Research Paper

Continuous Head-related Transfer Function Representation Based on Hyperspherical Harmonics

Adam SZWAJCOWSKI

Department of Robotics and Mechatronics
AGH University of Science and Technology
Kraków, Poland; e-mail: szwajcowski@agh.edu.pl

Expressing head-related transfer functions (HRTFs) in the spherical harmonic (SH) domain has been thoroughly studied as a method of obtaining continuity over space. However, HRTFs are functions not only of direction but also of frequency. This paper presents an extension of the SH-based method, utilizing hyperspherical harmonics (HSHs) to obtain an HRTF representation that is continuous over both space and frequency. The application of the HSH approximation results in a relatively small set of coefficients which can be decoded into HRTF values at any direction and frequency. The paper discusses results obtained by applying the method to magnitude spectra extracted from exemplary HRTF measurements. The HRTF representations based on SHs and HSHs exhibit similar reproduction accuracy, with the latter one featuring continuity over both space and frequency and requiring much lower number of coefficients. The developed HSH-based continuous functional model can serve multiple purposes, such as interpolation, compression or parametrization for machine-learning applications.

Keywords: hyperspherical harmonics; HRTF; spherical harmonics; continuous functional model; directivity.

1. Introduction

Continuous development of virtual and augmented reality applications rises the need for efficient binaural audio processing algorithms. Especially an important role in authentic recreation of an auditory scene is played by the directivity of human ears, which is commonly called head-related transfer functions (HRTFs). HRTFs are different for each individual as they depend on the shape of torso, head, and pinna. Even though there are some similarities among sets of HRTFs obtained for different people, application of individually measured or matched HRTFs has been proven to improve the localization abilities (Wenzel *et al.*, 1993; Begault *et al.*, 2001). The importance of HRTFs is especially prominent in recognizing the position of a sound source placed in sagittal planes, where simple binaural cues such as interaural level difference (ILD) and interaural time difference (ITD) are mostly invariant (Macpherson, Middlebrooks, 2002; Agterberg *et al.*, 2012).

HRTFs can be physically measured in two ways. Most measurements are performed by putting a pair of microphones inside the subject's ears and recording the response to the sound coming from different directions (e.g., (Andreopoulou, 2015a; Zhang, 2012)). Alternatively, one can make use of the Helmholtz reciprocity principle and place microspeakers inside the ears while the microphones are set around the subject (Zotkin, 2006). Either way, the results of the measurements are sets of HRTFs for a finite number of directions. They are usually stored in the form of discrete head-related impulse responses (HRIRs), for example using the Spatially Oriented Format for Acoustics (SOFA) (AES, 2015; Majdak *et al.*, 2013).

Since HRTFs are relatively large data sets, some attempts were made to develop an efficient model that would reduce the amount of data without a significant loss of accuracy. Initially, the research concerned approximating HRIRs, e.g., by expressing them as filters of either finite or infinite impulse response (e.g., (Kulkarni, Colburn, 1995; 2004)), but soon

the focus was moved to the spatial properties. Evans *et al.* (1998) proposed expressing HRTFs by means of spherical harmonics (SHs). Beside reducing the data size, this representation introduced an even more significant feature – continuity over space. Over the years, the SH-based method has been widely investigated, regarding, i.e., efficiency of different sampling schemes (Zhang *et al.*, 2012), preprocessing techniques (Brinkmann, Weinzierl, 2018) or mixed-order approximations (Ben-Hur *et al.*, 2019). Some alternative continuous representations were suggested, e.g., based on spherical wavelets (Hu *et al.*, 2019; Liu *et al.*, 2019) or Slepian functions (Bates *et al.*, 2015); however, SHs still prevail as the most popular basis functions to approximate not only HRTFs but also other directivity functions such as the sound source directivity (Szwajcowski *et al.*, 2021; Shabtai *et al.*, 2017) or the microphone directivity (Ziegler *et al.*, 2017).

