\text{in}}(\mathbf{r}, t)$ and a scattered wave $p_{\text{sc}}(\mathbf{r}, t)$ as

$$p(\mathbf{r}, t) = p_{\text{in}}(\mathbf{r}, t) + p_{\text{sc}}(\mathbf{r}, t). \tag{3}$$

The scattered wave is a solution of the free field wave equation

$$\frac{\partial^2 p_{\text{sc}}(\mathbf{r}, t)}{\partial^2 t} - c^2 \nabla^2 p_{\text{sc}}(\mathbf{r}, t) = 0, \tag{4}$$

where c is the speed of sound in air.

The incoming wave is given by a spherical wave propagating from the source position $\mathbf{r_S}$. For large values of $|\mathbf{r_S}|$, the spherical wave can be approximated with a plane wave. For a normalised monochromatic excitation at the angular frequency ω, we have, respectively,

$$p_{\text{in}}(\mathbf{r},t) = \frac{1}{4\pi|\mathbf{r}-\mathbf{r_S}|}e^{i(k|\mathbf{r}-\mathbf{r_S}|-\omega t)}, \qquad p_{\text{in}}(\mathbf{r},t) = e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}, \tag{5}$$

where $\mathbf{k}$ is the wave vector and $k = |\mathbf{k}| = \frac{\omega}{c}$ is the wave number.When Equation (4) is supplemented with boundary conditions at the surface Σ and with the Sommerfeld radiation condition at infinity, the scattering problem has a unique solution. For the sake of simplicity, Neumann boundary conditions, corresponding to the assumption of infinite impedance and, hence, hard surface, are frequently chosen. This approximation is valid in many cases, even though the acoustic impedance of human skin varies with frequency and across different individuals [39]. We assume Neumann boundary conditions in the remaining part of this section. An infinite impedance corresponds to a null normal derivative of the sound pressure field at points on Σ:

$$\frac{\partial p_{\text{in}}(\mathbf{r},t)}{\partial \mathbf{n}} = -\frac{\partial p_{\text{sc}}(\mathbf{r},t)}{\partial \mathbf{n}} \quad \mathbf{r} \in \Sigma, \tag{6}$$

where $\frac{\partial}{\partial \mathbf{n}}$ is the normal derivative to the surface Σ. The sound pressure field in Equation (6) is subdivided into the incoming and scattered waves, as in Equation (3). The Sommerfeld radiation condition corresponds to the requirement that the energy is radiated outward from the scatterer, without allowing inbound waves propagating from the infinity. In the time domain, the Sommerfeld condition is as follows:

$$\lim_{r\to\infty}\left[r\left(\frac{\partial p_{\text{sc}}(\mathbf{r},t)}{\partial t} + c\frac{\partial p_{\text{sc}}(\mathbf{r},t)}{\partial r}\right)\right] = 0, \tag{7}$$

where $r = \|\mathbf{r}\|$ is the radial coordinate. Together with conditions (6) and (7), Equation (4) gives a unique solution to the scattering problem.

The problem can be reformulated in the frequency domain via the Fourier transform. The Fourier transform of wave Equation (4) leads to the Helmholtz equation for the spatial part $u(\mathbf{r})$ of the scattered pressure field:

$$\nabla^2 u_{\text{sc}}(\mathbf{r}) + k^2 u_{\text{sc}}(\mathbf{r}) = 0, \tag{8}$$

where $k = \frac{\omega}{c}$ is the wave number.

The Neumann and the Sommerfeld radiation conditions of Equations (6) and (7) become, respectively,

$$\frac{\partial u_{\text{in}}(\mathbf{r})}{\partial \mathbf{n}} = -\frac{\partial u_{\text{sc}}(\mathbf{r})}{\partial \mathbf{n}}, \quad \mathbf{r} \in \Sigma, \tag{9}$$

$$\lim_{r\to\infty}\left[r\left(\frac{\partial u_{\text{sc}}(\mathbf{r})}{\partial r} - iku_{\text{sc}}(\mathbf{r},t)\right)\right] = 0. \tag{10}$$

The incoming wave for a source placed at a finite position is given, in the frequency domain, by the following:

$$u_{\text{in}}(\mathbf{r}) = G(|\mathbf{r}-\mathbf{r_S}|), \tag{11}$$

where

$$G(r) = \frac{1}{4\pi r}e^{ikr} \tag{12}$$

is the Green function for the wave equation, and $\mathbf{r_S}$ is the position of the source. For sources at infinity, the incoming wave becomes a plane wave of wave number k:

$$u_{\text{in}}(\mathbf{r}) = e^{ik\mathbf{r}\cdot\mathbf{r}}, \tag{13}$$

where **k** is the wave vector.

Once the scattering problem of Equations (8)–(10) has been solved in the frequency domain, the pressure field $p_R(\omega)$ at the receiver R is obtained by adding the solution of the scattering problem with the incoming wave of Equations (11) or (13) at R. The pressure field $p_O(\omega)$ at O is obtained by evaluating the incoming wave in O. These data enable us to compute the HRTF via Equation (1).

3. Measurement and Estimation

Binaural mannequins with standard dimensions can be employed for the measurement of HRTFs [40]. In the literature, several HRTF databases created by head and torso (HAT) simulators can be found [41]. These standard measurements help for the investigation of important aspects, such as the spatial resolution [42,43] and the listener distance from the loudspeaker [44,45]. However, standardisation does not catch individual differences between the HRTFs. The use of in-ear microphones placed in the ear canal of real subjects can solve this problem and allows to obtain individual (or personalised) HRTFs [18–21,46–48].

Two methods exist for HRTF measurement: deconvolution approaches and systems based on adaptive filters [49]. The most used methods are the ones based on deconvolution, and they can be categorised depending on the type of signal given as input, i.e., pseudo random sequences [47,50–58] or sweep signals [59–63]. Pseudo random sequences embrace maximum length sequences (MLSs) [51–55], inverse repeated sequences (IRSs) [56,57], and Golay codes [58]. The sweep signals can consider multiple exponential/logarithmic sweep [64,65] and multi-sine approaches [66]. In [67], a comparison of the most used deconvolution methods can be found, and in [68], the MLS method is compared with the sweep. Regarding the adaptive filtering approach, it is employed in HRTFs measurements in [69] for the first time, and then it is also applied in [45,70]. Usually, the normalised least mean square (NLMS) algorithm is applied for adapting the filter coefficients [71]. An accurate review of HRTFs measurement methods is reported in [49].

Various problems can occur and influence the measurement of HRTFs. Examples are the nonlinear distortions introduced by the acquisition system, noise produced by the environment, reflections, the type of sound source, and the temperature alteration [49,72]. The environment-based problems can be solved by using an anechoic chamber, and the nonlinear distortions can be prevented by applying the correct stimuli and technique [73–75]. Perfect periodic sequences (PPSs) and orthogonal periodic sequences (OPSs) are employed for measuring the HRTFs in [76] considering a real car environment, demonstrating robustness against non-linearities. PPSs can be applied for identifying Legendre non-linear (LN) filters [74,77] or Wiener non-linear (WN) filters [78,79], while the OPSs can determine functional link polynomial (FLiP) filters that include LN and WN filters [80–82].

The data acquisition process traditionally requires specialised equipment and controlled environments, which can limit its practical applicability. To address these challenges, in [83], the authors investigate the possibility of measuring the HRTF in a simplified experimental setup with a single sound source and a tracking system to acquire information about the head position and orientation in real-time. In this setup, the subject moves their head freely by sampling different directions while the source emits, in a non-anechoic environment, sequences of exponential frequency-modulated sweeps alternated with silences. Similarly, in [66], the authors propose a novel method for efficiently generating individualised HRTFs, in which a multisine signal is used to collect acoustic data as the listener moves freely, with a motion capture system tracking the head movements to ensure accurate positional data. This setup allows the acquisition of noisy, but plentiful, HRTF samples. Further details regarding this approach are provided in Section 6.

As an alternative approach to address the challenges associated with specialised equipment and controlled environments, the authors of [84] introduce a method for deriving personalised HRTFs using binaural recordings and head tracking data from consumer devices like earbuds with microphones and inertial measurement units (IMUs). By analysing how sound changes with head movement in various environments, the method estimates

personalised HRTFs. The results show that the estimated HRTFs closely match those measured in an anechoic chamber and significantly enhance sound localisation and reduce front–back confusion in virtual environments. Despite its effectiveness, there are limitations, including the need for stationary noise sources during recording, potential challenges with multiple or moving sound sources, and the impact of earbud placement on HRTF accuracy.

In [85], a method called UNIQ is presented for estimating a user's personal HRTF by leveraging the movement of a smartphone around the head, combined with mobile acoustic communication between the phone and earphones. Unlike traditional approaches that require controlled environments and precise equipment, UNIQ aims to provide a practical solution for estimating HRTFs at home, making it more accessible for applications such as immersive augmented reality (AR) and virtual reality (VR), earables, and smart hearing aids. The proposed system works by asking users to wear earphones and move a smartphone around their head while it emits pre-designed sounds. These sounds are recorded by the earphones, while the smartphone logs its inertial measurements during the motion. The algorithm then uses these inputs to estimate the user's personal HRTF. This method addresses two main challenges: first, accurately tracking the smartphone's location while accounting for head-related signal diffraction and multipath effects; and second, synthesising a far-field HRTF from near-field measurements, which differ due to variations in signal arrival angles. UNIQ tackles these challenges by modelling the 3D head geometry, applying diffraction effects, and deriving expected signal equations at the ears. It then compares these expectations against acoustic and IMU measurements to optimise the head parameters and phone locations. The near-field HRTF is generated and subsequently used to construct a far-field HRTF using physics-based modelling techniques.

As an alternative to direct measurement, estimation techniques rely on user feedback in response to specific auditory stimuli, allowing models to learn and adapt the HRTF based on subjective auditory perception. These techniques have become increasingly powerful tools in the field of HRTF personalisation. Researchers are leveraging various methods to predict and generate HRTFs without the need for direct physical measurements or detailed personal data. By relying on generalizable features or even user preferences, these techniques enable efficient and user-friendly HRTF customisation. Among these, it is particularly worth mentioning Yamamoto et al. (2017) [86], who proposed a method for personalising HRTFs using an adaptive variational autoencoder. The autoencoder is initially trained on a public HRTF dataset to learn general and individual-specific features. During calibration, users listen to pairs of audio signals and indicate their preferences. This feedback is used to tailor the HRTFs to the user's specific preferences without requiring detailed physical measurements. Despite the effectiveness and relatively fast optimisation of this method, it still requires about 30 min for calibration, and assessing the absolute quality of the final personalised HRTFs can be challenging. Gebru et al. (2021) [87] present a temporal convolutional network (TCN) that implicity models HRTFs. This method avoids the need for explicit measurements of the listener's physical features, instead learning the HRTF characteristics through training on perceptual data. The model's ability to generate high-quality HRTFs is evaluated through perceptual tests, showing promising results in terms of user satisfaction and sound localisation accuracy.

As a complement to HRTF measurements, some approaches require the acquisition of individual-specific data to further enrich HRTF datasets. These individual-specific data may range from simple inputs, such as facial or ear images and anthropometric measurements, to more complex and structured data, like 3D meshes. Indeed, a frequent preliminary problem concerns data acquisition for constructing a detailed model of an individual's head and torso. Such high-quality data are necessary when physical modelling HRTFs and for customisation and interpolation procedures. These 3D measurements are often provided in the form of 3D meshes, which consist of a set of vertices and polygons representing the head and upper torso of the subject. These measurements are typically obtained using 3D vision systems. By their nature, such systems have a limited field of view (FoV), meaning that only a restricted portion of the subject can be captured at any

given time. To overcome this limitation, two primary approaches are commonly employed. The first involves the use of multiple 3D vision devices positioned around the subject to capture different perspectives, followed by the fusion of the data from each device. Though offering the highest accuracy, this approach is also the most expensive due to the need for multiple devices. For example, the NoW dataset [88], which contains 3D head scans and high-resolution images of the subjects, belongs to this category. The second approach involves moving a single device around the subject to capture the necessary data. This method is more cost effective, as it eliminates the need for multiple devices, but it is also less accurate. One limitation is the potential for subject movement during the scanning process. Additionally, to achieve a complete 3D reconstruction, the movement of the sensor in space must be accurately estimated, which is necessary to align and merge the individual measurements. This is typically achieved through simultaneous localisation and mapping (SLAM) techniques followed by several refinement steps; these steps, however, tend to be less accurate compared to the multidevice approach. Datasets like HUTUBS [89], SONICOM [90], and 3D3A [91], which include 3D models and HRTF measurements, belong to this category. A detailed list of available HRTF datasets can be found in Appendix A.

4. Physical Modelling

We saw in Section 2 that an HRTF is a measure of a physical process, the scattering of a sound wave by a rigid body (the head and torso of an individual) in an anechoic, boundless environment. This process is completely governed by the wave equation with suitably chosen boundary conditions. This makes it possible to generate HRTFs through a numerical solution of a well-defined mathematical problem defined by Equations (4), (6) and (7) in the time domain, or Equations (8)–(10) in the frequency domain.

This section focuses primarily on the boundary element method (BEM), which is one of the most popular approaches to modelling scattering problems in acoustics, but other approaches are also considered.

4.1. Boundary Element Method

In this section, the boundary element method for the solution of scattering problems in acoustics is introduced. The aim of this part is to give a general overview of the method as a useful starting point for a more in-depth investigation. Furthermore, some key terms and concepts are introduced. For more details, we refer to [92–95], and the references therein. The appendix of the book [96] contains also a good starting point for understanding the theory of BEM.

The first step in BEM computation for an HRTF is to reformulate the problem described in Equations (8)–(10) as a boundary problem on Σ using Green's identity. The result is the Kirchhoff–Helmholtz integral equation for the scattered wave $u_{\mathrm{sc}}(\mathbf{r})$:

$$c(\mathbf{r})u_{\mathrm{sc}}(\mathbf{r}) = L[q] - M[u_{\mathrm{sc}}] \qquad q = \frac{\partial u_{\mathrm{sc}}}{\partial \mathbf{n}}, \tag{14}$$

where the functional operators L and M are defined as follows:

$$L[q] = \int_{\Sigma} q(\mathbf{r}')G(|\mathbf{r}' - \mathbf{r}|)d\Sigma(\mathbf{r}'), \quad M[u] = \int_{\Sigma} u(\mathbf{r}')\frac{\partial G(|\mathbf{r}' - \mathbf{r}|)}{\partial \mathbf{n}}d\Sigma(\mathbf{r}'), \tag{15}$$

where $\mathbf{r}$ is an arbitrary observation point in space, $\frac{\partial}{\partial \mathbf{n}}$ is the normal derivative to the surface Σ, and $G(r)$ is the Green function (12) for the Helmholtz equation. The normal to Σ is directed toward the external of Ω (the internal of the head) as established in Section 2.3. The factor $c(\mathbf{r})$ on the left-hand side of Equation (14) is one for $\mathbf{r}$ internal to Ω, $\frac{1}{2}$ for $\mathbf{r}$ on the boundary Σ and zero for $\mathbf{r}$ internal to the closed surface bounded by Σ. The integral Equation (14) gives a representation of the scattered pressure field in terms of its values on the boundary Σ. The numerical computation of an HRTF for different source positions requires repeated solutions of Equation (14). Hence, determining HRTFs in a sufficiently

dense set of directions may be an expensive computational task. Furthermore, only pressure field values at the receiver are retained. This computation can be made much more efficient using the reciprocal principle of acoustic: the pressure field at the receiver is not affected if the position S of the source and the position R of the receiver are swapped [97].

When the reciprocity principle is applied, the source is kept fixed at the entrance of the auditory channel and a single simulation enables obtaining the pressure fields at different positions of the receiver. The effectiveness of this approach has been experimentally validated using a rigid sphere and a Kemar dummy-head with a source located at the ear entrance and arrays of microphones in several positions in the laboratory [98]. However, this reversed approach requires that the source is placed on the boundary Σ, and this generates conceptual and computational difficulties that must be addressed. In [99], the authors show how the singular part of the scattered field can be separated using an image source internal to the region bounded by Σ.

Though computationally more efficient, reciprocity-based approaches produce a pressure field on the surface Σ that does not correspond to the propagation from a source placed in the true position S. Furthermore, the method based on the reciprocity principle is well suited for applications with point sources but cannot straightforwardly be applied with spatially complex sources. Hence, reciprocity-based approaches are not useful if the distribution of the pressure field on the head and torso is the objective of the simulation and in cases where the acoustic solicitation is complex.

BEM is a numerical method for solving the integral Equation (14) by means of a discretisation of the boundary Σ with a mesh (see Section 3 on the acquisition of meshes). In order to have high-quality simulations, the mesh should fulfil some requirements (see [99]):

- As a rough rule, for resolving a wavelength λ, the diameter d of elements of the mesh should be contained 6 to 10 times in λ. For instance, at a frequency of 6000 Hz and with a speed of sound of 343 m/s, the wavelength λ is approximately 5.7 cm, and the diameter d should not exceed a few millimetres. For a discussion of this requirement, see [100,101]. In order to resolve the geometry of the pinna, the elements should not exceed the length of 1 to 2 mm, depending on the local curvature of the surface.
- All the triangles are close to being equilateral. For planar quadrilateral meshes, the requirement is that the lengths of the edges of each element are as similar as possible.
- At each vertex, the number of concurring triangles is as close as possible to six.
- The rate of change of the area of the triangles across the mesh is gradual.

Once a high-quality mesh is available, the functional operators of Equation (15) in the Kirchhoff–Helmholtz integral Equation (14) can be discretised by associating discrete values of the fields to discrete elements of the mesh. A mesh provides a partition of the boundary Σ into N non-overlapping elements Σ_j, where j is a discrete index. The discretisation of the field u and its normal derivative q in Equation (14) are based on the choice of shape functions S_j associated with the elements Σ_j, which are typically piece-wise polynomials. A field ϕ on Σ, either u or its normal derivative q, is discretised as follows:

$$\phi(\mathbf{x}) = \sum_{j=1}^{N} S_j(\mathbf{x})\phi_j. \tag{16}$$

When a set of basis functions S_j has been chosen, there are several ways to obtain a discretised system of linear equations from the Kirchhoff–Helmholtz Equation (14). Here, we mention the following two methods:

- In the collocation schema, a set of collocation points x_j is chosen and Equation (14), with fields u and q expanded as in Equation (16), is required to hold at the collocation points.
- In the Galerkin schema, the problem is written in a weak form using a set of test functions, usually the same as the shape functions.

For the collocation schema, implementation is simpler and, for this reason, is frequently considered the first choice. When the observation point in Equation (14) is on the surface Σ, the coefficient $c(\mathbf{r})$ is equal to $\frac{1}{2}$, and the result of the discretisation process is one of the following linear problems, where the first equation refers to the collocation schema and the second one to the Galerkin schema:

$$\frac{1}{2}\mathbf{u}_{\mathrm{sc}} = G_{\mathrm{c}}q - H_{\mathrm{c}}\mathbf{u}_{\mathrm{sc}}, \qquad \frac{1}{2}B_{\mathrm{g}}\mathbf{u}_{\mathrm{sc}} = G_{\mathrm{g}}\mathbf{q} - H_{\mathrm{g}}\mathbf{u}_{\mathrm{sc}}, \tag{17}$$

where $\mathbf{u}_{\mathrm{sc}}$ is the N-dimensional vector of the pressure values associated with the N elements of the mesh, $\mathbf{q}$ is the corresponding vector associated with the normal derivatives of the pressure field, G_{c} and H_{c} are square matrices of dimension N corresponding to the operators L and M in Equation (15) in the collocation schema, G_{g} and H_{g} are the corresponding matrices for the Galerkin schema, and B_{g} is a real-valued, sparse mass matrix. The computation of the components of these matrices requires numerically integrating shape functions on the surface Σ. In both schemata, the integral operators in Equation (15) must be evaluated, by linearity, on the shape functions, which can be achieved via numerical methods. The linear problem can be solved by matrix inversion to obtain pressure field values on the surface of Σ:

$$\mathbf{u}_{\mathrm{c}} = \left(H_{\mathrm{c}} + \frac{1}{2}I\right)^{-1} G_{\mathrm{c}}\mathbf{q}, \qquad \mathbf{u}_{\mathrm{c}} = \left(H_{\mathrm{g}} + \frac{1}{2}B_{\mathrm{g}}\right)^{-1} G_{\mathrm{g}}\mathbf{q}. \tag{18}$$

where I is the identity matrix. The solutions of either the first (for collocation schema) or second (for the Galerkin schema) of Equations (18) can be used to derive the pressure field value at a generic position in space by numerically evaluating the integrals (15) in Equation (14).