Beside focusing on either only spatial or only time/frequency variations, several attempts were made to develop a model including both these dependencies. Kistler and Wightman (1992) employed the principal component analysis to an HRTF database. While resulting model indeed provided a good accuracy, it was discrete and required a priori knowledge of the database to determine optimal basis vectors. Chen *et al.* (1995) developed a representation based on thin-plate splines that was continuous in space but still discrete in frequency. In 2009, Zhang *et al.* described a model based on Fourier series and Bessel functions which was continuous over both frequency and horizontal angles (Zhang *et al.*, 2009) and which was later extended to cover the entire sphere and include distance dependence as well (Zhang *et al.*, 2010). Another proposition comes from authors who based their model on the fusion of infinite impulse response filters and Legendre polynomials (Shekarchi *et al.*, 2013). However, the authors focused purely on the compression of measured data and thus their model cannot be used to interpolate missing values. Zhang *et al.* (2015) created a functional model utilizing SHs for spatial dependency and complex exponentials for the frequency representation. Both mentioned fully continuous models (Zhang *et al.*, 2010; 2015) determined coefficients for basis functions partially by numerical integrating, which is computationally efficient, but provides worse accuracy than that obtained by fitting in the least-squares sense. Furthermore, these researches focused on retrieving high accuracy complex representations, even though it is known that humans are insensitive to fine spectral or spatial details of HRTFs (Kulkarni, Colburn, 1998; Romigh *et al.*, 2015; Breebaart, Kohlrausch, 2001; Xie, Zhang, 2010).

In this paper, a new approach to the HRTF dimensionality is presented: frequency is imagined as another spatial dimension. This way, the character of these functions becomes purely spatial with two truly spatial variables (horizontal and vertical angles) and one extra variable that represents frequency, but is also treated as a spatial dependence[1]. This approach emphasizes coupling between frequency and space in directivity functions and is inspired by the mathematical structure known as Minkowski space, where time and three-dimensional (3D) space are combined together to conveniently express some of the physical phenomena, most notably regarding the theory of special relativity. Assuming that SHs are a good choice for expressing directivity functions in the 3D space, including the fourth dimension would require an extention of the basis to a four-dimensional (4D) space. A similar problem was tackled in computer graphics, where standard 3D shape descriptors were insufficient for certain cases. Bonvallet *et al.* (2007) extended the popular approach to another dimension by replacing SHs with hyperspherical harmonics (HSHs). This method was later applied in medical imaging by Pasha Hosseinbor *et al.* (20015), providing further a proof that HSHs, previously reserved for theoretical chemistry and nuclear physics, can be successfully employed in engineering. However, to the best of our knowledge, HSHs have not yet been utilized in acoustics or any related field. Thus, the application of this basis to represent acoustical data is the main novel element of this paper.

The primary advantage of the proposed HSH representation over the state-of-the-art SH one is continuity over both space and frequency. Such a representation allows to extract the HRTF magnitude not only at a given direction but also at a given frequency without any additional operations (e.g., interpolation or resampling), thus being computationally attractive. Furthermore, varying the approximation parameters enables easy control of the balance between accuracy and amount of data; acknowledging psychoacoustical aspects in the process of deriving the functional model can lead to a significant reduction of data size by ignoring high-order HSHs responsible for imperceptible spectral details. Last but not least, a more holistic HRTF representation can be of great value in research requiring a thorough directivity parametrization method, e.g., for machine learning applications.

Section 2 provides the necessary theoretical background on 4D coordinate systems and HSHs. Section 3 describes conversion from raw measurement data to hyperspherical data and then to the HSH domain. Section 4 presents exemplary results of the conversion performed on a typical set of HRTFs and suggests further improvements of the process. Section 5 consists of general comments on the HSH representation, its accuracy and potential applications. Finally, Sec. 6 summarizes all the content of this paper.

---

[1]This paper concerns far-field HRTFs, where the radius dependence is neglected. This is why three- and four-dimensional spaces are described by only two and three variables, respectively.

2. Theoretical background

HRTFs are functions of direction, distance, and frequency. Under the assumption that the distance is greater than 1 m, the radial dependence can be dropped. Such simplified functions are called far-field HRTFs, but since they are measured and applied more commonly than full, distance-dependent HRTFs, they are often called just HRTFs. This is also the case in this paper; wherever HRTFs are mentioned, they mean far-field HRTFs, independent of distance.