Problem (17) is a computationally expensive and a memory-hungry task because the number of elements in the mesh ranges between tens of thousands to hundreds of thousands, and the matrices in (17) are fully populated and complex (except B_{g}, which is sparse). Furthermore, in collocation schemata, the matrices are not symmetric. The direct solution to the problem has complexity $O(N^3)$. The use of iterative methods can reduce the computational complexity from $O(N^3)$ to $O(N^2)$, but the cost of repeated iterations remains high. Originally developed in 1980s for dealing with many body problems governed by long-range interactions [102], the fast multipole method (FMM) can reduce computational complexity, at the cost of approximation errors. This means that computing HRTFs in a reasonable amount of time without the need for super-computational resources is feasible. Roughly speaking, the FMM is based on a multipole expansion of Green's function, which allows neighboring sources to be grouped together and treated as a single source. When applied to BEM, the method groups together elements of the mesh and partially substitutes element-to-element interactions with cluster-to-element and cluster-to-cluster interactions.

Methods where the FMM is applied to BEM are usually called fast multipole boundary methods (FMBEMs). A presentation of the method applied to the wave equation is given in [103], where a single-level application of the method is shown to reduce the computational complexity from $O(N^2)$ to $O(N^{\frac{3}{2}})$. This method is applied in [104], where a multilevel clustering of elements is introduced. In this paper, the method is tested on HRTF generation in a frequency range between 1000 Hz and 18,000 Hz. The use of a multilevel approach can reduce the computational complexity from $O(N^2)$ to $O(N\log(N))$ or $O(N)$, making it possible for the simulation of large-scale systems with reasonable computational resources. Kreuzer et al. [105] contains a short exposition of the multilevel FMM applied to BEM (see references therein for further details).

4.2. Implementations of BEM

The numerical implementation of BEM and other approaches in solving the wave equation is supported by several solvers. For instance, the Acoustic Module of COMSOL Multiphysics® [106], available starting at version 5.3a to the latest version 6.3, supports

BEM applications in acoustics. COMSOL is a proprietary software. Other proprietary tools supporting BEM for acoustics include FastBEMAcoustic® [107], suited for parallel computing and high-frequency simulations, and Ansys® [108].

Open-source software is also available, such as Mesh2HRTF [109], version 1.x, with its BEM core NumCalc. In [95], an overview of the numerical methods behind NumCalc is given. Many practical suggestions for implementing an effective simulation are shown throughout the paper. Furthermore, this paper contains examples of the accuracy of NumCalc in solving some benchmark problems; it is also useful for estimating the computational resources (CPU and memory) required in such simulations. Finally, the cited paper contains information concerning specific topics, e.g., the specific version of FMM and the implementation of frequency-dependent boundary conditions with frequency-dependent admittance. The references [101,110] contain information about the computational pipeline of Mesh2HRTF from pre- to post-processing stages through numerical simulation. Reference [111] contains an introduction to the theoretical background of Mesh2HRTF and a description of the software package.

Another open-source tool for solving boundary problems in homogeneous media, like the Kirchhoff–Helmholtz Equation (14), is BEM++ [112]. This software is a C++ library that can be used with Python scripts (latest version Python 3.13.0). The main citation of the Bempp library is provided by [113], while an introduction to solving boundary problems with BEM++ can be found in [94].

Many available solvers make use of a modified formulation of the problem (17) proposed by Burton and Miller [114] to prevent the possible degeneration of the boundary operators (15) at frequencies corresponding to resonances of the internal problem defined by Σ. These irregular frequencies may lead to the non-uniqueness in the solution and, even when the solution is unique, poor conditioning near these frequencies. The authors of [115] present a Helmholtz equation solver, making use of Burton and Miller's formulation to avoid problems arising from spurious resonances.

In the following, we discuss some applications of BEM to scattering problems in acoustics, and, more specifically, to HRTF computation.

In [105], FMBEM is used for HRTF generation over a wide frequency range. This work shows that, with FMBEM, it is possible to overcome the computational limitations at high frequencies due to the increased number of elements required to resolve short wavelengths. In [99], FMBEM is applied for the computation of HRTFs on different models (a rigid sphere, the head models Neumann KU-100 and Kemar and the torso and head model Kemar mannequin), all in the audible range up to 20 kHz. The computation is organised into two steps: the computation of the spherical harmonic representation of the HRTF Equations (26) and (27) and the subsequent HRTF evaluation. FMBEM is also used in [116] to numerically compute the HRTF of a dummy head in a wide range of frequencies between 20 Hz and 20 kHz. The aim of this study is to show how the FMBEM can be effectively used in this kind of problem, with emphasis on computational efficiency. In order to reduce the number of calls to the iterative solver, the authors make use of singular value decomposition (SVD) to express the excitation vectors related to the different positions of the source as linear combinations of a limited number of basis vectors.

In [117], BEM is used to generate the HRTF of the standard dummy head BHead210, which is designed with the typical anatomical characteristics of Chinese people, which may differ from those of Western people. The approach is first validated on a rigid sphere and then applied to BHead210. Simulated HRTFs are compared with measured ones and with the known ones of the Kemar dummy head. The authors report significant differences between Kemar and BHead210. HRTFs produced by FMBEM are also compared with the measured ones in [118]. The purpose is a cross-validation aimed at overcoming possible sources of error in computer simulations (e.g., accuracy in the generation of the mesh and choice of the boundary conditions) and laboratory measurements (e.g., inaccuracy of transducers and experimental errors in subject and device positionings). In this paper, the authors present both spectral and temporal analyses of the data collected

with measurements and simulations. The conclusion is that both methods have specific shortcomings, and that no single one is sufficient for providing a ground truth. The analysis identifies specific frequency bands where one method is more reliable than the other, making it possible to correct one with the other (cross-validation).

An optimisation in the implementation of a conventional BEM using a graphic processing unit (GPU) is presented in [119]. This study shows that the use of GPUs makes the fast generation of HRTFs by means of conventional BEM possible on a personal computer, without using the reciprocity priciple. Though more efficient, the reciprocity principle, as pointed out by the authors, may lead to inaccuracies in the final result. In addition, they investigate the perceptual effectiveness of the generated HRTFs. In [120], the authors propose a simulation framework for calculating room acoustic transfer functions in a low-frequency range, where traditional methods based on geometric acoustics do not prove to be effective. The task is required for the auralisation of virtual spaces. The framework includes the directionality of both the sound source and receiver. The sound field in the vicinity of the receiver and source is expanded into spherical harmonics (see Equations (26) and (27). The BEM is used to calculate the coefficients of the expansion of the fields in the vicinity of the receiver.

In order to exploit the best of both methods, a system that combines BEM with deep neural networks (DNNs) is proposed in [121]. Instead of training the network to directly predict HRTFs from data, they condition the network to predict differences between measures and simulations. The method has been proven to outperform BEM and SPCA-DNN approaches in all the audible range, as measured with spectral distortion. SPCA-DNN is a special framework introduced by the authors that uses neural networks in an approach based on a spatial principal component analysis (SPCA) factorisation of the HRTF.

In [122], Braren and Fels construct a database of HRTFs of children aged 5–10 years using BEM as a supporting tool. This knowledge is important in the auralisation of virtual environments appropriate for children, but data are scarce and, as shown in this paper, the available dataset and Kemar models for adults are significantly inadequate. BEM simulations are performed with the Mesh2HRTF [109] simulation tool and partially validated with COMSOL Multiphysics® [106]. The simulations are used to evaluate the quality of the dataset and the effect of head movement on measurements. In fact, it is more difficult to carry out controlled measures with children than with adults.

In [123], the authors propose a method for HRTF personalisation based on the reconstruction of ear meshes from high-quality images. The reconstructed meshes are integrated with a model of the human body to generate the personalised HRTF via BEM simulations run with COMSOL Multiphysics®.

4.3. Other Approaches

BEM-based approaches to solving the Helmholtz equation are especially attractive because only the values of the pressure field on the boundary Σ of the domain Ω are considered. This avoids the discretisation of large volumes, which results in cubic growth in the number of elements. Another appealing feature of BEM is that the Sommerfeld radiation condition (7) is already considered in the integral formulation (8) of the problem. Nevertheless, approaches with FEM, finite differences in the time domain (FDTD), and adaptive rectangular decomposition (ARD) are also considered. In this section, we mention some of these alternative approaches. We also mention some other approaches, for instance, ray tracing, that are only indirectly connected to the wave equation.

As stated above, the integral formulation in the frequency domain of the scattering problem in Equation (14) incorporates the Sommerfeld radiation condition (7), making an exact formulation of the problem for an unbounded domain possible. On the contrary, approaches in the time domain or based on FEM require a truncation of the unbounded domain with an external surface equipped with boundary conditions suited to absorb outgoing waves. Perfectly matched layers (PMLs) [124], where a fictitious absorbing material is placed at the external boundary of the computational domain, provide such conditions.

PMLs are applied in [125] to absorb outgoing waves at the truncated computational domain of the unbounded medium. The proposed system is based on FDTD to simulate the scattering of a sound wave on a simplified spherical model of a human head and a more realistic model up to a frequency of 7 kHz. The model is validated with the analytical solution for a spherical surface. It is also applied to compute an HRTF in the presence of a wall. In [126], an FDTD approach is used to compute HRTFs from head shapes to investigate the mechanism generating peaks and notches. This study shows that the pinna is responsible for generating such spectral structures. The method used in computer simulations is also explained in [127]. In [128], the effect of varying the complexity of the head and torso model is investigated by means of FDTD simulations. The objective is to assess the accuracy of the computed HRTFs with respect to the complexity of the geometry. In this paper, the simulations with a single sphere, a double sphere (the so-called snowman model), and a simplified model of a human head and torso are compared. The study shows that, for the single and double spheres, the simulations are accurate up to relatively high frequencies, whereas for the more complex model of a human head and torso, the same accuracy can be reached only at lower frequencies. In [129], the quality and reliability of FDTD-based approaches in HRTF generation are investigated. Limitations in boundary modelling are also faced in this work. The study is conducted both in the near and far fields.

An efficient algorithm for the solution of the wave equation in the time domain is adaptive rectangular decomposition (ARD) [130]. This algorithm is based on the well-known analytical solution of the wave equation in homogeneous media and in a rectangular decomposition of the simulation domain by means of large rectangular regions of empty space. The methods can be efficiently implemented on GPUs, achieving considerable performance. Furthermore, these approaches reduce the errors accumulated during computation. In [131], the authors propose an efficient computation framework for HRTF generation making use of ARD. Accurate 3D meshes of the head and torso are generated with the use of consumer-grade digital cameras. The meshes are subsequently used in an efficient computational pipeline for the numerical generation of HRTFs. In order to overcome the high computational cost associated with large volumes, they limit the simulation domain to a small cuboid enclosing the head and torso. In this way, they can gain an estimate of the pressure field on the scattering surface. This enables them to use the Kirchhoff–Helmholtz integral representation of the sound pressure field (14) to propagate the sound field in space.

In [132], a special form of the FEM tailored to the solution of the wave equation, the ultra-weak variational formulation (UWVF), is implemented. This method is especially efficient at high frequencies and it enables the authors to extend the computation over the entire frequency range. It is also applied to the computation of HRTFs of a standard head–torso mannequin. In [133], the authors apply and validate a FEM model for HRTF computation in the near field up to a frequency of 6 kHz. The computation domain is delimited with a triangular head mesh and a triangulated outer sphere. Volumes are meshed with tetrahedral elements. The requirement of 6 elements per wavelength leads to a number of about 1.4 million of elements. The main objective of the paper is to prove that FEM can be efficiently used in HRTF computation. Furthermore, the method allows for investigating features like diffractive region determination, which is possible with other methods.

In [134], the authors propose and validate two computational tools based on finite element methods for generating distance-dependent HRTFs. Experimental measurements of distance-dependent HRTFs are rare due to the complexity of experimental procedures. Therefore, computer simulations can help to fill this gap. The method proves to be computationally efficient and accurate over a frequency range of 0.2 to 20 kHz. Furthermore, the approach is robust against mesh quality reduction. The tool is versatile enough to incorporate bio-acoustic attributes such as the middle ear wall reflections. In [135], the authors face the problem of the high costs in computing HRTFs with FEM due to geometrical complexity, fine spatial discretisation, and densely sampled frequencies. They apply a model order reduction (MOR) based on the Krylov subspace [136] to a large-scale FEM, obtaining a considerable reduction in the computation time while preserving accuracy

below a frequency of 6813 Hz. The MOR reduces the complexity of large-scale models by constructing accurate surrogate models of lower complexity. Generally speaking, the MOR is based on detecting patterns in the solution and restricting the solution to these patterns.

Finally, some examples of geometrical approaches that are only indirectly associated with the solution of the wave equation are reported in the following. Acoustic ray tracing is based on the fact that, for wavelengths shorter than the size of the spatial inhomogeneities, the sound propagates in rays following geometrical laws; this is common in both acoustics and optics. This makes the use of computer graphics approaches in computing impulse responses at the location of a receiver when a pulse is sent from a source possible. In [137], the authors use acoustical ray tracing to simulate the HRIR of a standardised head described through a high-resolution mesh. Their approach is also able to include effects due to diffraction. The model is able to predict notches in the HRTF due to pinna reflections. In [138], the authors use ray tracing to correct an HRTF for taking into account torso reflections. HRTF computation usually considers the head and torso as a rigid body, because taking into account the relative motion of the two would be too costly. The authors use a simplified model of a head, neck and torso and optimize the parameters in such a way that the reflections lags computed with ray tracing fit those from dynamical HRTF measures. These time lags can be computed efficiently and are thus well suited for real-time applications.

To conclude this section, we mention some simplified models, in which the head is approximated with simple geometrical shapes like spheres and ellipsoids. The most extreme simplification consists of considering a pair of ears without a head. In this model, the HRTF is reduced to the pure, frequency-independent ITD and ILD generated by the propagation of the sound from the source. For this reason, the model is called a free-field HRTF model. This approximation provides partially accurate results at only large wavelengths with respect to the typical size of the head. The free-field model is applied, for instance, in [139,140]. Spherical heads are also considered in constructing simplified models. In fact, an HRTF can be exactly computed for a hard sphere (see [141,142] for an analytical derivation of the scattering of a plane wave from a rigid sphere). Exact solutions can be useful benchmarks for validating experimental procedures. Simplified head shapes also enable simple applications of ray tracing approaches to ITD computation. When applied to a spherical head with antipodal ears, these approaches lead to the well-known Woodworth formula. Extended versions of this formula, together with an explanation of the original model of Woodworth, can be found in [143], where the formula is adapted to describe situations where the ears are located in non-antipodal positions and the direction of the incoming wave is such that more than a single path contributes to the ITD. In [144], the Woodworth formula is also applied to elliptical heads. All these ray tracing-based methods foresee frequency-independent ITD and work properly at higher wavelengths, where diffractive contributions are less significant. Other approaches are based on the idea that simple geometrical shapes can model parts of an individual's body like the head, torso, and pinnae. In [145], for instance, regular spheres, oblate and prolate spheroids, and elliptical discs are used to model the head, torso, and ears. One example of these simplified shapes is given by the snowman model [146], where the head and torso are modelled as a couple of spheres or ellipsoids.

5. Anthropometric and Morphological Approaches for HRTF Personalisation

Alongside the physical modelling techniques, numerous data-driven methodologies for obtaining personalised HRTFs have been explored, spanning a spectrum of approaches that differ in both complexity and accuracy. These methods can be broadly defined as anthropocentric and morphological, with further subdivisions shown in Figure 4. Anthropometric approaches are based on the idea that the HRTF of an individual is influenced by somatic features of parts of the body like the size and shape of the torso, head, and pinnae. Additionally, the use of anthropometric data enables a less intrusive and more scalable approach, where individualised HRTFs can be derived.

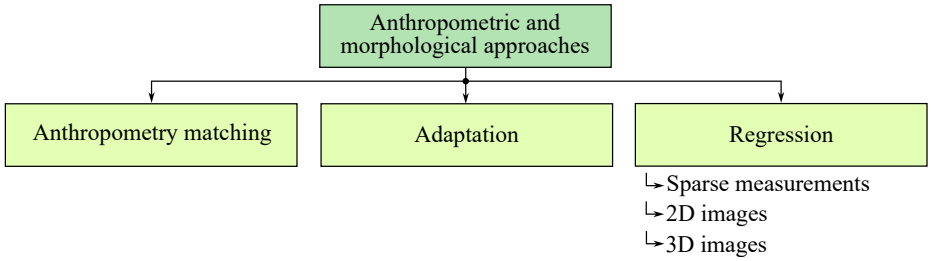


Figure 4. Overview of methodologies for obtaining personalised HRTFs using anthropometric and morphological approaches. The arrows under 'Regression' represent the types of input data used: sparse measurements, 2D images, and 3D images.

Many of the early studies on HRTF personalisation based on anthropometric features have relied on a sparse set of measures introduced in the CIPIC dataset [147]. This dataset includes measurements from 45 individuals, with a total of 37 anthropometric features recorded for each subject. Specifically, 17 of these features pertain to the upper body, while 10 additional measurements are dedicated to each ear, encompassing a comprehensive range of parameters relevant to auditory perception. Furthermore, the subsequent HUTUBS dataset [89], published later in 2019, utilises the same 37 measurements. Over the years, researchers have developed methodologies that leverage either all or a subset of these parameters to personalize HRTFs more effectively. These approaches vary in complexity, aiming to balance accuracy with computational efficiency. Table 1 provides a detailed overview of the key anthropometric parameters from the CIPIC dataset.

Somatic features can also be derived from a set of landmarks corresponding to acoustically significant points of the body extracted from 2D images, or even learnt automatically from a 3D representation using machine learning techniques. Each of these methods presents distinct advantages and challenges. In the following sections, we will examine these approaches in detail, exploring their specific use cases and potential for personalised auditory rendering. Anthropometric approaches to HRTF personalisation try to exploit these data to individualize an already available, standardised HRTF or to generate a new one starting from the available data. The main approaches include anthropometry matching, adaptation, and regression.

Table 1. List of anthropometric measures in the CIPIC dataset.