2.1. Hyperspherical coordinate system

Assuming that the spatial dependence of HRTFs is to be described in the spherical coordinate system, an extension of this system is needed to capture variability over another dimension. This can be achieved by adding another linear or angular dimension, resulting in either spherindrical or 4D hyperspherical coordinate system (HCS), respectively. The HCS has already proven to be successful in similar research problems in other fields (BONVALLET *et al.*, 2007; PASHA HOSSEINBOR *et al.*, 2015) and its geometrical properties enhance natural physical properties of acoustic directivity characteristics (see Subsec. 2.2.1 and 3.1.1 for explanation). For these reasons, HCS was chosen over the spherindrical coordinate system for this research.

HCS consists of hyperspherical radius ρ and three angles: $\varphi \in [0, 2\pi)$, $\theta \in [0, \pi)$, and $\psi \in [0, \pi)$. The angles φ and θ correspond to azimuth and inclination defined in the spherical coordinate system, while ψ is an extra angle representing another dimension. Since there are no unified names for angles in 4D space, within this paper they are referred to just by their symbols or colloquially as the spatial angles (meaning angles φ and θ) and the frequency angle (meaning angle ψ) (the details of utilizing an angle to describe the frequency are provided in Subsec. 3.1.1). The relation between HCS and 4D Cartesian system (x, y, z, w) is following:

$$
\begin{aligned}
x &= \rho \sin \psi \sin \theta \sin \varphi, \\
y &= \rho \sin \psi \sin \theta \cos \varphi, \\
z &= \rho \sin \psi \cos \theta, \\
w &= \rho \cos \psi.
\end{aligned}
\tag{1}
$$

2.2. Hyperspherical harmonics definition

HSHs can be defined for any multidimensional space, but within this paper only 4D HSHs are considered and simply referred to the HSHs. They can be defined as following (DOMOKOS, 1967):

$$
Z_{nl}^m(\varphi, \theta, \psi) \equiv N(n,l) \sin^l \psi \, C_{n-l}^{l+1}(\cos \psi) \, Y_l^m(\varphi, \theta), \tag{2}
$$

where $N(n,l)$ is the normalization factor, $C_\nu^\alpha(x)$ are the Gegenbauer polynomials, and $Y_l^m(\varphi, \theta)$ are the

SHs, while n, l, and m are the integer parameters limited as following:

$$
\begin{aligned}
n &\geq 0, \\
0 &\leq l \leq n, \\
-l &\leq m \leq l.
\end{aligned}
\tag{3}
$$

$N(n,l)$ is the normalization factor, making the HSH basis not only orthogonal but orthonormal. It is given by the formula:

$$
N(n,l) \equiv 2^{l+\frac{1}{2}}(l+1)! \sqrt{\frac{2(n+1)(n-l+1)!}{\pi(n+l+2)!}}. \tag{4}
$$

It is worth noting, that the SHs in Eq. (2) are also normalized, which is why $N(n,l)$ does not depend on the parameter m – this dependency is entirely included in the SH normalization factor.

$C_\nu^\alpha(x)$ are the Gegenbauer polynomials, also known as ultraspherical polynomials. They can be defined in multiple equivalent ways; however, for the computations, the most useful one is that given by the following recurrence relation (Ultraspherical polynomials, n.d.):

$$
\begin{aligned}
C_0^\alpha(x) &= 1, \\
C_1^\alpha(x) &= 2\alpha x, \\
C_\nu^\alpha(x) &= \frac{1}{\nu}\big(2x(\nu + \alpha - 1)C_{\nu-1}^\alpha(x) \\
&\quad - (\nu + 2\alpha - 2)C_{\nu-2}^\alpha(x)\big).
\end{aligned}
\tag{5}
$$

In the HSH definition (Eq. (2)), the Gegenbauer polynomials, together with the factor of $\sin^l \psi$, are responsible for variation among ψ angle (the frequency angle).

$Y_l^m(\varphi, \theta)$ are SHs. SHs can be defined either as complex- or real-valued functions. This paper focuses on their application to express magnitude spectra of HRTFs (see Subsec. 3.1 for explanation) and thus the real form is used, making the entire HSH basis real as well. Real SHs are defined as following[2] (VARSHALOVICH *et al.*, 1998):

$$
Y_l^m(\phi, \theta) \equiv \begin{cases} N_Y(l,m) P_l^m(\cos\theta) \cos(m\varphi), \\ \qquad\qquad\qquad \text{if} \quad m \geq 0, \\ N_Y(l,m) P_l^{|m|}(\cos\theta) \sin(|m|\varphi), \\ \qquad\qquad\qquad \text{if} \quad m < 0, \end{cases}
\tag{6}
$$

where P_l^m are the associated Legendre functions and $N_Y(l,m)$ is the normalization factor for SHs defined as:

$$
N_Y(l,m) \equiv \sqrt{(2 - \delta_{m0})\frac{2l+1}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}}, \tag{7}
$$

where δ is the Kronecker delta.

---

[2]Sometimes in real SH definitions there is also $(-1)^m$ factor included, but it is irrelevant for approximation purposes as it merely changes the sign of some coefficients.

2.2.1. Hyperspherical poles

Hyperspheres, by being extensions of 2-spheres (regular 3D spheres), display some analogous features. On a 2-sphere, there are two poles lying at inclinations $\theta \in \{0, \pi\}$, at which all meridians converge to the same point. In other words, at these poles, the actual direction is the same for every azimuthal angle φ. In HSC, such poles exist for every ψ, but there are also two major poles (hyperpoles) at $\psi \in \{0, \pi\}$, where the direction is dependent on neither φ nor θ. Now, applying the interpretation that ψ is the frequency angle, this means that values for the limit frequencies have to be constant for every physical direction. This is inline with how the directivity (including HRTFs) behaves at low frequency – the lower the frequency, the less variance along the directions can be observed, as the size and shape of objects become less and less relevant when compared to the length of acoustic waves corresponding to the analyzed frequency. This coincidence is used in frequency-to-angle mapping, so that the hyperpole convergence can be perceived as an advantage rather than a limitation of the proposed model (see Subsec. 3.1.1).

The presence of hyperpoles can also be seen as a part of general property of HSHs, which are meant to represent the hypersphere boundary evenly. Since the angular distance between two points on the hypersphere gets lower as the hyperpoles are approached (in the same vein as meridians get closer near the poles on a sphere), the ability to capture fine HRTF details decreases. For SHs, the same effect is often achieved by decreasing the maximum order for lower frequencies (e.g., (Zhang *et al.*, 2015; Li *et al.*, 2021)). This is not feasible when continuous representations are used; however, HSHs provide the same effect in a different way, which further shows that they fit well for handling directivity data.

3. Hyperspherical harmonic approximation

In order to express HRTFs in the HSH domain, two steps have to be taken: first, the measurement data have to be converted to HCS and then the HSH coefficients need to be determined so that the weighted sum of HSHs approximates the data in HCS as accurately as possible. Both these steps are described in details in the following subsections.

3.1. HRIRs to data in HCS

As stated in Introduction, HRTFs are commonly stored in the form of HRIRs. It might seem natural to try to find a way to efficiently represent the data in the time domain rather than in the frequency domain. However, there are several arguments in favour of the latter. Firstly, in previous research, both approaches were studied and models based on the frequency representation resulted in lower approximation errors using objective measures (Evans *et al.*, 1998; Hartung *et al.*, 1999). Secondly, the human auditory system, analyzes sounds mostly in the frequency domain and thus accurate reproduction of the impulse response shape is irrelevant to our ears, as opposed to the shape of the frequency spectrum. Furthermore, HRIRs usually have much sharper shapes than the corresponding HRTFs, which makes representing them as a sum of basis functions much harder, especially when the basis is truncated. Last but not least, HRIRs include information on phase, which is of little importance; it is widely acknowledged that phase spectra can be ignored as long as the interaural time difference is preserved (Kistler, Wightman, 1992; Romigh *et al.*, 2015; Kulkarni *et al.*, 1999)[3], although some contrary results have also been presented (Rasumow *et al.*, 2014). Basing a model on the frequency-domain representation thus allows the information on phase to be dropped, reducing the efficient amount of data twice without a significant loss as far as the applicatory aspect is concerned. However, HSHs can be defined as complex functions as well, so it is possible to include the phase information as well, if desired.