Symbol	Name of the Parameter	Symbol	Name of the Parameter
x_1	Head Width	x_{15}	Seated Height
x_2	Head Height	x_{16}	Head Circumference
x_3	Head Depth	x_{17}	Shoulder Circumference
x_4	Pinna Offset Down	d_1	Cavum Concha Height
x_5	Pinna Offset Back	d_2	Cymba Concha Height
x_6	Neck Width	d_3	Cavum Concha Width
x_7	Neck Height	d_4	Fossa Height
x_8	Neck Depth	d_5	Pinna Height
x_9	Torso Top Width	d_6	Pinna Width
x_{10}	Torso Top Width	d_7	Intertragal Incisure Width
x_{11}	Torso Top Depth	d_8	Cavum Conca Depth
x_{12}	Shoulder Width	θ_1	Pinna Rotation Angle
x_{13}	Head Offset Forward	θ_2	Pinna Are Angle
x_{14}	Height		

5.1. Anthropometry Matching

Anthropometry matching relies on a dataset where the available HRTFs are associated with corresponding anthropometric measures. When a new individual is considered, the method selects the HRTF in the dataset, minimising the suitably defined distance between the anthropometric measures of the subject and those in the dataset. In 2010,

Schönstein et al. [148] addressed the challenge of selecting individualised HRTFs for virtual audio environments using easily obtainable data, like photos of a listener's ear. By analysing a database of 46 measured HRTFs, the study identifies key morphological parameters through data reduction and data mining techniques. Using these parameters, it predicts a listener's preferred HRTF via linear regression, which is validated through listening tests.

In 2015, Torres-Gallegos et al. [149] introduced a novel approach utilising active shape models (ASMs) to automatically detect and extract key anthropometric features from 2D images, making it a more efficient solution compared to previous methods. The ASM is trained on a set of annotated images and is capable of identifying relevant anatomical features such as details of the ears, head, and torso from photographs of new subjects. The system begins by using computer vision to identify key anthropometric points on the subject's face and body. Once these points are located, the distances between them are measured and converted into metric units, generating a set of anthropometric parameters that describe the subject's unique anatomy. This information is then used to find the best HRTF match in the CIPIC HRTF database.

In 2022, Warnecke et al. [150] proposed a matching model based on a dataset of 3D ear meshes associated with landmark annotations and HRTFs. The authors used the Euclidean distance between the landmarks to create a ranking of distances from a selected individual and all the others in the dataset. Furthermore, they introduced different metrics to compare and rank the HRTFs of the selected user with all the others in the dataset: two objective measures, the spectral difference error (SDE) and the sagittal-plane localisation model (see Section 2.2), and a subjective measure based on the performances of real individuals in virtual reality sound detection task. The hypothesis is that the ear with the shortest distance from the ear of the selected individual corresponds also to the shortest distance between the corresponding HRTFs.

In the same year (2022), Zhi et al. [151] proposed a streamlined method for personalising HRTFs using a convolutional neural network (CNN) that predicts ear landmarks from a single photo. Their approach requires minimal inputs, photos of both ears and head dimensions, and involves several steps: identifying ear landmarks, calculating anthropometric distances, transforming ear shapes into a standard format, and adjusting HRTFs based on head and ear scales. To train the CNN, the authors create a dataset of 741 hand-labelled ear photos, expanding and augmenting these to produce a total of 2964 images, including synthetic ones generated from computer-rendered ear meshes. They employ a nearest-neighbour search to match input ears based on four computed distances. Once a matching database ear is found, the corresponding HRTF is scaled to account for differences in ear and head size by resampling based on the relative ear lengths and adjusting the ITD. A head-and-torso (HAT) model is used to further fine-tune the personalised HRTFs, improving spatial accuracy.

Most matching techniques rely on traditional methods, while the adaptation and regression approaches described in the following sections are based on machine learning. Results consistently show that these ML methods outperform conventional techniques. However, ML models are generally data hungry, with their effectiveness highly dependent on access to large, high-quality datasets, while generating personalised HRTFs with the techniques introduced in Section 4 can be relatively straightforward, noted minor inconsistencies between generated and measured HRTFs have been noted, underscoring the importance of comprehensive datasets like SONICOM [90] in ensuring model reliability and accuracy. Works like [152] emphasize how ample training data are crucial for these models, as increased data consistently enhance model performance.

5.2. Adaptation

The adaptation methods start with a non-individualised HRTF and try to reshape it based on the specific anthropometry of the subject. One of the first works in this domain is that of Zotkin et al. [153]. The authors propose an approach that utilises specific anthropometric ear parameters (specifically, cavum concha height, cymba concha height, cavum concha width, fossa height, pinna height, pinna width, and intertragal incisure width; see Table 2), combined with a low-frequency head-and-torso model, to match and adapt HRTFs from the CIPIC database [147]. Their results indicate that this method improves both the localisation accuracy and subjective perception of virtual auditory scenes.

Another approach, proposed by [154], is a hybrid matching-adaptation technique. The authors proposes a method for customising HRTFs using computer vision to address the limitations of traditional direct measurement-based techniques, particularly at low frequencies. Traditional HRTF measurements often struggle with obtaining reliable low-frequency data due to the long signal lengths required, which can be influenced by room reflections and measurement equipment. To mitigate this issue, the paper integrates an analytical head-and-torso model for synthesising low-frequency HRTFs. These synthesised HRTFs are blended with measured HRTFs from the Kemar dataset [41] by performing a matching process based on anthropometric features derived from a computer vision analysis of the user's head, torso, and external ears.

In 2017, Zhu et al. [155] developed an efficient method for HRTF personalisation, utilising a weighted sparse representation of anthropometric features. These features are weighted according to their importance in accurately modelling HRTFs, with greater emphasis placed on those features that have a stronger influence on acoustic characteristics. The authors assume that both HRTFs and anthropometric data could be represented by the same sparse vector, following previous sparse representation techniques. The normalised sparse vector is applied to the log-scale HRTF data from the CIPIC database to synthesize personalised HRTFs.

5.3. Regression

Regression methods aim to generate an HRTF directly from anthropometric measures using a trained regression model. Given that anthropometric features can be represented as sparse measurements, two-dimensional images, three-dimensional meshes, or a combination of these, it is worthwhile categorising existing studies in the literature based on the type of input data used.

5.3.1. Sparse Measurements

When speaking of anthropometric features in terms of sparse measurements, one of the first works is Grindlay et al.'s 2007 study [156]. They employ a regression model to map anthropometric features to the "people" vector space generated by N-mode SVD. Subsequently, when anatomical measurements from a subject not present in the database are provided, this regression model is used to determine that individual's position in the people space. This representation is then combined with the N-mode SVD results to produce a customised set of HRTFs.

In 2008, Hu et al. [157] introduced a method for customising HRTFs using a neural network approach based on specific anatomical measurements. The model maps selected anthropometric parameters to estimate individualised HRTFs without requiring complex setups. Selected parameters for the model include head width, head depth, shoulder width, cavum concha height and width, fossa height, pinna height, and pinna width. A three-layer feed-forward neural network is used to capture the non-linear relationship between these parameters and key HRTF characteristics like principal components and ITD. Results indicate that the localisation accuracy of the personalised HRTFs is increased compared to non-individualised HRTFs.

In 2013, to develop a personalised model, Li et al. [158] applied High-Order Singular-Value Decomposition (HOSVD) to extract an individual core tensor representing an HRTF,

which serves as the output. Key anthropometric parameters are selected using the Laplacian score and correlation analysis, which act as inputs for a Radial Basis Function (RBF) neural network. The RBF-based model is then employed to predict individual HRTFs using these selected anthropometric features.

In 2014, Bilinski et al. [159] proposed a method for synthesising HRTF magnitudes using a sparse representation of anthropometric features. The proposed method tackles the problem of HRTF synthesis by determining a sparse representation of a subject's anthropometric data relative to the original ones present in the training set. By learning a sparse vector that expresses a subject's features as a linear combination of the training data, this method then utilises the same representation to generate the corresponding HRTF magnitudes. This approach has been benchmarked against ridge regression, a generic mannequin model, and various classifiers, showing enhanced localisation accuracy compared to non-individualised HRTFs.

Later, in the era of deep learning, Chun et al. [160] introduced a deep neural network (DNN) approach for personalising HRTFs using all measurements from the CIPIC dataset. The model outputs a personalised HRIR and demonstrates improved accuracy over averaged HRTFs, validated through root mean square error (RMSE) and log-spectral distortion (LSD) metrics.

In 2019, Chen et al. [161] proposed a model based on an autoencoder and a multilayer perceptron (MLP). First, the autoencoder is trained on HRTF data to extract a bottleneck vector representing key features. This bottleneck vector is then used, like in a distillation task, as a target to train the MLP, with anthropometric features as input. During the test phase, the MLP generates the bottleneck vector for a new subject's anthropometric features, and the decoder reconstructs the corresponding HRTFs. The model mitigates the overfitting observed in direct MLP-based estimation by using the autoencoder to reduce the dimensionality of the data and augmenting the training set with additional HRTF samples. The model leverages 27 anthropometric features derived from the CIPIC database, focusing on left ear measurements while excluding data related to the right pinna. Although the CIPIC database includes ITD, HRIRs, and anthropometry data, only magnitude HRTFs are considered in this work. In 2020, Miccini and Spagnoli expanded upon and improved these ideas in a similar investigation [162].

Still in 2020, ref. [163] introduced a method for global HRTF personalisation using spherical harmonics (SHs). The approach extracts low-dimensional SH features from HRTFs across different frequencies, incorporating perceptual auditory characteristics like critical bands and loudness sensitivity. A deep learning model maps anthropometric measurements, such as head, torso, and ear dimensions, to SH coefficients. By using the HUTUBS HRTF database for training, this method predicts personalised HRTFs for all spatial directions through a single model, improving upon previous techniques that used separate models for each elevation.

To summarize this section, Table 2 encapsulates all the parameters employed in the approaches discussed thus far, aiming to assist the reader in assessing which measurements from the CIPIC dataset have been considered effective by researchers in the personalised HRTF generation task.

Table 2. List of approaches using CIPIC anthropometric measures and the selection of the used parameters. Starting from the deep learning era, the trend has been to utilize all 27 measurements available. Please refer to Table 1 for a detailed description of the parameters.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	θ_1	θ_2
Schönstein et al. [148]	✓	✓	✓					✓			✓	✓					✓				✓	✓					
Torres-Gallegos et al. [149]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓						✓	✓	✓	✓	✓	✓	✓			
Zhi et al. [151]																		✓	✓	✓	✓						
Zotkin et al. [153]	✓						✓					✓						✓	✓	✓	✓	✓	✓	✓			
Mohan et al. [154]																		✓	✓	✓	✓	✓	✓	✓		✓	
Zhu et al. [155]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Grindlay et al. [156]	✓																	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hu et al. [157]	✓		✓			✓		✓	✓	✓		✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		
Li et al. [158]		✓												✓	✓	✓	✓	✓		✓	✓		✓	✓	✓		✓
Bilinski et al. [159]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Chun et al. [160]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Chen et al. [161]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wang et al. [163]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

5.3.2. Two-Dimensional Images

In the category of two-dimensional images, machine learning models focus on generating HRTFs using two-dimensional data, such as images or anthropometric measurements derived from pictures. These techniques leverage 2D representations to infer spatial auditory cues, providing a less complex but effective means of HRTF generation. One notable study is that of Spagnol [164]. In their approach, HRTFs from the CIPIC [147] dataset are processed to extract the pinna reflection transfer function (PRTF). This is derived by isolating the spectral features related to pinna reflections from the measured HRTFs using a series of filtering techniques. Specifically, the model focuses on the central frequencies of spectral notches in the PRTF, which are linked to different reflection paths on the pinna surface. To identify these reflection paths, the model also incorporates geometric data extracted from 2D images of the subject's pinna, including manually traced contours representing distinct anatomical features such as the helix border, concha outer border, and antihelix. These contours are then used to map the observed notch frequencies to specific reflection surfaces on the pinna, allowing the model to predict the effects of its anatomy on HRTF variations. Years later, in 2023, Pirard [165] discussed the use of convolutional neural networks to predict HRTFs from images of a listener's ear. The method extracts anthropometric features from ear images to find a matching set of HRTFs. The process involves training a CNN to identify ear landmarks, computing key distances, and matching them with a database to retrieve personalised HRTFs. The approach aims to simplify HRTF individualisation with minimal user input, though further research is needed to optimise and refine the process.

5.3.3. Three-Dimensional Meshes

More recent methods approach the task from a 3D perspective. These approaches offer a higher degree of personalisation by capturing fine-grained spatial features, resulting in more accurate and individualised HRTFs. In 2021, Fantini et al. [166] proposed a regression method based on two steps: the extraction of anthropometric measures from a 3D mesh of the head followed by HRTF individualisation. The anthropometric parameters are automatically extracted from the mesh using the active shape model (ASM) algorithm trained on an available dataset. These parameters are employed as input for generalised regression neural networks (GRNNs) trained to predict the HRTF. The GRNNs, one for each value of the elevation angle of the source in the median plane, are trained using the spectral distortion (SD) to measure the distance between the current HRTF and the estimated one. Though it is widely used in HRTF personalisation studies, spectral distortion poorly correlates with subjective perception performance results. Thus, the authors train a second set of GRNNs using the sagittal-plane localisation model (see Section 2.2) as a metric.

As another notable study, ref. [167] investigates the use of deep neural networks to predict individualised HRTFs from 3D ear shapes. Utilising data from 645 human subjects, 3D meshes of their heads and upper torsos are scanned and used to simulate ear HRTFs with the finite-difference time-domain (FDTD) method. The study evaluates two models: CNN-Reg, which employs a series of convolutional layers, and a 3D UNet model (UNet-Reg), which also predicts HRTFs from 3D inputs, both of which are trained separately for each frequency bin. Even though the two models achieve good results in minimising the spectral distance errors in the predicted HRTF magnitude spectra, the approach of training a separate model for each frequency bin presents limitations, making it less suitable for complex, real-time applications, where continuous frequency prediction is required. Wang et al. [168] investigate the application of convolutional neural networks (CNNs) in predicting individualised HRTFs using detailed head scan data. In this approach, the input to the CNN consists of truncated spherical cap harmonic (SCH) coefficients, which efficiently represent the pinna geometry, an essential feature for accurate acoustic scattering, along with additional anthropometric measurements of the head and torso. The HRTFs are predicted as spherical harmonic (SH) coefficients, serving as the output of the model. Experimental results demonstrate that this method achieves a significantly

lower log-spectral distortion (LSD) compared to the BEM when evaluated on the HUTUBS HRTF database [89].

In 2023, Ko et al. [169] introduced PRTFNet, a CNN model designed to reconstruct accurate spectral cues in HRTFs by reducing the influence of the head and torso. It uses a compact pinna-related transfer function and includes HRTF phase personalisation, adjusting phases based on the listener's head width. Validated with the HUTUBS HRTF dataset [89], the model transforms HRIRs into compact PRTFs to preserve spectral cues. In 2024, Zhao et al. [170] utilised 3D mesh features to predict HRTF spectra across vertical planes, ensuring spectral continuity and correlation across adjacent grids and frequencies. The evaluation results show that the model outperforms state-of-the-art methods in terms of LSD and enhances the prominence of peaks and notches in HRTF spectra.

6. Interpolation

Interpolation techniques are essential for enhancing the smoothness and continuity of spatial audio experiences, especially when dealing with incomplete or sparse HRTF datasets. Various papers discuss the development of interpolation algorithms to improve HRTF accuracy and performance. The interpolation methods in the literature can be classified into three main categories, i.e., nearest-neighbour approaches, techniques based on functional models, and neural network-based interpolation. An overview of these categorisations is shown in Figure 5 and Table 3.

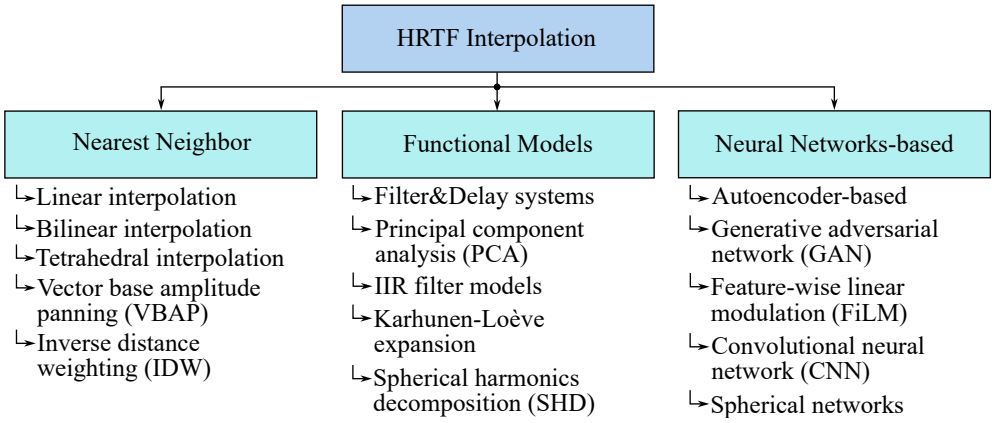


Figure 5. Categorisation of the methods for HRTF interpolation presented in this paper.

Table 3. Overview of the HRTF interpolation methods.

Nearest Neighbour	Functional Models	Neural Networks-Based
Linear interpolation [27,33,171–173]	Filter&Delay systems [174–180]	Autonecoder-based [181,182]
Bilinear interpolation [25,26,183,184]	Principal component analysis (PCA) [185–187]	Generative adversarial network (GAN) [188]
Tetrahedral interpolation [23,24]	IIR filter model [189,190]	Feature-wise linear modulation (FiLM) [37]
Vector base amplitude panning (VBAP) [34,191]	Karhunen–Loève expansion [192]	Convolutional neural network (CNN) [66,193–195]
Inverse-distance weighting (IDW) [28,196]	Spherical harmonics decomposition (SHD) [29,30,197–205]	Spherical network [206,207]

6.1. Nearest Neighbour Interpolation

The nearest neighbour interpolation techniques are based on calculating the interpolated HRIR $\hat{h}(n)$ as the combination of the nearest measured impulse responses as follows:

$$\hat{h}(n) = \sum_{i=1}^{N} w_i h_i(n), \tag{19}$$

where N is the number of the measured impulse responses h_i considered in the interpolation, and w_i are the weights that can be estimated in several ways.

The most simple interpolation technique is linear interpolation [171], in which two impulse responses are involved ($N = 2$), and the weights are computed as

$$w_1 = \frac{x_2 - \hat{x}}{x_2 - x_1}, \quad \text{and} \quad w_2 = \frac{\hat{x} - x_1}{x_2 - x_1}, \tag{20}$$

where $\hat{x}$ is the position of the interpolated point, and x_1 and x_2 are the positions of the two measured impulse responses. In the HRTF dataset, where the measurements depend on the azimuth and the elevation, x resembles the azimuth or the elevation angles, and HRIRs should have the same elevation or the same azimuth, respectively. In 1993, Wenzel and Foster employed linear interpolation for a perceptual study conducted on individual HRTFs [171]. In 2018, Garcia-Gomez used linear interpolation as part of the entire algorithm applied to binaural room responses, considering a horizontal measurement distribution [33]. This approach was also used by Bruschi et al. in 2020, who considered a spherical distribution [172]. Finally, in 2023, Bruschi et al. applied linear interpolation in combination with a frequency-warping approach to interpolate HRTF for non-linear 3D audio systems [27]. Frequency-warping is applied to reduce the resolution at high frequencies, improving binaural reproduction performance.

One widespread technique is bilinear interpolation, as presented by Savioja and Lokki in 1999 [183]. In this case, $N = 4$ impulse responses are involved, and the four weights are calculated as follows:

$$w_1 = (1 - c_\theta)(1 - c_\phi), \quad w_2 = c_\theta(1 - c_\phi), \quad w_3 = c_\theta c_\phi, \quad \text{and} \quad w_4 = (1 - c_\theta)c_\phi, \tag{21}$$

where the coefficients c_θ and c_ϕ are computed as

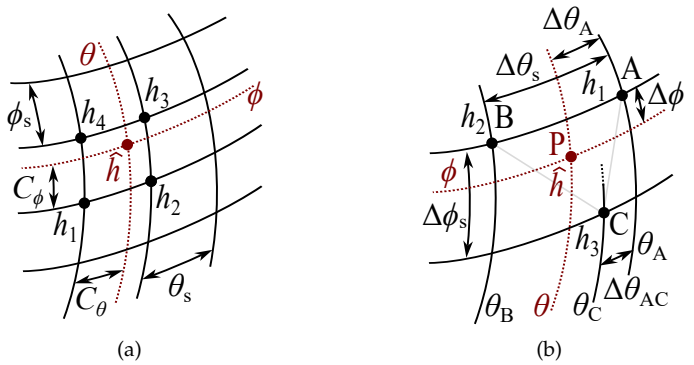
$$c_\theta = \frac{C_\theta}{\theta_s} = \frac{\theta \text{mod} \theta_s}{\theta_s}, \quad \text{and} \quad c_\phi = \frac{C_\phi}{\phi_s} = \frac{\phi \text{mod} \phi_s}{\phi_s}, \tag{22}$$

where θ_s and ϕ_s are the angular steps, and C_θ and C_ϕ are the relative positions for azimuth and elevation, respectively, as shown in Figure 6a. The angular steps θ_s and ϕ_s must be fixed, so the positions of the dataset should be homogeneously distributed on a grid on the sphere surface.

In 2002, Biscainho et al. improved the bilinear technique by adding a new auxiliary inter-positional transfer function (IPTF) and reducing the number of points needed to three, resulting in a triangular interpolation in which the weights are calculated as follows [25,184]

$$w_1 = 1 - w_2 - w_3, \quad w_2 = \frac{\Delta\theta_A - w_3 \Delta\theta_{AC}}{\Delta\theta_s}, \quad \text{and} \quad w_3 = \frac{\Delta\phi}{\Delta\phi_s}, \tag{23}$$

where the angular distances are determined as $\Delta\phi = \phi - \phi_A$, $\Delta\phi_s = \phi_C - \phi_A$, $\Delta\theta_s = \theta_B - \theta_A$, $\Delta\theta_A = \theta - \theta_A$, and $\Delta\theta_{AC} = \theta_C - \theta_A$ as shown in Figure 6b. In this case, the measurement points can be more scattered, and fixed angular steps are unnecessary, accepting more freedom in the HRTF measurement positions. In 2004, Freeland et al. proposed a hybrid interpolation method that combines the bilinear interpolation of [25,184] with spatial feature extraction [26]. This method presents a lower computational complexity and performs similarly to direct bilinear interpolation.



(a) (b)

Figure 6. Setup of (**a**) the bilinear interpolation of [183] and (**b**) the triangular interpolation of [25].

Another technique that can be used for calculating weights is vector base amplitude panning (VBAP), which was introduced by Pulkki in 1997 [191] for loudspeaker reproduction and used by Queiroz et al. in 2011 [34] for HRTF interpolation. Three-dimensional VBAP involves $N = 3$ measured impulse responses, and the weights are obtained as follows:

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \mathbf{L}^{-1}\mathbf{p} = \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \tag{24}$$

where $\mathbf{L}$ is the matrix containing the direction of the three measured points in the Cartesian coordinates, and $\mathbf{p}$ is the vector that describes the position of the virtual sound source, i.e., the interpolated point.

In 1999, Hartung et al. [28] applied inverse-distance weighting (IDW) for HRTF interpolation. IDW is a spatial interpolation that can consider different numbers of measured HRTFs N according to the number of nearby points. This method is also applied by Bruschi et al. in [196], and the equation for the weights computation is as follows:

$$w_i = \frac{d_i^{-q}}{\sum_{i=1}^{N} d_i^{-q}}, \tag{25}$$

where $d_i = \cos^{-1}[\sin\theta_i \sin\theta + \cos\theta_i \cos\theta \cos(\phi_i - \phi)]$ is the distance of the ith point (θ_i, ϕ_i), and q is the distance-decay parameter. In [28], $N = 4$ and $q = 1$ are imposed, while in [196], $N = 2$ and different values of q are considered. All the methods described above can be implemented in both the time and frequency domains by applying the procedures depicted in Figure 7. When the interpolation is performed in the time domain, the impulse responses must be aligned to guarantee the effectiveness of the interpolation as shown in the scheme of Figure 7a. This requirement is assured when minimum-phase (MP) impulse responses are employed [183]. Figure 8 compares the HRTF interpolated with the described methods and the measured HRTF, when non-minimum-phase responses are involved. The IDW is applied by imposing $N = 8$ and $q = 0.5$. In this case, bilinear interpolation better fits the measured frequency response. Figure 9 shows the results obtained with minimum-phase responses. The great improvement in all the algorithms that can correctly represent the measured HRTF is evident. The SD values obtained using Equation (2) and the methods in Figures 8 and 9 are calculated in the range [20 Hz; 30 kHz]; these are shown in Table 4. As verified by the frequency response plots, bilinear interpolation gives the best SD values; however, in general, the SD decreases when minimum-phase responses are involved.

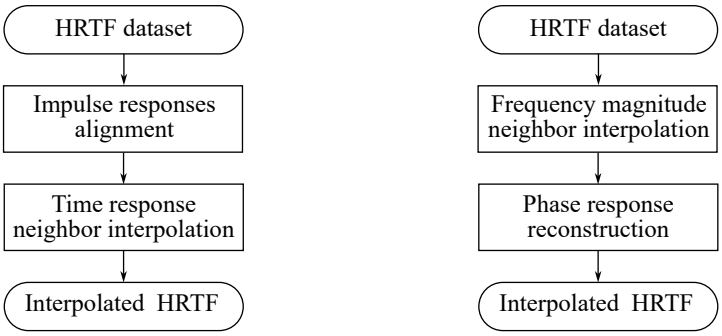


Figure 7. Scheme of the nearest neighbour interpolation method applied (**a**) in the time and (**b**) frequency domains.

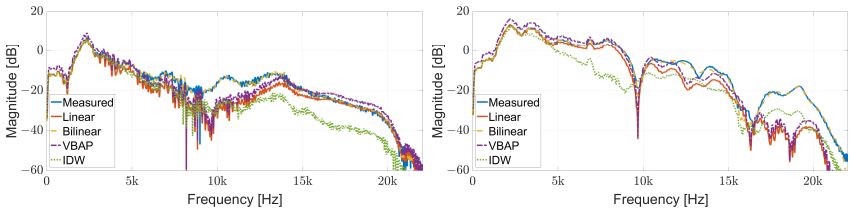


Figure 8. Comparison of HRTF interpolation methods with non-minimum-phase HRTF, considering (**a**) the left ear and the (**b**) right ear for the position $(\theta = 45°, \phi = 10°)$ using the MIT-KEMAR database of [41].

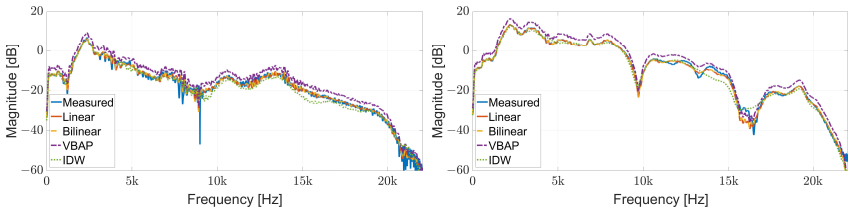


Figure 9. Comparison of HRTF interpolation methods with minimum-phase HRTF, considering (**a**) the left ear and the (**b**) right ear for the position $(\theta = 45°, \phi = 10°)$ using the MIT-KEMAR database of [41].

Table 4. Comparison of the SD values obtained with the methods of Figure 8, for non-minimum-phase responses and Figure 9, for minimum-phase ones. The lowest values are highlighted in bold text.

	Non-Minimum-Phase		**Minimum-Phase**	
	Left	**Right**	**Left**	**Right**
Linear	6.6 dB	9.2 dB	**1.5 dB**	**1.4 dB**
Bilinear	**1.5 dB**	**1.4 dB**	**1.5 dB**	**1.4 dB**
VBAP	5.1 dB	7.7 dB	3.4 dB	3.0 dB
IDW	9.2 dB	9.1 dB	2.3 dB	2.2 dB

When information related to HRIR delays or phases is important, such as when different distances are considered, HRIRs cannot be considered minimum phase. This also occurs when the measurements are performed in real rooms, so the effect of the room response is included in the HRIR, resulting in what is called binaural room responses (BRIRs) [33]. In these cases, every interpolation technique applied in the time domain

requires pre-processing to align the impulse responses. In this context, in 2018, Garcia-Gomez and Lopez [33] proposed a new interpolation method applied to BRIRs that divides the impulse responses into early reflections and reverberant tail. The idea of splitting the impulse responses was already proposed by Kearney et al. in [208] for room impulse responses. In that case, the dynamic time warping (DTW) algorithm is applied for the temporal alignment, and the reverberant part is synthesised following the approach of [209]. In a different way, Garcia-Gomez and Lopez [33] apply the time-splitting to binaural room impulse responses, where the early reflections part is divided into two frequency bands. The low-frequency portions are linearly interpolated, while peak detection and alignment algorithms are reserved for high-frequency parts, which contain most of the information. This method presents a lower computational cost than the DTW algorithm of [208]. In 2020, Bruschi et al. [172] improved the approach of Garcia-Gomez by changing the peak detection algorithm and evaluating the performance on semi-anechoic measurements. Recently, Bau et al. proposed a spherical array interpolation by time alignment (SARITA) for upsampling the BRIR dataset [173]. Similarly, signal time alignment is performed by separating early reflections and the direct sound into overlapping blocks, and a linear interpolation is then applied [173].

When nearest neighbour interpolation is applied in the frequency domain, Equation (19) can be directly applied to the complex HRTF; however, this method does not always guarantee good performance, depending on the employed HRTFs. For this reason, some approaches suggest applying the nearest neighbour interpolation to the frequency magnitude response before reconstructing the phase response, as shown in the scheme of Figure 7b. In [28], Hartung et al. performed a comparison between IDW and spherical spline interpolation [210], both implemented in time and frequency domains. The results have shown the better performance of the frequency domain approaches, particularly with the spherical spline technique, which involves not only the nearby positions but the entire dataset, thus requiring a higher computational cost. Another frequency domain technique was proposed by Gamper in 2013 [24]. In this case, the interpolation is performed in azimuth, elevation, and also distance, and an interpolated HRTF is obtained by applying proper gains to the nearby HRTFs of the database, forming a tetrahedron. Interpolation is only executed only on the magnitude frequency response, and a spherical head model is applied to derive the phase response. In 2017, Hugeng et al. compared tetrahedral interpolation with triangular and rectangular bilinear interpolation [23]. The results show better performance with minimum-phase responses and with tetrahedral interpolation. However, the latter requires a 3D database that varies with distance and has a higher computational complexity.

6.2. Functional Models for HRTF Interpolation

The first attempt at HRTF interpolation was functional representation, i.e., the calculation of a mathematical model that depends on frequency and direction, derived from a set of HRTF measurements. The first functional approach was proposed in 1967 by Batteau, who modelled the external ear as a three-channel, two-delay system [174]. The same model was then used in [175,176] to investigate localisation cues. In 1986, Genuit proposed a model based on a 16-channel filterbank, in which the filter parameters are related to the shape of the external ear [177]. Based on Genuit's idea [177], Brown and Duda developed a model that applies proper filters and delays to reproduce the effect of the head, shoulders, and pinna [179,180]. The model proposed in [179] consists of different modules obtained by combining properly parameterised delays and filters; each component of the model affects at least one of the three spatial dimensions, i.e., azimuth θ, elevation ϕ, and distance r. Based on this idea, in 2007, Freeland et al. [211] proposed a structure dependent on θ and ϕ by generalising the continuous variable digital delay (CVDD) method introduced in [212]. The system performs polynomial interpolation, optimising the coefficients inside a given region of the sphere [211]. In 1992, Chen et al. modelled the external ear as a multi-

sensor broadband beamformer [178]. This approach is limited by the high computational complexity and numerical instability.

Aiming to reduce the computational complexity of the model, Martens introduced principal components analysis (PCA) in 1987 [186]. Kistler and Wightman extended this method in 1992 demonstrating the efficiency of representing a low-order HRTF [185]. PCA was also applied by Middlebrooks, who considered different HRTF measurements captured by in-ear microphones [187]. HRTF functions can also be simplified also by modelling low-order IIR filters [189,190]. In this case, starting from the measured HRTF, the optimal coefficients of the IIR filter of any desired order that represents the HRTF are calculated [189]. Moreover, interpolation techniques, such as bilinear or triangular interpolation, can be applied to the IIR coefficients [190]. The methods discussed above can reduce the order of the HRTF, decreasing the computational cost. However, the weights are only optimised for measured directions, and the final model is not described by a continuous function. In 1995, Chen et al. proposed a model based on Karhunen–Loève expansion that defines HRTFs in the complex frequency domain as a series of orthogonal eigentransfer functions (EFs) [192]. This solution offers a continuous function that can predict the phase and magnitude information of unmeasured directions.

Another more widespread and effective method that guarantees a continuous and orthogonal representation is the spherical harmonics decomposition (SHD). Following the Helmoltz reciprocity principle, SHD models the HRTF spectrum as the weighted sum of spherical harmonics [198]. In this way, the spectrum of a generic HRTF function in the spatial domain can be represented as follows:

$$H(\theta, \phi, f) = \sum_{n=0}^{N-1} \sum_{m=-n}^{n} c_n^m(f) Y_n^m(\theta, \phi), \tag{26}$$

where N is the truncation order of the SHD, $c_n^m(f)$ are the SH coefficients, and $Y_n^m(\theta, \phi)$ are the spherical harmonic basis function, defined as follows [213]:

$$Y_n^m(\theta, \phi) = (-1)^m \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}} P_n^{|m|}(\cos\theta) e^{im\phi}, \tag{27}$$

where $n = 0, 1, .., N$, $m = -n, ..., n$, and $P_n^m(\cdot)$ is the associated Legendre function. Starting with the measured HRTFs, the SH coefficients can be analytically calculated by inverting Equation (26), or by applying an optimisation technique, as the least square (LS) method [200]. Once $c_n^m(f)$ are computed, the HRTF of any position can be calculated using Equation (26). The first work based on SHD was presented by Evans et al. in 1998 [197]. In the following years, several approaches using SHD were applied. The performance of this method varies with the order of the SHD and the employed measurement grid. Noting N as the SHD order, at least N^2 measured HRTFs are necessary to guarantee an accurate representation, especially at higher frequencies [200]. Moreover, measurement point distribution, i.e., the grid, is also important for reconstruction accuracy. Thus, this method requires the entire dataset, not only the nearest neighbour positions. In this way, interpolation performance can be improved at the expense of the computational complexity, especially for dense datasets. In 2008, Zhang et al. explored the challenge of missing data at low elevations by proposing an iterative algorithm and using a fourth-order decomposition [199]. However, results showed that an order of four does not produce sufficiently accurate responses. In 2009, Zotkin et al. tested SHD interpolation performance with different types of grids, proving that the best results are obtained with closed grids, i.e., the measurement points cover the entire sphere surface [200]. In 2012, Aussal et al. applied spherical harmonics decomposition to interpolate ITD values of HRTFs measured with a Kemar mannequin; then, they used PCA to derive a full-sphere model [30,201]. In 2018, Brinkmann and Weinzierl investigated different pre-processing algorithms for time alignment applied before SHD [202]. The effectiveness of the pre-processing allows allows for a

reduction in the order of the SHD and the number of HRTF measurements. Results show that the error is below the just noticeable difference (see Section 2.2) for orders ≥ 4 [202].

In 2020, Pörschmann et al. compared different HRTF interpolation methods, including nearest-neighbour techniques and SHD. The results demonstrate that the performance of the SHD increases with the spatial resolution of the measurements [203]. Thus, the nearest-neighbour interpolation performs better with a low spatial resolution, while the accuracy is comparable to SHD when a higher spatial resolution is considered [203]. In 2021, Skarha obtained similar conclusions by comparing the SHD with the binaural interpolation of the four closest and evaluating the reconstruction error and the computational complexity [204]. The study shows that the bilinear interpolation outperforms the SHD when sparse HRTF datasets are involved, while SHD exhibits better performance with dense measurement grids. This emphasises the importance of selecting appropriate interpolation methods depending on the specific application, such as real-time audio rendering or offline audio production. Skarha also discusses the computational trade-offs associated with these methods, highlighting the balance between interpolation accuracy and processing speed [204]. In the same year, Arend et al. proved that SH representation is affected by spatial aliasing and truncation errors [205]. These problems are reduced by introducing a time-alignment procedure prior to interpolation [205]. In 2023, the same authors presented a new approach in which a magnitude correction procedure is applied after a time-aligned interpolation method, thus improving performance [29].

6.3. Neural Network-Based Interpolation

In recent years, machine learning techniques and neural networks have also been applied to HRTF interpolation. Cobos et al. [214] provide a comprehensive overview of data-driven methods, highlighting the machine learning techniques used for HRTF interpolation and capturing and processing spatial audio in real-time environments. In 2021, Lu and Qi proposed pre-trained models for HRTF interpolation, focusing on real-time applications [181]. This approach exhibits minimal latency, making it suitable for VR/AR interactive applications. This study also explores the limitations of such models, particularly in terms of generalization to different head and ear shapes that were not included during training [181]. In 2022, Siripornpitak et al. [188] explored the use of generative adversarial networks (GANs) for the spatial up-sampling of HRTF datasets. GANs are particularly suited to creating high-fidelity interpolations in sparse datasets, where traditional methods might struggle to maintain audio quality. This approach demonstrates significant improvements in the perceived spatial accuracy of audio, especially in complex 3D environments [188]. In the same year, Ito et al. proposed an HRTF interpolation approach using an autoencoder for spatially sparse measurement datasets [182]. The network is designed so that only the encoder and decoder depend on the position. This characteristic also allows good performance for sparse measurements and subjects not included in the training [182]. In 2023, Zurale and Dubnov proposed a 3D CNN based on a vector-quantised variational autoencoder (VQ-VAE) to spatially upsample sparse datasets [194].

In the same year, Lee et al. designed a neural network based on feature-wise linear modulation (FiLM) for incorporating HRTFs from datasets with different coordinate systems [37]. This approach is based on evaluating the neighbour HRTFs and anthropometric measurements, employing two conditioning methods. Results showed good generalisation with unknown datasets, outperforming the baseline [37]. In recent years, convolutional neural networks have often been used for HRTF interpolation [193,195]. In [66], a CNN was employed to interpolate HRTF samples collected using a multi-sine signal while the listener moves freely (as introduced in Section 3). The process begins by scanning the listener's upper body and head using 3D sensors, creating a detailed mesh that serves as the basis for generating spatially dense and direction-dense HRTFs. The experimental results obtained with this approach are depicted in Figure 10. The methods based on CNN described above require a plane–sphere projection, which may exclude some HRTF features. In 2023, Chen et al. investigated the problem of the coordinate system by proposing a

spherical convolutional neural network to solve the plane–sphere projection with spherical harmonics representation [206]. Based on this, in 2024, also Thuillier et al. also introduced a spherical convolutional conditional neural process (SConvCNP) that can manage spherical data typical of the HRTF geometry. This method has shown good results with the time-aligned HRTF spectrum; however, it introduces a high computational cost that makes the approach inapplicable to real-time solutions [207].