The magnitude spectra typically are represented either in linear or in logarithmic scale (in dBs). Referring again to the psychoacoustic aspects of sound perception, the logarithmic scale is relative and thus it is more reflective of how the sound pressure is perceived by the human auditory system. Since the approximation method of the model proposed within this paper is based on the least-squares fitting (see Subsec. 3.2.2 for more details), it seems reasonable to use a logarithmic scale, so that the solver minimizes relative errors (Kulkarni, Colburn, 2004; Romigh *et al.*, 2015; Hartung *et al.*, 1999; Blommer, Wakefield, 1997).

The reference value for the logarithmic scale does not matter as long as it is uniform within a given HRTF set; this value is linked only with the coefficient for Z_{00}^0 (the first HSH invariant along all the angles), but otherwise has no impact whatsoever. Therefore, to make it simple, the reference value was chosen to be 1.

3.1.1. Frequency mapping

As signaled earlier, in order to express HRTFs in HCS, frequency has to be mapped to the ψ angle. The simplest way of doing it would be to set the frequency $f = 0$ at $\psi = 0$ and the frequency $f = f_s/2$ to $\psi = \pi$, where f_s denotes the sampling frequency of HRIRs in the HRTF set. The remaining frequencies would be then mapped to ψ linearly, following the formula:

---

[3]Phase spectra are still important for low frequencies, but they can be replaced by a linear phase derived from interaural time differences, without any noticeable damage to the localization abilities.

$$\psi_k = \frac{2\pi f_k}{f_s}, \qquad (8)$$

where f_k is the center frequency of the k-th frequency bin and ψ_k is the corresponding value of the frequency angle. However, such mapping implies convergence to uniform values at hyperpoles, i.e., at $f \in \{0, f_s/2\}$. When approaching 0 Hz, HRTFs tend to be omnidirectional anyway, but this is not the case for the highest frequencies, especially in the logarithmic scale (see Fig. 1). While this might seem to be of little importance, as the convergence to the hyperspherical pole at the Nyquist frequency would likely affect mostly the highest frequencies which are inaudible anyway, it is possible to avoid such distortions; $f = f_s/2$ can be set to $\psi = \pi/2$ instead of $\psi = \pi$, so that all the HRTFs can be mapped to effectively only half a hypersphere. This requires twice as large resolution of HSH along ψ since the spectra are squeezed to fit on only half of the full range of the frequency angle; however, such mapping allows to ignore all HSHs that are not symmetric about the hyperplane at $\psi = \pi/2$. The application of only ψ-symmetric HSHs (the HSHs that exhibit the aforementioned symmetry about $\psi = \pi/2$) removes the convergence at $f = f_s/2$ and makes the spectra for the entire range of $\psi \in [0, \pi)$ symmetric about the Nyquist frequency in the same manner as typical magnitude spectra obtained by performing the discrete Fourier transform on a real signal. With $f = f_s/2$ set to $\psi = \pi/2$, Eq. (8) takes the form of:

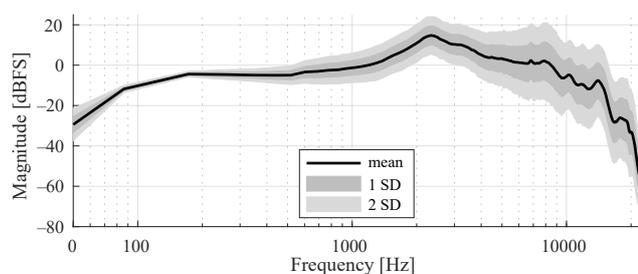$$\psi_k = \frac{\pi f_k}{f_s}, \qquad (9)$$