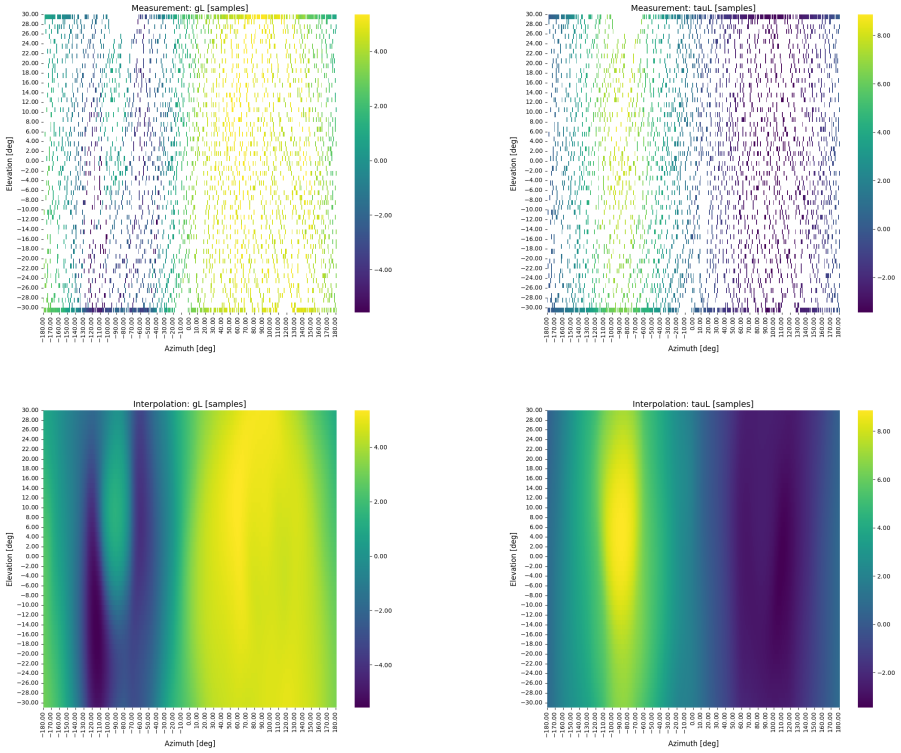


Figure 10. Interpolation of gain g and phase delay τ for the left ear at 1500 Hz. The first row represents the sparse lab data, while the second shows the result of the interpolation [66].

7. Technological Integration and Evaluation

The integration of these advanced HRTF techniques into consumer products, as well as their evaluation in real-world scenarios is a crucial aspect in ensuring their practical applicability. With the continuous evolution of audio technologies, companies such as Apple and Dolby, alongside academic projects like EarFish from the Active Perception Lab at the University of Antwerp, have pioneered the use of personalised HRTF solutions. These efforts bridge the gap between research and consumer adoption, aiming to enhance spatial audio experiences in consumer devices, from headphones to home theatre systems.

7.1. Consumer Technology Integration

The work by Liu [215] discusses the increasing integration of consumer technology, such as smartphones and VR headsets, into the process of capturing and utilising personalised HRTFs. They present a novel approach that integrates consumer technology, such as the iPhone's FaceID feature, to capture 3D meshes of a user's head. This method demonstrates the feasibility of using consumer-grade devices to generate personalised HRTFs. The study details the process of extracting head-related features from these 3D scans and discusses the accuracy and practicality of such an approach compared to traditional methods with specialised equipment. The ability to capture and fine-tune HRTFs using readily available consumer devices has democratised access to high-quality spatial audio experiences.

In the realm of commercial products, Apple's AirPods Pro and AirPods Max have been pivotal in integrating personalised HRTFs. Their dynamic head-tracking feature adjusts spatial audio in real-time, providing a highly immersive experience for music, movies, and VR content. The system uses machine learning to analyze users' unique auditory profiles, thereby delivering a personalised listening experience. Dolby's Dolby Atmos also integrates HRTF-based spatial audio technologies to create life-like 3D sound environments across media platforms, setting the standard for immersive audio in both consumer and professional markets.

The EarFish project, developed by the Active Perception Lab at the University of Antwerp, offers an accessible and affordable HRTF measurement kit that simplifies the personalisation process of spatial audio experiences. By utilising minimal and cost-effective equipment, such as compact microphones and user-friendly software, the EarFish system enables individuals to capture their personal HRTF measurements without the need for expensive hardware or extensive technical expertise. This democratisation of HRTF measurement empowers non-experts to benefit from accurate sound localisation and enhanced immersion in applications like virtual reality, gaming, and immersive media. Additionally, EarFish plays a crucial role in research settings by providing an affordable tool for collecting HRTF data from a larger and more diverse pool of participants, thereby facilitating studies on spatial hearing and audio perception. By reducing the cost and complexity associated with HRTF personalisation, the EarFish project contributes significantly to advancing spatial audio technologies and promoting the wider adoption of individualised audio experiences in both research and consumer applications.

The work by Binelli et al. [216] introduces a practical solution for individualised HRTFs in VR environments, specifically for playing 360-degree video content with Ambisonics spatial audio on head-mounted displays (HMDs). Compatible with both desktop and portable platforms, their system integrates personalised HRTFs through a straightforward, open-source approach. This flexibility enhances spatial audio immersion in HMDs without extensive equipment, demonstrating the potential of individualised HRTFs in consumer VR applications and making advanced spatial audio more accessible in virtual environments.

Tomasetti et al. [217] compared six tools for implementing spatial audio experiences in Extended Reality (XR) applications, emphasising the compatibility of these tools with HRTFs and the WebXR Device API. They highlight the challenges of integrating spatial audio into immersive experiences and the lack of standardised frameworks for delivering consistent audio performance. This study is critical for understanding how consumer-facing XR applications can benefit from HRTF personalisation and integration.

Incorporating individualised HRTFs into speech signal processing can significantly enhance the intelligibility and spatial localisation of spoken audio within immersive environments. This advancement is particularly beneficial for applications such as virtual meetings, telepresence, and assistive listening devices, where precise auditory cues are essential. For instance, research has demonstrated that personalised HRTFs improve speech intelligibility by providing accurate spatial cues, thereby reducing cognitive load and enhancing user experience in complex auditory scenes [218].

7.2. Evaluation in Real-World Applications

Real-world applications and evaluations are vital for validating the effectiveness of HRTF technologies. Andersen et al. [219] performed a detailed evaluation of individualised HRTFs within a dynamic 3D shooter game environment, emphasising the importance of testing personalised HRTFs in dynamic environments such as gaming. Their findings demonstrate that players equipped with individualised HRTFs experience enhanced spatial awareness, sound localisation, immersion, and improved reaction times compared to those using generic profiles. This research provides insights into how well these personalised profiles translate to real-world applications. It also supports the broader adoption of HRTF technologies in interactive and immersive environments, such as gaming and virtual

reality, where precise sound localisation significantly enhances the overall user experience. The study highlights the potential of personalised spatial audio in interactive applications.

Lu and Qi [181] evaluated various HRTF interpolation methods to determine their computational efficiency in real-time audio rendering systems. Their research highlights the trade-offs between achieving high-quality audio reproduction and maintaining low computational costs. This balance is critical for consumer products, where real-time spatial audio processing must operate with minimal latency while ensuring that sound quality remains high. This study also provides insights for developers of real-time spatial audio systems in terms of optimising performance with limited processing power.

López and Gutierrez-Parera [220] evaluate the frequency response of consumer headphones and their role in individualising HRTFs. They introduce a technique using a chain of peak filters, which models the frequency response of headphones, allowing for more personalised HRTF outputs. This approach reduces the need for extensive equipment and offers a practical solution for integrating personalised HRTFs into consumer devices. This work provides a foundation for understanding how spatial audio can be democratised, making high-quality HRTF personalisation more accessible to a wider audience.

Dolby Atmos and similar spatial audio solutions used in home theatre systems have been evaluated for their performance in various acoustic environments. These evaluations shed light on how HRTF-based technologies perform in everyday contexts, providing avenues for future refinements to improve user experience across different audio applications.

8. Conclusions and Future Directions

Due to the individual variability in human head and ear shapes, generating and optimising HRTFs for accurate and personalised spatial audio experiences remains a significant research challenge. In this paper, a review about the generation of head-related transfer functions for spatial audio is presented.

Starting from the mathematical definitions and settings used for HRTFs, a deep analysis of various methods for their generation is reported. HRTFs can be first obtained by direct measurements conducted using standard HAT simulators or human subjects equipped with in-ear microphones. These measurements can be conducted using several techniques characterised by different excitation signals. However, the measurement procedure is long and time consuming and requires professional equipment. As an alternative to direct measurements, physical modelling can be employed to generate HRTFs by numerically solving the wave equation in the time domain or the Helmholtz equation in the frequency domain. In this context, the boundary element method (BEM) is one of the most widely used techniques. An elementary explanation of the mathematical model based on BEM is provided. Additionally, other approaches, such as the finite element method (FEM), finite-difference time-domain (FDTD) method, and adaptive rectangular decomposition (ARD), are also discussed as alternative solutions.

Predictive models have a high computational cost, and they are not able to capture the full anatomical diversity of listeners, reducing precision for individuals with less typical features. Personalised HRTFs can significantly improve spatial binaural performance. They can be obtained by the direct measurement of human subjects using in-ear microphones or by applying anthropometric and morphological approaches. Several anthropometric parameters can be defined and measured to obtain personalised HRTFs through anthropometry matching, adaptation methods, or trained regression models.

In some cases, when high-quality HRTF datasets are available and the computational resources are limited, the HRTF interpolation can be applied for enhancing the spatial audio experience. Starting from a limited set of measured HRTFs, interpolation enables spatial upsampling, allowing the generation of HRTFs for positions that were not included in the original measurement points. The interpolated HRTF can be calculated as a linear combination of the nearest neighbors, as a result of functional models that can be derived from several approaches, such as SHD, or using neural networks trained on sparse measurements.

Finally, the integration of these HRTF generation techniques is examined, with a focus on their incorporation into consumer technologies and their evaluation in real-world applications. In particular, evaluations of HRTF technologies confirm their benefits in real-world applications, especially in gaming and virtual reality. Personalised HRTFs improve spatial awareness and immersion, while efficient interpolation methods help maintain audio quality with low latency. Techniques for matching headphone responses make HRTFs more accessible for consumer devices.

The proposed review can be a valuable reference for future research in this field since it investigates the innovations and the limitations of the current techniques. Despite the significant advancements introduced in the field of HRTF generation, several challenges remain. Achieving real-time performance without compromising accuracy requires further innovation in computational methods. Additionally, the integration of machine learning techniques, while promising, poses challenges related to generalisation across different listener anatomies and environments. Future research should focus on developing scalable methods incorporating low-latency computational models that can operate in real-time. Adaptive machine learning algorithms can be investigated for developing an immersive audio system capable of dynamically adjusting themselves to anatomical and environmental changes. Moreover, exploring hybrid approaches that combine perceptual metrics with statistical optimisation could yield more robust and user-friendly solutions for personalised spatial audio experiences.

Author Contributions: Conceptualisation, S.C. and T.L.; investigation, V.B., A.V., A.Q. and L.G.; writing—original draft preparation, V.B. and S.C.; writing—review and editing, V.B., L.G., N.A.D., A.Q., A.V., T.L. and S.C.; supervision, S.C. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by the European Union – Next Generation EU. Project Code: ECS00000041; Project CUP: C43C22000380007; Project Title: Innovation, digitalisation and sustainability for the diffused economy in Central Italy–VITALITY.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR	Augmented Reality
ARD	Adaptive Rectangular Decomposition
ASM	Active Shape Model
ASM	Active Shape Model
BEM	Boundary Element Method
BRIR	Binaural Room Impulse Response
CNN	Convolutional Neural Network
CVDD	Continuous Variable Digital Delay
DNN	Deep Neural Network
DTW	Dynamic Time Warping
EF	Eigentransfer Function
FDTD	Finite-Difference Time-Domain
FEM	Finite Element Method
FIR	Finite Impulse Response
FLiP	Functional Link Polynomial
FMBEM	Fast Multipole Boundary Method
FMM	Fast Multipole Method
FoV	Field of View
GAN	Generative Adversarial Network
GPU	Graphic Processing Unit
GRNN	Generalised Regression Neural Network
HAT	Head and Torso
HOSVD	High-Order Singular Value Decomposition

HRIR	Head-Related Impulse Response
HRTF	Head-related Transfer Function
IDW	Inverse-Distance Weighting
IIR	Infinite Impulse Response
ILD	Interaural Level Difference
IMU	Inertial Measurement Unit
IPTF	Inter-Positional Transfer Function
IRS	Inverse Repeated Sequence
ITD	Interaural Time Difference
JND	Just Noticeable Difference
LN	Legendre Non-linear
LS	Least Square
LSD	Log-Spectral Distortion
LTI	Linear Time-Invariant
MLP	Multilayer Perceptron
MLS	Maximum Length Sequence
MP	Minimum Phase
MSE	Mean Square Error
NLMS	Normalised Least Mean Square
OPS	Orthogonal Periodic Sequence
PCA	Principal Component Analysis
PML	Perfectly Matched Layer
PPS	Perfect Periodic Sequence
PRTF	Pinna-Related Transfer Function
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SARITA	Spherical Array Interpolation by Time Alignment
SCH	Spherical Cap Harmonic
SConvCNP	Spherical Convolutional Conditional Neural Process
SD	Spectral Distortion
SDE	Spectral Difference Error
SH	Spherical Harmonic
SHD	Spherical Harmonic Decomposition
SLAM	Simultaneous Localisation and Mapping
SPCA	Spatial Principal Component Analysis
SVD	Singular Value Decomposition
TCN	Temporal Convolutional Network
UWVF	Ultra-Weak Variational Formulation
VBAP	Vector base Amplitude Panning
VQ-VAE	Vector-Quantised Variational Autoencoder
VR	Virtual Reality
WN	Wiener Non-linear
XR	Extended Reality

Appendix A

This appendix presents an overview of the publicly available HRTF databases. Table A1 summarises the discussed databases, providing key information regarding the number of measurement points N, the range of considered elevations, azimuth, distance, the type of sound source, and the measurement method used for HRIR acquisitions. For measurements taken from human subjects, the number of subjects, and the type of microphone are reported. For measurements performed with mannequins, the type of head and torso (HAT) simulator is provided.

Table A1. List of the publicly available databases of HRTFs.

Name	N	Elevation	Azimuth	Distance	N Subj	Microphone	HAT	Sound Source	Meas. Method
MIT-KEMAR (1994) [41]	710	$[-40°; 90°]$	$[0°; 360°]$	1.4 m	-	-	KEMAR	Realistic Optimus Pro 7	MLS
CIPIC (2001) [147]	1250	$[-45°; 230.625°]$	$[0°; 360°]$	1 m	43	Etymotic Research ER-7C	KEMAR	Bose Acoustimass	Golay-code
LISTEN (2003) [221]	187	$[-45°; 90°]$	$[0°; 360°]$	1.95 m	51	Knowles FG3329	-	Tannoy system 600	Logarithmic sweep
PKU&IOA (2009) [222]	6344	$[-40°; 90°]$	$[0°; 360°]$	[0.2 m; 1.6 m]	-	-	KEMAR	Spark gap BDMS1-040528	Spark gap
FIU (2010) [223]	72	$[-36°; 54°]$	$[-150°; 180°]$	1 m	15	HeadZap Blocked-Meatus	-	HeadZap speaker	Golay-code
SYMARE (2013) [224]	393	$[-40°; 90°]$	$[0°; 360°]$	1 m	61	Sennheiser KE4-211-2	-	Vifa D26TG-35	Golay-code
TUM-LDV (2013) [225]	2160	$[-50°; 230°]$	$[0°; 360°]$	1.3 m	35	Knowles FG-23629	KEMAR	KS digital C5 tiny	Dynamic NLMS
ARI (2014) [13]	1550	$[-30°; 80°]$	$[0°; 360°]$	1 m	>250	Sennheiser KE-4-211-2	-	Vifa 10 BGS	Multiple exponential sweep
RIEC (2014) [226]	936	$[-80°; 90°]$	$[0°; 360°]$	1.5 m	105	Knowles FG3329	-	Fostex FE83E	OATSP signal
BiLi (2015) [227]	1680	$[-51°; 86°]$	$[0°; 360°]$	1 m	56	Knowles FG26107	-	ELAC 301	Exponential sweep
SADIE (2015) [228]	1550/170	$[-90°; 90°]$	$[0°; 360°]$	1.5 m	20	Knowles FG-23329-D65	KU100, KEMAR	Equator D5	Logarithmic sweep
ITA (2016) [229]	2304	$[-66°; 90°]$	$[0°; 360°]$	1.2 m	48	Sennheiser KE3	-	1" broadband speakers	Multiple exponential sweep
3D3A (2017) [91]	648	$[-30°; 75°]$	$[0°; 360°]$	0.76 m	38	Theoretica Appl. Physics BACCH-BM Pro	-	Genelec 8030A	Exponential sweep
SADIE II (2018) [16]	8802/2818/2114	$[-90°; 90°]$	$[0°; 360°]$	1.2 m	29	Knowles FG-23329-C05	KU100, KEMAR	Genelec 8010S	Exponential sweep
HUTUBS (2019) [89]	440	$[-90°; 90°]$	$[0°; 360°]$	1.47	94	Knowles FG 23329	FABIAN	Peerless NE65-04	NLMS
VIKING (2019) [230]	1513	$[-45°; 90°]$	$[0°; 360°]$	1 m	-	-	KEMAR	Genelec 8020CPM-6	Logarithmic sweep
SONICOM (2023) [90]	828	$[-45°; 90°]$	$[0°; 180°]$	1.5 m	120	Knowles FG-23329-P07	-	Peerless 830987	Multiple exponential sweep
LORIS (2024) [66]	8186	$[-42°; 42°]$	$[0°; 360°]$	1.5 m	1	OKM II	-	Adam A5X	Multisine

The database MIT-KEMAR (1994) [41] is obtained using a Kemar mannequin placed at a distance of 1.4 m from a Realistic Optimus Pro 7 loudspeaker inside the MIT anechoic chamber. A total number of measurement points of 710 is provided, considering the elevations from -40° to 90°, with a step of 10°, and azimuth from 0° to 360°, with variable steps $\geq 5^\circ$. The impulse responses have a length of 512 samples and are measured using maximum-length pseudo-random binary sequences (MLS) and a sampling frequency of 44.1 kHz.

The database CIPIC (2001) [147] is a subset of the HRTFs of more than 90 subjects measured at the CIPIC Interface Laboratory. The publicly available dataset consists of 43 real subjects and the Kemar mannequin with large and small pinnae, resulting in a total of 45 subjects. The dataset includes the impulse responses with a length of 200 samples for 1250 directions and anthropometric measurements. A total of 64 Bose Acoustimass loudspeakers were uniformly mounted along the hoop at 1 m to the subject to consider the elevations sampled in $360/64 = 5.625^\circ$ steps from -45° to 230.625°. The azimuth angles were sampled non-uniformly considering steps $\geq 5^\circ$. For real subjects measurements, Etymotic Research ER-7C probe microphones were used. The impulse responses were calculated with a sampling frequency of 44.1 kHz using Golay-code signals. Anthropometric measurements are also contained in the dataset.

The LISTEN database (2003) [221] consists of 187 locations per subject and 51 subjects. The HRIRs, 512 samples long , were measured at different azimuth resolutions for 10 respective elevation angles spaced at 15° from -45° to 90° with a sampling frequency of 44.1 kHz. A Tannoy system 600 loudspeaker was installed at a distance of 1.95 m and Knowles FG3329 microphones were used for the acquisitions using logarithmic sweep signals.

The PKU&IOA database (2009) [222] was created by using the Kemar mannequin and a spark gap type BDMS1-040528 as the sound source. The signal generated by the spark gap was about 0.5 ms, corresponding to a length of 32 samples with a sample rate of 65,536 Hz. The elevations were sampled from -40° to 90° with a 10° increment and a full 360° azimuth loop was considered, resulting in 793 locations, repeated for eight different distances, i.e., 20, 30, 40, 50, 75, 100, 130, and 160 cm, for a total of 6344 positions.

The FIU database (2010) [223] consists of individual HRTFs of 15 different subjects considering 6 elevations, i.e., -36°, -18°, -0°, 18°, 38°, and 54°, and 12 azimuths spaced of 30° from -150° to 180°, resulting in 72 HRIRs for each subject with a length of 256 samples, acquired with a sample rate of 96 kHz using Golay codes. The measurements were carried out using the AUSIM3D HeadZap system, specifically designed for HRTF acquisitions in reflective environments. Anthropometric measurements of the various parts of the pinna are included in the database.

The SYMARE database (2013) [224] provides HRIRs of 61 subjects for 393 directions, including elevations from -40° to 90° and azimuth from 0° to 360° with a resolution of 10°. The SYMARE database provides both acoustically measured HRIRs and HRIRs simulated using the BEM. Multiple high-resolution surface mesh models of the upper torso, head, and ears are provided.

The TUM-LDV database (2013) [225] contains the HRIRs and the anthropometric parameters of 35 human subjects and two Kemar mannequins. The acoustic measurement per elevation is achieved by a continuous counter-clockwise turn from -30° to 370° covering a full loop with an azimuth resolution of 1° or 0.2° considering 6 elevations in the range from -50° to 230°. The HRIRs have a length of 1024 samples, measured with a dynamic NLMS approach using a sampling frequency of 48 kHz.

The ARI database (2014) [13] is acquired by mounting twenty-two Vifa 10 BGS loudspeakers at fixed elevations from -30° to 80°. The subject is seated in the centre of the arc with a radius of 1 m, has in-ear Sennheiser KE-4-211-2 microphones, and can rotate, varying the azimuth with variable steps $\geq 2.5^\circ$. The dataset covers 1550 directions of the sphere and includes more than 250 subjects. The HRIRs are measured using the multiple exponential sweep method (MESM) using a sample rate of 48 kHz and a length of 256 samples. The anthropometric data of 60 listeners are also provided.

The database RIEC (2014) [226] includes the HRIRs of 105 subjects and 936 directions covered by a vertical circular loudspeaker array consisting of 35 Fostex FE83E loudspeakers. The loudspeakers are arranged at intervals of 10° of the elevation angle from −80° to 90°. The circular loudspeaker array can automatically rotate around the vertical axis; a resolution of 5° is chosen for the azimuth. The distance between the centre of the subject's head and each loudspeaker is 1.5 m. The HRIRs are measured using the optimised Aoshima's time-stretched pulse (OATSP) [231] as a source signal with a length of 8192 samples and a sampling frequency of 48 kHz. Anthropometric data for 39 subjects are also included.

For the BiLi (2015) [227] dataset, the head-related impulse responses of 56 subjects are measured in the IRCAM's anechoic chamber using the exponential sweep sine technique, non-coaxial two-way ELAC 301 loudspeakers, and a sampling rate of 96 kHz. Knowles FG26107 microphones are positioned at the entrance of the blocked ear canal. The spatial sampling scheme is based on a Gaussian grid and includes 1680 directions with a full azimuth range, and elevation ranging from −51° to 86°. The angular step size is approximately 6° in both dimensions. A database of 3D morphological data is also included.

The database SADIE (2015) [228] contains measurements from 20 human subjects and two dummy heads, i.e., a Neumann KU100 and a Kemar mannequin. For the subjects, 170 measurement points are considered, while 1550 positions are used for the mannequins. The distribution of the points covers the overall sphere surface and satisfies Ambisonic decoding up to the fifth order. They also contain reference measurements at azimuthal increments of 5°, only at 0° elevation for the subjects, and in steps of elevation of 10° for the mannequins. The HRIRs are 256 samples long that were measured using logarithmic sine sweeps at a sampling frequency of 44.1 kHz.

The ITA database (2016) [229] contains the 256-samples HRIRs of 48 subjects measured in 2304 positions with Sennheiser KE3 microphones using a sample rate of 44.1 kHz. The azimuth and the elevation have a resolution of 5°, and the elevation varies from −66° to 90°. A vertical arc of 64 1" broadband speakers is used for the measurements, and HRIRs are measured using the multiple-exponential sweep method. Three-dimensional models of the ear geometry, reconstructed from magnetic resonance imaging scans, are included.

The database 3D3A (2017) [91] contains HRIRs of 38 subjects considering 648 positions with the elevation ranging from −30° to 75° and azimuth from 0° to 360° with a resolution of 5°. The measurements are conducted with an arc of eight Genelec 8030A loudspeakers with a radius of 0.76 ± 0.005 and binaural microphones from Theoretica Applied Physics BACCH-BM Pro. All measurements are conducted at a sampling rate of 96 kHz, and the sine sweep signals have a duration of 500 ms. The HRIRs have a duration of 10 ms, corresponding to a length of 960 samples. The 3D morphological scans of the torso, head, and ears of the subjects are available.

The database SADIE II (2018) [16] is a follow-on from the original database SADIE [228]. It contains the HRIRs of 29 human subjects and the two HAT simulators Neumann KU100 and Kemar. A regular latitude–longitude distribution (15° elevation and variable azimuthal resolution) is considered. Twenty-three Genelec 8010S loudspeakers are installed at 23 unique elevations at a radius of 1.2 m. With the mannequins, 8802 directions are involved, and with subjects 2818 or 2114. The HRIRs have a length of 512 samples and are measured with exponential sine sweep using a sample rate of 96 kHz. However, sample rates of 48 kHz and 44.1 kHz are also available. Anthropometric data, such as 3D head scans and high-resolution pictures of the ear, are included.

The HUTUBS database (2019) [89] considers 94 human subjects and a FABIAN HAT simulator [232]. It consists of 440 measurement points, including elevations from −90° to 90° with 10° increment and different azimuth resolutions depending on the elevation ($\geq$10°). The measurement system consists of 37 Peerless NE65-04 2-inch drivers. The distance from the centre of the array to the membrane of the speakers is 1.47 m. HRIRs, 256 samples long, are then measured under continuous rotation of the subject using NLMS adaptive filters and a sampling frequency of 44.1 kHz. In total, 93 sets of anthropometric measures and 58 head meshes are available.

The database VIKING (2019) [230] comprises full-sphere HRTFs measured on 1513 positions with a Kemar mannequin with 20 different pairs of artificial pinnae attached. The Kemar mannequin is mounted on a 360° rotating cylindrical stand, and a Genelec 8020CPM-6 loudspeaker is mounted on an L-shaped rotating arm. Elevations from $-45°$ to 90° uniformly sampled in 5° are considered, and azimuths are sampled in different increments ($\geq 5°$). The HRIRs of 128-samples length are measured with a logarithmic sweep using a sample rate of 48 kHz.

The SONICOM database (2023) [90] considers 120 subjects and 828 positions. An arch of 23 full-range Peerless 830987 3-in drivers is used between $-45°$ and 225° of elevation, spaced every 10° between $-30°/30°$ and $150°/210°$ and spaced every 15° in the other positions. The distance between the loudspeaker drivers and the centre of the arch is 1.5 m. The HRIRs are measured with multiple exponential sweeps with a sample rate of 96 kHz, and widowed to a length of 512 samples. In addition, 3D models of ears, head, and torso are also available.

The database LORIS (2024) [66] is built from a non-uniform set of dynamic data measurements taken from a single subject. The elevations have a range from $-42°$ to 42°, approximatively, and the azimuths are non-uniformly distributed in the 360° full range for a total of 8186 positions. The HRIRs have a length of 8192 samples and are acquired with a sample rate of 48 kHz. A 3D model of the head is also provided.

References

1. Iida, K. *Head-Related Transfer Function and Acoustic Virtual Reality*; Springer: Singapore, 2019. [CrossRef]
2. Rafaely, B.; Tourbabin, V.; Habets, E.; Ben-Hur, Z.; Lee, H.; Gamper, H.; Arbel, L.; Birnie, L.; Abhayapala, T.; Samarasinghe, P. Spatial audio signal processing for binaural reproduction of recorded acoustic scenes–review and challenges. *Acta Acust.* **2022**, *6*, 47. [CrossRef]
3. Garas, J. *Adaptive 3D Sound Systems*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 566.
4. Cecchi, S.; Palestini, L.; Peretti, P.; Piazza, F.; Bettarelli, F. Sub-band adaptive crosstalk cancellation: A novel approach for immersive audio. In Proceedings of the 124th Audio Engineering Society Convention, Amsterdam, The Netherlands, 17–20 May 2008.
5. Pelzer, S.; Masiero, B.; Vorländer, M. 3D Reproduction of room acoustics using a hybrid system of combined crosstalk cancellation and ambisonics playback. In Proceedings of the International Conference on Software Architecture (ICSA), Honolulu, HI, USA, 21–28 May 2011. [CrossRef]
6. Cecchi, S.; Primavera, A.; Piazza, F.; Bettarelli, F.; Li, J. Advanced audio spatializer combined with a multipoint equalization system. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 3569–3576. [CrossRef]
7. Muhammad, I.; Jang, H.S.; Jeon, J.Y. Virtual sound field immersions by beamforming and effective crosstalk cancellation using wavelet transform analysis. In Proceedings of the Forum on Acoustic Waves and Applications, Karkow, Poland, 7–12 September 2014. [CrossRef]
8. Bruschi, V.; Nobili, S.; Cecchi, S. Real-time binaural synthesis of moving sound sources over loudspeakers. In Proceedings of the Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 8–10 September 2021. [CrossRef]
9. Kabzinski, T.; Jax, P. A causality-constrained frequency-domain least-squares filter design method for crosstalk cancellation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 2942–2956. [CrossRef]
10. Vancheri, A.; Leidi, T.; Heeb, T.; Grossi, L. Geometrical acoustics approach to cross talk cancellation. In Proceedings of the 152nd Audio Engineering Society Convention, Hague, The Netherlands, 7–8 May 2022.
11. Vancheri, A.; Leidi, T.; Heeb, T.; Grossi, L. Multiband time-domain crosstalk cancellation. In Proceedings of the 153rd Audio Engineering Society Convention, Hague, The Netherlands, 7–8 May 2022.
12. Vancheri, A.; Leidi, T.; Heeb, T.; Grossi, L. Dynamic adaptation in geometrical acoustic CTC. In Proceedings of the 154th Audio Engineering Society Convention, New York, NY, USA, 25–27 October 2023.
13. Majdak, P.; Masiero, B.; Fels, J. Sound localization in individualized and non-individualized crosstalk cancellation systems. *J. Acoust. Soc. Am.* **2013**, *133*, 2055–2068. [CrossRef] [PubMed]
14. Brinkmann, F.; Lindau, A.; Weinzierl, S. On the authenticity of individual dynamic binaural synthesis. *J. Acoust. Soc. Am.* **2017**, *142*, 1784–1795. [CrossRef] [PubMed]
15. Møller, H.; Sørensen, M.F.; Jensen, C.B.; Hammershøi, D. Binaural technique: Do we need individual recordings? *J. Audio Eng. Soc.* **1996**, *44*, 451–469.
16. Armstrong, C.; Thresh, L.; Murphy, D.; Kearney, G. A perceptual evaluation of individual and non-individual HRTFs: A case study of the SADIE II database. *Appl. Sci.* **2018**, *8*, 2029. [CrossRef]

17. Usher, J.; Martens, W.L. Perceived naturalness of speech sounds presented using personalized versus non-personalized HRTFs. In Proceedings of the International Conference on Auditory Display (ICAD), Montreal, QC, Canada, 26–29 June 2007; Georgia Institute of Technology: Atlanta, GA, USA.
18. Hirahara, T.; Sagara, H.; Toshima, I.; Otani, M. Head movement during head-related transfer function measurements. *Acoust. Sci. Technol.* **2010**, *31*, 165–171. [CrossRef]
19. Masiero, B.; Pollow, M.; Fels, J. Design of a fast broadband individual head-related transfer function measurement system. In Proceedings of the Forum Acusticum, Aalborg, Denmark, 27 June–1 July 2011; pp. 2197–2202.
20. Bolaños, J.G.; Pulkki, V. HRIR database with measured actual source direction Data. In Proceedings of the 133rd Audio Engineering Society Convention, San Francisco, CA, USA, 26–29 October 2012.
21. Carpentier, T.; Bahu, H.; Noisternig, M.; Warusfel, O. Measurement of a head-related transfer function database with high spatial resolution. In Proceedings of the 7th Forum Acusticum (EAA), Krakow, Poland, 7–12 September 2014.
22. Avendano, C.; Duda, R.O.; Algazi, R. Modeling the contralateral HRTF. In Proceedings of the 16th AES International Conference on Spatial Sound Reproduction, Rovaniemi, Finland, 10–12 April 1999.
23. Hugeng, H.; Anggara, J.; Gunawan, D. Implementation of 3D HRTF interpolation in synthesizing virtual 3D moving sound. *Int. J. Technol.* **2017**, *8*, 186–195. [CrossRef]
24. Gamper, H. Head-related transfer function interpolation in azimuth, elevation, and distance. *J. Acoust. Soc. Am.* **2013**, *134*, 547–553. [CrossRef]
25. Biscainho, L.W.P.; Freeland, F.P.; Diniz, P.S.R. Using inter-positional transfer functionsin 3D-sound. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Orlando, FL, USA, 13–17 May 2002; pp. 1961–1964. [CrossRef]
26. Freeland, F.P.; Biscainho, L.W.P.; Diniz, P.S.R. Interpolation of head-related transfer functions (HRTFs): A multi-source approach. In Proceedings of the 12th European Signal Processing Conference (EUSIPCO), Vienna, Austria, 6–10 September 2004; IEEE: Piscataway, NJ, USA, 2004; pp. 1761–1764.
27. Bruschi, V.; Dourou, N.; Carini, A.; Cecchi, S. A New HRTF interpolation approach for nonlinear 3D audio systems. In Proceedings of the Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 8–10 September 2021; IEEE: Piscataway, NJ, USA, 2023; pp. 1–9. [CrossRef]
28. Hartung, K.; Braasch, J.; Sterbing, S.J. Comparison of different methods for the interpolation of Head-Related Transfer Functions. In Proceedings of the 16th Audio Engineering Society Conference on Spatial Sound Reproduction, Rovaniemi, Finland, 10–12 April 1999.
29. Arend, J.M.; Pörschmann, C.; Weinzierl, S.; Brinkmann, F. Magnitude-corrected and time-aligned interpolation of head-related transfer functions. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2023**, *31*, 3783–3799. [CrossRef]
30. Aussal, M.; Alouges, F.; Katz, B.F.G. ITD interpolation and personalization for binaural synthesis using spherical harmonics. In Proceedings of the 25th Audio Engineering Society Conference, York, UK, 25–27 March 2012; Volume 4.
31. Tu, W.; Hu, R.; Wang, H.; Chen, W. Measurement and analysis of just noticeable difference of interaural level difference cue. In Proceedings of the International Conference on Multimedia Technology, Kaifeng, China, 24–25 April 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1–3.
32. Andreopoulou, A.; Katz, B.F. Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees. *J. Multimodal User Interfaces* **2016**, *10*, 259–271. [CrossRef]
33. Garcia-Gomez, V.; Lopez, J.J. Binaural room impulse responses interpolation for multimedia real-time applications. In Proceedings of the 144th Audio Engineering Society Convention, Milan, Italy, 23–26 May 2018.
34. Queiroz, M.; Sousa, G. Efficient binaural rendering of moving sound sources using HRTF interpolation. *J. New Music Res.* **2011**, *40*, 239–252. [CrossRef]
35. Katz, B.F.; Parseihian, G. Perceptually based head-related transfer function database optimization. *J. Acoust. Soc. Am.* **2012**, *131*, EL99–EL105. [CrossRef]
36. International Telecommunication Union. *Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems*; Recommendation ITU-R BS.1534-3; ITU: Geneva, Switzerland, 2015.
37. Lee, J.W.; Lee, S.; Lee, K. Global HRTF interpolation via learned affine transformation of hyper-conditioned features. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2013; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5. [CrossRef]
38. Baumgartner, R.; Majdak, P.; Laback, B. Modeling sound-source localization in sagittal planes for human listeners. *J. Acoust. Soc. Am.* **2014**, *136*, 791–802. [CrossRef] [PubMed]
39. Plaskota, P. Research of acoustical impedance of human skin. *Vib. Phys. Syst.* **2019**, *30*, 337–342.
40. Burkhard, M.D.; Sachs, R.M. Anthropometric manikin for acoustic research. *J. Acoust. Soc. Am.* **1975**, *58*, 214–222. [CrossRef] [PubMed]
41. Gardner, B.; Martin, K. HRTF measurements of a KEMAR dummy-head microphone. *MIT Media Lab Perceptual Computing Technical Report 280*. 1994. Available online: https://sound.media.mit.edu/resources/KEMAR.html (accessed on 24 November 2024).

42. Bovbjerg, B.P.; Christensen, F.; Minnaar, P.; Chen, X. Measuring the Head-Related Transfer Functions of an artificial head with a high-directional resolution. In Proceedings of the 109th Audio Engineering Society Convention, Los Angeles, CA, USA, 22–25 September 2000.
43. Richter, J.G.; Fels, J. On the influence of continuous subject rotation during high-resolution head-related transfer function measurements. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *27*, 730–741. [CrossRef]
44. Wierstorf, H.; Geier, M.; Spors, S. A free database of head related impulse response measurements in the horizontal plane with multiple distances. In Proceedings of the 130th Audio Engineering Society Convention, London, UK, 13–16 May 2011.
45. Li, Y.; Preihs, S.; Peissig, J. Acquisition of continuous-distance near-field head-related transfer functions on KEMAR using adaptive filtering. In Proceedings of the 152nd Audio Engineering Society Convention, Hague, The Netherlands, 7–8 May 2022.
46. Møller, H.; Sørensen, M.F.; Hammershøi, D.; Jensen, C.B. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.* **1995**, *43*, 300–321.
47. Ye, Q.; Dong, Q.; Zhang, Y.; Li, X. Fast head-related transfer function measurement in complex environments. In Proceedings of the 20th International Congress on Acoustics, Sydney, Australia, 23–27 August 2010; pp. 23–27.
48. Yu, G.; Wu, R.; Liu, Y.; Xie, B. Near-field head-related transfer-function measurement and Ddatabase of human subjects. *J. Acoust. Soc. Am.* **2018**, *143*, 194–198. [CrossRef]
49. Li, S.; Peissig, J. Measurement of head-related transfer functions: A review. *Appl. Sci.* **2020**, *10*, 5014. [CrossRef]
50. Gardner, B.; Martin, K. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.* **1995**, *97*, 3907–3908. [CrossRef]
51. Briggs, P.A.N.; Godfrey, K.R. Pseudorandom signals for the dynamic analysis of multivariable systems. *Proc. Inst. Electr. Eng.* **1966**, *113*, 1259–1267. [CrossRef]
52. MacWilliams, F.J.; Sloane, N.J.A. Pseudo-random sequences and arrays. *Proc. IEEE* **1976**, *64*, 1715–1729. [CrossRef]
53. Schroeder, M.R. Integrated-impulse method measuring sound decay without using impulses. *J. Acoust. Soc. Am.* **1979**, *66*, 497–500. [CrossRef]
54. Borish, J.; Angell, J.B. An efficient algorithm for measuring the impulse response using pseudorandom noise. *J. Audio Eng. Soc.* **1983**, *31*, 478–488.
55. Mommertz, E.; Müller, S. Measuring impulse responses with digitally pre-emphasized pseudorandom noise derived from maximum-length sequences. *Appl. Acoust.* **1995**, *44*, 195–214. [CrossRef]
56. Ream, N. Nonlinear identification using inverse-repeatm sequences. *Proc. Inst. Electr. Eng. IET* **1970**, *117*, 213–218. [CrossRef]
57. Dunn, C.; Hawksford, M.J. Distortion immunity of MLS-derived impulse response Measurements. *J. Audio Eng. Soc.* **1993**, *41*, 314–335.
58. Golay, M. Complementary series. *IRE Trans. Inf. Theory* **1961**, *7*, 82–87. [CrossRef]
59. Müller, S.; Massarani, P. Transfer-function measurement with sweeps. *J. Audio Eng. Soc.* **2001**, *49*, 443–471.
60. Müller, S. Measuring transfer-functions and impulse responses. In *Handbook of Signal Processing in Acoustics*; Springer: New York, NY, USA, 2008; pp. 65–85. [CrossRef]
61. Rothbucher, M.; Veprek, K.; Paukner, P.; Habigt, T.; Diepold, K. Comparison of head-related impulse response measurement approaches. *J. Acoust. Soc. Am.* **2013**, *134*, 223–229. [CrossRef] [PubMed]
62. Heyser, R.C. Acoustical measurements by time delay spectrometry. *J. Audio Eng. Soc.* **1967**, *15*, 370–382.
63. Farina, A. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *J. Audio Eng. Soc.* **2000**, *108*, 5093.
64. Majdak, P.; Balazs, P.; Laback, B. Multiple exponential sweep method for fast measurement of head-related transfer functions. *J. Audio Eng. Soc.* **2007**, *55*, 623–637.
65. Dietrich, P.; Masiero, B.; Vorländer, M. On the optimization of the multiple exponential sweep method. *J. Audio Eng. Soc.* **2013**, *61*, 113–124.
66. Quattrini, A.; Vancheri, A.; Leidi, T.; Heeb, T.; Grossi, L.; Oldano, G.; Spagnoli, N. An approach for mesh-based generation of spatially dense, lowpass-filtered, individualized HRTFs using dynamic data acquisition. In Proceedings of the 156th Audio Engineering Society Convention, Madrid, Spain, 15–17 June 2024.
67. Stan, G.B.; Embrechts, J.J.; Archambeau, D. Comparison of different impulse response measurement techniques. *J. Audio Eng. Soc.* **2002**, *50*, 249–262.
68. Pulkki, V.; Laitinen, M.V.; Sivonen, V. HRTF measurements with a continuously moving loudspeaker and swept sines. In Proceedings of the 128th Audio Engineering Society Convention, London, UK, 22–25 May 2010.
69. Enzner, G. Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, NV, USA, 31 March–4 April 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 393–396. [CrossRef]
70. Correa, C.K.; Li, S.; Peissig, J. Analysis and comparison of different adaptive filtering algorithms for fast continuous HRTF measurement. In Proceedings of the Tagungsband Fortschritte der Akustik—DAGA, Kiel, Germany, 6–9 March 2017; pp. 6–9.
71. Haykin, S.S. *Adaptive Filter Theory*; Pearson Education: Rajam, Andhra Pradesh, India , 2002. [CrossRef]
72. Xie, B.; Zhong, X. *Head-Related Transfer Function and Virtual Auditory Display*; InTech: Penang, Malaysia, 2014. [CrossRef]
73. Carini, A.; Cecchi, S.; Romoli, L.; Sicuranza, G.L. Perfect periodic sequences for Legendre nonlinear filters. In Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 1–5 September 2014; pp. 2400–2404.

74. Carini, A.; Cecchi, S.; Romoli, L. Room impulse response estimation using perfect sequences for Legendre nonlinear filters. In Proceedings of the 23rd European Signal Processing Conference (EUSIPCO), Nice, France, 31 August–4 September 2015. [CrossRef]
75. Carini, A.; Romoli, L.; Cecchi, S.; Orcioni, S. Perfect periodic sequences for nonlinear Wiener filters. In Proceedings of the 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 28 August–2 September 2016. [CrossRef]
76. Cecchi, S.; Bruschi, V.; Nobili, S.; Terenzi, A.; Carini, A. Using periodic sequences for HRTFs measurement robust towards nonlinearities in automotive audio applications. In Proceedings of the IEEE International Workshop on Metrology for Automotive (MetroAutomotive), Modena, Italy, 4–6 July 2022; pp. 99–104. [CrossRef]
77. Carini, A.; Cecchi, S.; Romoli, L. Robust room impulse response measurement using perfect sequences for Legendre nonlinear filters. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2016**, *24*, 1969–1982. [CrossRef]
78. Carini, A.; Cecchi, S.; Terenzi, A.; Orcioni, S. On room impulse response measurement using perfect sequences for Wiener nonlinear filters. In Proceedings of the 26th European Signal Processing Conference (EUSIPCO), Eternal, Italy, 3–7 September 2018; pp. 982–986. [CrossRef]
79. Carini, A.; Cecchi, S.; Orcioni, S. Robust room impulse response measurement using perfect periodic sequences for Wiener nonlinear filters. *Electronics* **2020**, *9*, 1793. [CrossRef]
80. Carini, A.; Orcioni, S.; Terenzi, A.; Cecchi, S. Orthogonal periodic sequences for the identification of functional link polynomial filters. *IEEE Trans. Signal Process.* **2020**, *68*, 5308–5321. [CrossRef]
81. Carini, A.; Orcioni, S.; Cecchi, S. On room impulse response measurement using orthogonal periodic sequences. In Proceedings of the 27th European Signal Processing Conference (EUSIPCO), Coruna, Spain, 2–6 September 2019; pp. 1–5. [CrossRef]
82. Carini, A.; Cecchi, S.; Terenzi, A.; Orcioni, S. A room impulse response measurement method robust towards nonlinearities based on orthogonal periodic Sequences. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 3104–3117. [CrossRef]
83. Reijniers, J.; Partoens, B.; Steckel, J.; Peremans, H. HRTF measurement by means of unsupervised head movements with respect to a single fixed speaker. *IEEE Access* **2020**, *8*, 92287–92300. [CrossRef]
84. Jayaram, V.; Kemelmacher-Shlizerman, I.; Seitz, S.M. HRTF estimation in the wild. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, San Francisco, CA, USA, 29 October–1 November 2023; pp. 1–9. [CrossRef]
85. Yang, Z.; Choudhury, R. Personalizing head related transfer functions for earables. In Proceedings of the 2021 ACM SIGCOMM Conference, Virtual Event, 9 August 2021; pp. 137–150. [CrossRef]
86. Yamamoto, K.; Igarashi, T. Fully perceptual-based 3D spatial sound individualization with an adaptive variational autoencoder. *ACM Trans. Graph. (TOG)* **2017**, *36*, 1–13. [CrossRef]
87. Gebru, I.D.; Marković, D.; Richard, A.; Krenn, S.; Butler, G.A.; De la Torre, F.; Sheikh, Y. Implicit HRTF modeling using temporal convolutional networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 3385–3389. [CrossRef]
88. Sanyal, S.; Bolkart, T.; Feng, H.; Black, M.J. Learning to regress 3D face shape and expression from an image without 3D supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7763–7772. [CrossRef]
89. Brinkmann, F.; Manoj, D.; Pelzer, R.; Wohlgemuth, J.J.; Seipel, F.; Voss, D.; Grosche, P.; Weinzierl, S. *The HUTUBS Head-Related Transfer Function (HRTF) Database*; TU Berlin: Berlin, Germany, 2019. [CrossRef]
90. Engel, I.; Daugintis, R.; Vicente, T.; Hogg, A.O.T.; Pauwels, J.; Tournier, A.J.; Picinali, L. The sonicom HRTF dataset. *J. Audio Eng. Soc.* **2023**, *71*, 241–253. [CrossRef]
91. Sridhar, R.; Tylka, J.; Choueiri, E. A database of head-related transfer function and morphological measurements. In Proceedings of the 143rd Audio Engineering Society Convention, New York, NY, USA, 18–21 October 2017.
92. Costabel, M. Principles of boundary element methods. *Comput. Phys. Rep.* **1987**, *6*, 243–274. [CrossRef]
93. Kirkup, S. The boundary element method in acoustics: A survey. *Appl. Sci.* **2019**, *9*, 1642. [CrossRef]
94. Śmigaj, W.; Betcke, T.; Arridge, S.; Phillips, J.; Schweiger, M. Solving boundary integral problems with BEM++. *ACM Trans. Math. Softw. (TOMS)* **2015**, *41*, 1–40. [CrossRef]
95. Kreuzer, W.; Pollack, K.; Brinkmann, F.; Majdak, P. NumCalc: An open-source BEM code for solving acoustic scattering problems. *Eng. Anal. Bound. Elem.* **2024**, *161*, 157–178. [CrossRef]
96. Bai, M.R.; Ih, J.G.; Benesty, J. *Acoustic Array Systems: Theory, Implementation, and Application*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
97. Kahana, Y.; Nelson, P.A.; Petyt, M.; Choi, S. Numerical modelling of the transfer functions of a dummy-head and of the external ear. In Proceedings of the 16th Audio Engineering Society International Conference on Spatial Sound Reproduction, Rovaniemi, Finland, 10–12 April 1999; pp. 1–6.
98. Zotkin, D.N.; Duraiswami, R.; Grassi, E.; Gumerov, N.A. Fast head-related transfer function measurement via reciprocity. *J. Acoust. Soc. Am.* **2006**, *120*, 2202–2215. [CrossRef]
99. Gumerov, N.A.; O'Donovan, A.E.; Duraiswami, R.; Zotkin, D.N. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *J. Acoust. Soc. Am.* **2010**, *127*, 370–386. [CrossRef]
100. Marburg, S. Six boundary elements per wavelength: Is that enough? *J. Comput. Acoust.* **2002**, *10*, 25–51. [CrossRef]

101. Kreuzer, W.; Pollack, K.; Majdak, P.; Brinkmann, F. Mesh2HRTF/NumCalc: An open-source project to calculate HRTFs and wave scattering in 3D. In Proceedings of the Euroregio BNAM Joint Acoustics Conference, Aalborg, Denmark, 9–11 May 2022; pp. 443–452.
102. Rokhlin, V. Rapid solution of integral equations of classical potential theory. *J. Comput. Phys.* **1985**, *60*, 187–207. [CrossRef]
103. Coifman, R.; Rokhlin, V.; Wandzura, S. The fast multipole method for the wave equation: A pedestrian prescription. *IEEE Antennas Propgation Mag.* **1993**, *35*, 7–12. [CrossRef]
104. Chen, Z.S.; Waubke, H.; Kreuzer, W. A formulation of the fast multipole boundary method (FMBEM) for acoustic radiation and scattering from three dimensional structures. *J. Comput. Acoust.* **2008**, *16*, 303–320. [CrossRef]
105. Kreuzer, W.; Majdak, P.; Chen, Z. Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. *J. Acoust. Soc. Am.* **2009**, *126*, 1280–1290. [CrossRef] [PubMed]
106. COMSOL Multiphysics®. Available online: https://www.comsol.com (accessed on 24 November 2024).
107. FastBEMAcoustics®. Available online: https://www.fastbem.com (accessed on 24 November 2024).
108. Ansys ®. Available online: https://www.ansys.com (accessed on 24 November 2024).
109. Mesh2HRTF. Available online: https://www.mesh2hrtf.org (accessed on 24 November 2024).
110. Brinkmann, F.; Kreuzer, W.; Thomsen, J.; Dombrovskis, S.; Pollack, K.; Weinzierl, S.; Majdak, P. Recent advances in an open software for numerical HRTF calculation. *J. Audio Eng. Soc.* **2023**, *71*, 502–514. [CrossRef]
111. Ziegelwanger, H.; Kreuzer, W.; Majdak, P. Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions. In Proceedings of the 22nd International Congress on Sound and Vibration, Florence, Italy, 12–16 July 2015. [CrossRef]
112. Bempp, Boundary Element Method Python Package. Available online: https://bempp.com/ (accessed on 24 November 2024).
113. Betcke, T.; Scroggs, M. Bempp-cl: A fast python based just-in-time compiling boundary element library. *J. Open Source Softw.* **2021**, *6*, 2879. [CrossRef]
114. Burton, A.J.; Miller, G.F. The application of the integral equation methods to the numerical solution of some exterior boundary-value problems. *Proc. R. Soc. Lond. A* **1971**, *323*, 201–210. [CrossRef]
115. Gumerov, D. A broadband fast multipole accelerated boundary element method for the 3D Helmholtz equation. *J. Acoust. Soc. Am.* **2009**, *125*, 191–205. [CrossRef]
116. Fiala, P.; Huijssen, J.; Pluymers, B.; Hallez, R.; Desmet, W. Fast multipole BEM modelling of head related transfer functions of a dummy head and torso. In Proceedings of the ISMA 2010 Conference, Katoomba, Australia, 30–31 August 2010.
117. Tang, L.; Fu, Z.H.; Xie, L. Numerical calculation of the head-related transfer functions with chinese dummy head. In Proceedings of the 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, Taiwan, 29 October–1 November 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1–4. [CrossRef]
118. Brinkmann, F.; Lindau, A.; Müller-Trapet, M.; Vorländer, M.; Weinzierl, S. *Cross-Validation of Measured and Modeled Head-Related Transfer Functions*; Technische Universität Berlin: Berlin, Germany, 2019.
119. Fan, Z.; Arce, T.; Lu, C.; Zhang, K.; Wu, T.W.; McMullen, K. Computation of head-related transfer functions using graphics processing units and a perceptual validation of the computed hrtfs against measured HRTFs. In Proceedings of the Audio Engineering Society International Conference on Headphone Technology, San Francisco, CA, USA, 27–29 August 2019.
120. Hargreaves, J.A.; Rendell, L.R.; Lam, Y.W. A framework for auralization of boundary element method simulations including source and receiver directivity. *J. Acoust. Soc. Am.* **2019**, *145*, 2625–2637. [CrossRef]
121. Zhang, M.; Wang, J.H.; James, D.L. Personalized HRTF modeling using DNN-augmented BEM. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 451–455. [CrossRef]
122. Braren, H.S.; Fels, J. Towards child-appropriate virtual acoustic environments: A database of high-resolution HRTF measurements and 3D-scans of children. *Int. J. Environ. Res. Public Health* **2021**, *19*, 324. [CrossRef]
123. Huang, X.Y.; Wang, Y.; Liu, Y.; Ni, B.; Zhang, W.; Liu, J.; Li, T. AudioEar: Single-view ear reconstruction for personalized spatial audio. In Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI 2023), Washington, DC, USA, 7–14 February 2023. [CrossRef]
124. Liu, Q.H.; Tao, J. The perfectly matched layer for acoustic waves in absorptive media. *J. Acoust. Soc. Am.* **1997**, *102*, 2072–2082. [CrossRef]
125. Xiao, T.; Liu, H.Q. Finite difference computation of head-related transfer function for human hearing. *J. Acoust. Soc. Am.* **2003**, *113*, 2434–2441. [CrossRef] [PubMed]
126. Takemoto, H.; Mokhtari, P.; Kato, H.; Nishimura, R.; Iida, K. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *J. Acoust. Soc. Am.* **2012**, *132*, 3832–3841. [CrossRef] [PubMed]
127. Mokhtari, P.; Takemoto, H.; Nishimura, R.; Kato, H. Computer simulation of KEMAR's head-related transfer functions: Verification with measurements and acoustic effects of modifying head shape and pinna concavity. In *Principles and Applications of Spatial Hearing*; World Scientific: Singapore, 2011; pp. 205–215. [CrossRef]
128. Meyer, J.; Picinali, L. Comparison of simulated head-related transfer functions accuracy for different model complexities using the finite-difference time-domain method. In Proceedings of the 10th Convention of the European Acoustics Association (EAA), Torino, Italy, 11–15 September 2023. [CrossRef]

129. Prepelita, S. Verification and Validation of Wave-Based Simulations of Head-Related Transfer Functions. Ph.D. Thesis, Aalto University, Espoo, Finland, 2020.
130. Raghuvanshi, N.; Narain, R.; Lin, M.C. Efficient and accurate sound propagation using adaptive rectangular decomposition. *IEEE Trans. Vis. Comput. Graph.* **2009**, *15*, 789–801. [CrossRef] [PubMed]
131. Meshram, A.; Mehra, R.; Yang, H.; Dunn, E.; Franm, J.M.; Manocha, D. P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound. In Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 10–12 September 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 53–61. [CrossRef]
132. Huttunen, T.; Seppälä, E.T.; Kirkeby, O.; Kärkkäinen, A.; Kärkkäinen, L. Simulation of the transfer function for a head-and-torso model over the entire audible frequency range. *J. Comput. Acoust.* **2007**, *15*, 429–448. [CrossRef]
133. Ma, F.; Wu, J.H.; Huang, M.; Zhang, W.; Hou, W.; Bai, C. Finite element determination of the head-related transfer function. *J. Mech. Med. Biol.* **2015**, *15*, 1550066. [CrossRef]
134. Kailas, G.; Tiwari, N. Efficient computational techniques for evaluating distance-dependent head-related transfer functions. *Acoust. Aust.* **2022**, *50*, 231–245. [CrossRef]
135. van de Par, S.; Bechtold, T.; Schütz, A.; Roden, R.; Blau, M. Highly efficient computation of HRTFs by Krylov subspace-based model order reduction for virtual acoustical rendering. In Proceedings of the 10th Convention of the European Acoustics Association (EAA), Torino, Italy, 11–15 September 2023. [CrossRef]
136. Hetmaniuk, U.; Tezaur, R.; Farhat, C. Review and assessment of interpolatory model order reduction methods for frequency response structural dynamics and acoustics problems. *Int. J. Numer. Methods Eng.* **2012**, *90*, 1636–1662. [CrossRef]
137. Röber, N.; Andres, S.; Masuch, M. *HRTF Simulations Through Acoustic Raytracing*; Universitäts-und Landesbibliothek Sachsen-Anhalt: Halle, Germany, 2006.
138. Reijniers, J.; Partoens, B.; Peremans, H. Torso reflection model for dynamic head-related transfer function. In Proceedings of the 10th Convention of the European Acoustics Association (EAA), Torino, Italy, 11–15 September 2023. [CrossRef]
139. Cecchi, S.; Primavera, A.; Virgulti, M.; Bettarelli, F.; Li, J.; Piazza, F. An efficient implementation of acoustic crosstalk cancellation for 3D audio rendering. In Proceedings of the 2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP), Xi'an, China, 9–13 July 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 212–216.
140. Lee, K.S.; Lee, S.P. A real-time audio system for adjusting the sweet spot to the listener's position. *IEEE Trans. Consum. Electron.* **2010**, *56*, 835–843. [CrossRef]
141. Partridge, C.J. Sound Wave Scattering from a Rigid Sphere. Unknown 1993. Available online: https://apps.dtic.mil/sti/tr/pdf/ADA264182.pdf (accessed on 24 November 2024).
142. Turley, S. Acoustic scattering from a sphere. In *Class Notes, Department of Physics and Astronomy*; Brigham Young University: Provo, UT, USA, 2006.
143. Aaronson, N.L.; Hartmann, W.M. Testing, correcting, and extending the Woodworth model for interaural time difference. *J. Acoust. Soc. Am.* **2014**, *135*, 817–823. [CrossRef]
144. Shin, K.h.; Park, Y. Modeling of non-individualized head-related transfer functions for nearby sources. In Proceedings of the 9th Western Pacific Acoustics Conference (WESPAC), Seoul, Republic of Korea, 26–28 June 2006; pp. 164–172.
145. Sottek, R.; Genuit, K. *Physical Modeling of Individual Head-Related Transfer Functions*; Deutsche Gesellschaft für Akustik, Ed.; Tagung der Deutschen Arbeitsgemeinschaft für Akustik: Potsdam, Germany, 1999.
146. Algazi, V.R.; Duda, R.O.; Duraiswami, R.; Gumerov, N.A.; Tang, Z. Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. Am.* **2002**, *112*, 2053–2064. [CrossRef] [PubMed]
147. Algazi, V.R.; Duda, R.O.; Thompson, D.M.; Avendano, C. The cipic HRTF database. In Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575), New Paltz, NY, USA, 21–24 October 2001; IEEE: Piscataway, NJ, USA, 2001; pp. 99–102. [CrossRef]
148. Schönstein, D.; Katz, B.F.G. HRTF selection for binaural synthesis from a database using morphological parameters. In Proceedings of the International Congress on Acoustics (ICA), Sydney, Australia, 23–27 August 2010.
149. Torres-Gallegos, E.A.; Orduña-Bustamante, F.; Arámbula-Cosío, F. Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database. *Appl. Acoust.* **2015**, *97*, 84–95. [CrossRef]
150. Warnecke, M.; Jamison, S.; Prepelita, S.; Calamia, P.; Ithapu, V.K. HRTF personalization based on ear morphology. In Proceedings of the Audio Engineering Society Conference: Audio for Virtual and Augmented Reality, Redmond, WV, USA, 15–17 August 2022.
151. Zhi, B.; Zotkin, D.N.; Duraiswami, R. Towards fast And cConvenient end-To-end HRTF personalization. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Virtual, 7–13 May 2022; pp. 441–445. [CrossRef]
152. Tsui, B. HRTF Generation for Data Demanding Machine Learning Algorithms. Ph.D. Thesis, University of York, York, UK, 2023.
153. Zotkin, D.Y.N.; Hwang, J.; Duraiswaini, R.; Davis, L.S. HRTF personalization using anthropometric measurements. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (Cat. No.03TH8684), New Paltz, NY, USA, 19–22 October 2003; IEEE: Piscataway, NJ, USA, 2003; pp. 157–160. [CrossRef]
154. Mohan, A.; Duraiswami, R.; Zotkin, D.N.; DeMenthon, D.; Davis, L.S. Using computer vision to generate customized spatial audio. In Proceedings of the International Conference on Multimedia and Expo, Baltimore, MD, USA, 6–9 July 2003; ICME'03. Proceedings (Cat. No. 03TH8698); IEEE: Piscataway, NJ, USA, 2003; Volume 3, pp. III–57. [CrossRef]

155. Zhu, M.; Shahnawaz, M.; Tubaro, S.; Sarti, A. HRTF personalization based on weighted sparse representation of anthropometric features. In Proceedings of the 2017 International Conference on 3D Immersion (IC3D), Brussels, Belgium, 11–12 December 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–7. [CrossRef]
156. Grindlay, G.; Vasilescu, M.A.O. A multilinear approach to HRTF personalization. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Honolulu, HI, USA, 15–20 April 2007. [CrossRef]
157. Hongmei, H.; Zhou, L.; Ma, H.; Wu, Z. HRTF personalization based on artificial neural network in individual virtual auditory space. *Appl. Acoust.* **2008**, *69*, 163–172. [CrossRef]
158. Li, L.; Huang, Q. HRTF personalization modeling based on RBF neural network. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 3707–3710. [CrossRef]
159. Bilinski, P.; Ahrens, J.; Thomas, M.R.P.; Tashev, I.J.; Platt, J.C. HRTF magnitude synthesis via sparse representation of anthropometric features. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 4468–4472. [CrossRef]
160. Chun, C.J.; Moon, J.M.; Lee, G.W.; Kim, N.K.; Kim, H.K. Deep neural network based HRTF personalization using anthropometric measurements. In Proceedings of the 143rd Audio Engineering Society Convention, New York, NY, USA, 18–21 October 2017.
161. Chen, T.Y.; Kuo, T.H.; Chi, T.S. Autoencoding HRTFS for DNN based HRTF personalization using anthropometric features. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 271–275. [CrossRef]
162. Miccini, R.; Spagnol, S. HRTF individualization using deep learning. In Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 390–395. [CrossRef]
163. Wang, Y.; Zhang, Y.; Duan, Z.; Bocko, M. Global HRTF personalization using anthropometric measures. In Proceedings of the Audio Engineering Society International Conference on Audio for Virtual and Augmented Reality, Online, 17–19 August 2020.
164. Spagnol, S.; Geronazzo, M.; Avanzini, F. On the relation between pinna reflection patterns and head-related transfer function features. *IEEE Trans. Audio Speech, Lang. Process.* **2013**, *21*, 508–519. [CrossRef]
165. Pirard, L. Spatial audio and individualized HRTFs using a convolutional neural network (CNN). *arXiv* **2023**, arXiv:2311.13397.
166. Fantini, D.; Avanzini, F.; Ntalampiras, S.; Presti, G. HRTF individualization based on anthropometric measurements extracted from 3D head meshes. In Proceedings of the Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 8–10 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–10. [CrossRef]
167. Zhou, Y.; Jiang, H.; Ithapu, V.K. On the predictability of HRTFs from ear shapes using deep networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 441–445. [CrossRef]
168. Wang, Y.; Zhang, Y.; Duan, Z.; Bocko, M. Predicting global head-related transfer functions from scanned head geometry using deep learning and compact representation. *arXiv* **2022**, arXiv:2207.14352. [CrossRef]
169. Ko, B.Y.; Lee, G.T.; Nam, H.; Park, Y.H. PRTFNet: HRTF individualization for accurate spectral cues Using a compact PRTF. *IEEE Access* **2023**, *11*, 96119–96130. [CrossRef]
170. Zhao, J.; Yao, D.; Gu, J.; Li, J. Efficient prediction of individual head-related transfer functions based on 3D meshes. *Appl. Acoust.* **2024**, *219*, 109938. [CrossRef]
171. Wenzel, E.M.; Foster, S.H. Perceptual consequences of interpolating head-related transfer functions during spatial synthesis. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Minneapolis, MN, USA, 27–30 April 1993; IEEE: Piscataway, NJ, USA, 1993; pp. 102–105. [CrossRef]
172. Bruschi, V.; Nobili, S.; Cecchi, S.; Piazza, F. An innovative method for binaural room impulse responses interpolation. In Proceedings of the 148th Audio Engineering Society Convention, Virtual, 2–5 June 2020.
173. Bau, D.; Himmelein, H.; Pörschmann, C. Comparison of nonparametric interpolation techniques for sparsely measured binaural room impulse responses. *J. Audio Eng. Soc.* **2024**, *72*, 479–492. [CrossRef]
174. Batteau, D.W. The role of the pinna in human localization. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **1967**, *168*, 158–180. [CrossRef]
175. Wright, D.; Hebrank, J.H.; Wilson, B. Pinna reflections as cues for localization. *J. Acoust. Soc. Am.* **1974**, *56*, 957–962. [CrossRef] [PubMed]
176. Watkins, A.J. The monaural perception of azimuth: A synthesis approach. In *Localization of Sound: Theory and Applications*; Amphora Press: Groton, CT, USA, 1982; pp. 194–206.
177. Genuit, K. A description of the human outer ear transfer function by elements of communication theory. In Proceedings of the 12th International Congress on Acoustics, Tokyo, Japan, 7–11 April 1986; p. B6-8.
178. Chen, J.; Van Veen, B.D.; Hecox, K.E. External ear transfer function modeling: A beamforming approach. *J. Acoust. Soc. Am.* **1992**, *92*, 1933–1944. [CrossRef] [PubMed]
179. Brown, C.P.; Duda, R.O. An efficient HRTF model for 3-D sound. In Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 19–22 October 1997; IEEE: New Paltz, NY, USA 1997; p. 4. [CrossRef]
180. Brown, C.P.; Duda, R.O. A structural model for binaural sound synthesis. *IEEE Trans. Speech Audio Process.* **1998**, *6*, 476–488. [CrossRef]

181. Lu, J.; Qi, X. Pre-trained-based individualization model for real-time spatial audio rendering system. *IEEE Access* **2021**, *9*, 128722–128733. [CrossRef]
182. Ito, Y.; Nakamura, T.; Koyama, S.; Saruwatari, H. Head-related transfer function interpolation from spatially sparse measurements using autoencoder with source position conditioning. In Proceedings of the 2022 International Workshop on Acoustic Signal Enhancement (IWAENC), Bamberg, Germany, 5–8 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5. [CrossRef]
183. Savioja, L.; Huopaniemi, J.; Lokki, T.; Väänänen, R. Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.* **1999**, *47*, 675–705.
184. Freeland, F.P.; Biscainho, L.W.P.; Diniz, P.S.R. Efficient HRTF interpolation in 3D moving sound. In Proceedings of the 22nd Audio Engineering Society International Conference on Virtual, Synthetic, and Entertainment Audio, Espoo, Finland, 15–17 June 2002.
185. Kistler, D.J.; Wightman, F.L. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.* **1992**, *91*, 1637–1647. [CrossRef]
186. Martens, W.L. *Principal Components Analysis and Resynthesis of Spectral Cues to Perceived Direction*; University of Michigan Library: Ann Arbor, MI, USA, 1987.
187. Middlebrooks, J.C.; Green, D.M. Observations on a principal components analysis of head-related transfer functions. *J. Acoust. Soc. Am.* **1992**, *92*, 597–599. [CrossRef]
188. Siripornpitak, P.; Engel, I.; Squires, I.; Cooper, S.J.; Picinali, L. Spatial up-sampling of HRTF sets using generative adversarial networks: A pilot study. *Front. Signal Process.* **2022**, *2*, 904398. [CrossRef]
189. Kulkarni, A.; Colburn, H.S. Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. Am.* **2004**, *115*, 1714–1728. [CrossRef]
190. Queiroz, M.; De Sousa, G.H.M. Structured IIR models for HRTF interpolation. In Proceedings of the International Computer Music Conference, New York, NY, USA, 1–5 June 2010.
191. Pulkki, V. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.* **1997**, *45*, 456–466.
192. Chen, J.; Van Veen, B.D.; Hecox, K.E. A spatial feature extraction and regularization model for the head-related transfer function. *J. Acoust. Soc. Am.* **1995**, *97*, 439–452. [CrossRef] [PubMed]
193. Jiang, Z.; Sang, J.; Zheng, C.; Li, A.; Li, X. Modeling individual head-related transfer functions from sparse measurements using a convolutional neural network. *J. Acoust. Soc. Am.* **2023**, *153*, 248–259. [CrossRef] [PubMed]
194. Zurale, D.; Dubnov, S. Spatial upsampling of sparse head related transfer functions—A VQ-VAE & Transformer based approach. In Proceedings of the Audio Engineering Society International Conference on Audio for Virtual and Augmented Reality, Redmond, WV, USA, 15–17 August 2022.
195. Hogg, A.O.T.; Jenkins, M.; Liu, H.; Squires, I.; Cooper, S.J.; Picinali, L. HRTF upsampling with a generative adversarial network using a gnomonic equiangular projection. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2024**, *32*, 2085–2099. [CrossRef]
196. Bruschi, V.; Dourou, N.; Nobili, S.; Cecchi, S. A new method for HRTFs interpolation based on inverse distance weighting. In Proceedings of the 154th Audio Engineering Society Convention, Espoo, Finland, 13–15 May 2023.
197. Evans, M.J.; Angus, J.A.S.; Tew, A.I. Analyzing head-related transfer function measurements using surface spherical harmonics. *J. Acoust. Soc. Am.* **1998**, *104*, 2400–2411. [CrossRef]
198. Duraiswami, R.; Zotkin, D.N.; Gumerov, N.A. Interpolation and range extrapolation of HRTFs [head related transfer functions]. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, USA, 17–21 May 2004; IEEE: Piscataway, NJ, USA, 2004; Volume 4, p. iv. [CrossRef]
199. Zhang, W.; Kennedy, R.A.; Abhayapala, T.D. Iterative extrapolation algorithm for data reconstruction over sphere. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 3733–3736. [CrossRef]
200. Zotkin, D.N.; Duraiswami, R.; Gumerov, N.A. Regularized HRTF fitting using spherical harmonics. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 18–21 October 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 257–260. [CrossRef]
201. Aussal, M.; Alouges, F.; Katz, B. A study of spherical harmonics interpolation for HRTF exchange. *Proc. Meet. Acoust.* **2013**, *19*, 050010. [CrossRef]
202. Brinkmann, F.; Weinzierl, S. Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition. In Proceedings of the Audio Engineering Society International Conference on Audio for Virtual and Augmented Reality, Redmod, WA, USA, 20–22 August 2018.
203. Pörschmann, C.; Arend, J.M.; Bau, D.; Lübeck, T. Comparison of spherical harmonics and nearest-neighbor based interpolation of head-related transfer functions. In Proceedings of the Audio Engineering Society International Conference on Audio for Virtual and Augmented Reality, Online, 17–19 August 2020.
204. Skarha, M. *Performance Tradeoffs in HRTF Interpolation Algorithms for Object-Based Binaural Audio*; McGill University (Canada): Montreal, QC, Canada, 2021.
205. Arend, J.M.; Brinkmann, F.; Pörschmann, C. Assessing spherical harmonics interpolation of time-aligned head-related transfer functions. *J. Audio Eng. Soc.* **2021**, *69*, 104–117. [CrossRef]

206. Chen, X.; Ma, F.; Zhang, Y.; Bastine, A.; Samarasinghe, P.N. Head-related transfer function interpolation with a spherical CNN. *arXiv* **2023**, arXiv:2309.08290.
207. Thuillier, E.; Jin, C.; Välimäki, V. HRTF interpolation using a spherical neural process meta-learner. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2024**, *32*, 1790–1802. [CrossRef]
208. Kearney, G.; Masterson, C.; Adams, S.; Boland, F. Dynamic time warping for acoustic response interpolation: Possibilities and limitations. In Proceedings of the 17th European Signal Processing Conference (EUSIPCO), Glasgow, UK, 24–28 August 2009; IEEE: New York, NY, USA, 2009; pp. 705–709.
209. Masterson, C.; Kearney, G.; Boland, F. Acoustic impulse response interpolation for multichannel systems using dynamic time warping. In Proceedings of the 35th Audio Engineering Society International Conference on Audio for Games, London, UK, 11–13 February 2009.
210. Wahba, G. Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Stat. Comput.* **1981**, *2*, 5–16. [CrossRef]
211. Freeland, F.P.; Biscainho, L.W.P.; Diniz, P.S.R. HRTF interpolation through direct angular parameterization. In Proceedings of the IEEE International Symposium on Circuits and Systems, New Orleans, LA, USA, 27–20 May 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1823–1826. [CrossRef]
212. Farrow, C.W. A continuously variable digital delay element. In Proceedings of the IEEE International Symposium on Circuits and Systems, Espoo, Finland, 7–9 June 1988; IEEE: Piscataway, NJ, USA, 1988; pp. 2641–2645. [CrossRef]
213. Gumerov, N.A.; Duraiswami, R. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*; Elsevier: Amsterdam, The Netherlands, 2005; pp. 49–55. [CrossRef]
214. Cobos, M.; Ahrens, J.; Kowalczyk, K.; Politis, A. An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction. *EURASIP J. Audio, Speech, Music Process.* **2022**, *2022*, 10. [CrossRef]
215. Liu, W. Generating Personalized Head-Related Transfer Function (HRTF) Using Scanned Mesh from iPhone FaceID. Master's Thesis, Chalmers University of Technology, Gothenburg, Sweden, 2023.
216. Binelli, M.; Pinardi, D.; Nili, T.; Farina, A. Individualized HRTF for playing VR videos with Ambisonics spatial audio on HMDs. In Proceedings of the Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality, Redmod, WA, USA, 20–22 August 2018.
217. Tomasetti, M.; Boem, A.; Turchet, L. How to spatial audio with the WebXR API: A comparison of the tools and techniques for creating immersive sonic experiences on the browser. In Proceedings of the Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 5–7 September 2023; pp. 1–9. [CrossRef]
218. Hiroaki, S. Introduction. In *Head-Related Transfer Function and Virtual Auditory Display*; Springer: Singapore, 2016; pp. 1–5. [CrossRef]
219. Andersen, J.S.; Miccini, R.; Serafin, S.; Spagnol, S. Evaluation of individualized HRTFs in a 3D shooter game. In Proceedings of the Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 8–10 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–10. [CrossRef]
220. López, J.J.; Gutierrez-Parera, P. Evaluation of the frequency response of consumer headphones and its influence towards a binaural reproduction with HRTF individualization. In Proceedings of the 22nd International Congress on Acoustics, Buenos Aires, Argentina, 5–9 September 2016.
221. Warusfel, O. Listen HRTF Database. 2003. Available online: http://recherche.ircam.fr/equipes/salles/listen/ (accessed on 24 November 2024).
222. Qu, T.; Xiao, Z.; Gong, M.; Huang, Y.; Li, X.; Wu, X. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. *IEEE Trans. Audio, Speech Lang. Process.* **2009**, *17*, 1124–1132. [CrossRef]
223. Gupta, N.; Barreto, A.; Joshi, M.; Agudelo, J.C. HRTF database at FIU DSP lab. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, TX, USA, 15–19 March 2010; pp. 169–172. [CrossRef]
224. Jin, C.T.; Guillon, P.; Epain, N.; Zolfaghari, R.; Van Schaik, A.; Tew, A.I.; Hetherington, C.; Thorpe, J. Creating the Sydney York morphological and acoustic recordings of ears database. *IEEE Trans. Multimed.* **2013**, *16*, 37–46. [CrossRef]
225. Rothbucher, M.; Paukner, P.; Stimpfl, M.; Diepold, K. *The TUM-LDV HRTF Database*; Technical Report; Lehrstuhl für Datenverarbeitung: Munchen, Germany, 2014.
226. Watanabe, K.; Iwaya, Y.; Suzuki, Y.; Takane, S.; Sato, S. Dataset of head-related transfer functions measured with a circular loudspeaker array. *Acoust. Sci. Technol.* **2014**, *35*, 159–165. [CrossRef]
227. Rugeles Ospina, F.; Emerit, M.; Katz, B.F.G. The three-dimensional morphological database for spatial hearing research of the BiLi project. *Proc. Meet. Acoust.* **2015**, *23*, 050001. [CrossRef]
228. Kearney, G.; Doyle, T. An HRTF database for virtual loudspeaker rendering. In Proceedings of the 139th Audio Engineering Society Convention, New York, NY, USA, 29 October–1 November 2015.
229. Bomhardt, R.; de la Fuente Klein, M.; Fels, J. A high-resolution head-related transfer function and three-dimensional ear model database. *Proc. Meet. Acoust.* **2016**, *29*, 050002. [CrossRef]
230. Spagnol, S.; Purkhús, K.B.; Unnthórsson, R.; Bjornsson, S. The viking HRTF dataset. In Proceedings of the 16th Sound & Music Computing Conference, Malaga, Spain, 28–31 May 2019; pp. 55–60.

231. Suzuki, Y.; Asano, F.; Kim, H.Y.; Sone, T. An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses. *J. Acoust. Soc. Am.* **1995**, *97*, 1119–1123. [CrossRef]
232. Lindau, A.; Hohn, T.; Weinzierl, S. Binaural resynthesis for comparative studies of acoustical environments. In Proceedings of the 122nd Audio Engineering Society Convention, Vienna, Austria, 5–8 May 2007.