Fig. 1. Statistical distribution of an exemplary set of HRTFs (described in details in Sec. 4) for all directions. Solid black line is the mean spectrum, while the dark and light grey areas cover regions of 1 and 2 standard deviations, respectively. First frequency bin was shifted from 0 to 50 Hz, to be able to fit on the logarithmic axis. At the last frequency bin, standard deviation is especially large due to the occurrence of zero points that were converted to floating-point accuracy in MATLAB to avoid $-\infty$ values. The plot was shifted to fluctuate around 0 dB.

Alternatively, different frequency mappings could be used. Both limit values and the character of scale can be changed, e.g., to a logarithmic frequency scale, which is quite popular in audio engineering, or the mel scale, which is based on a psychoacoustic model of frequency perception. This paper concerns only the linear frequency mapping and leaves the employment of other scales for potential future research.

3.2. Data in HCS to HSH coefficients

Once the 4D data are prepared in the form of a set of triplets of angles (φ, θ, ψ) and the corresponding magnitudes, these data need to be approximated by a finite number of HSHs. The approximation comes down to determining the coefficients $\alpha_{n,l}^m$ for the weighted sum so that approximated function $\widehat{H}(\varphi, \theta, \psi)$ of the form[4]:

$$\widehat{H}(\varphi, \theta, \psi) = \sum_{n=0}^{n_{\max}} \sum_{l=0}^{n} \sum_{m=-l}^{l} \alpha_{nl}^m Z_{nl}^m(\varphi, \theta, \psi) \qquad (10)$$

is as close to the original function $H(\varphi, \theta, \psi)$ as possible. In the analytical approach, when an infinite number of basis functions can be used ($n_{\max} = \infty$), consecutive coefficients can be determined by calculating the dot product of the approximated function and complex conjugate of the given HSH:

$$\alpha_{n,l}^m = \int_{\mathbb{S}^3} H(\varphi, \theta, \psi) \overline{Z_{nl}^m}(\varphi, \theta, \psi) \, \mathrm{d}\chi, \qquad (11)$$

where $\int_{\mathbb{S}^3} \mathrm{d}\chi$ means integrating over the surface of the unit 3-sphere:

$$\int_{\mathbb{S}^3} \mathrm{d}\chi = \int_0^{2\pi} \int_0^{\pi} \int_0^{\pi} \sin^2 \psi \, \mathrm{d}\psi \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi. \qquad (12)$$

However, when the number of basis functions is truncated, the accuracy of the approximation decreases. For this reason, least-squares fitting of the coefficient values is preferred over the numerical integration of Eq. (11) to find the optimal set of coefficients.

3.2.1. Determination of the number of HSHs used

The choice of the number of basis functions is one of the most important decisions affecting the quality of the approximation. In general, increasing this number improves the overall accuracy of the approximation. However, using too many basis functions requires more computational resources and can lead to overfitting. It is thus important to determine what is the optimal number of basis functions for a given problem.

HSHs are described by three parameters: n, l, and m. Analyzing the definition given in Eq. (2), one can notice that these parameters are responsible for variations along the angles ψ, θ, and φ, respectively. Limiting just n results in the same spatial frequency

---

[4]Equation (10) assumes approximation using all the HSHs up to $n = n_{\max}$. However, all of the parameters n, l and m can have independent limits (see Subsec. 3.2.1).

along all the angles. However, some extra limitations for l and m can be imposed to obtain different resolution along angles θ and φ. Equation (10) then takes the form of:

$$\widehat{H}(\varphi,\theta,\psi) = \sum_{n=0}^{n_{\max}} \sum_{l=0}^{\min(n,l_{\max})} \sum_{m=-\min(l,m_{\max})}^{\min(l,m_{\max})}$$
$$\cdot \alpha_{nl}^{m} Z_{nl}^{m}(\varphi,\theta,\psi), \tag{13}$$

where the limits have to satisfy the following relation[5]:

$$n_{\max} \geq l_{\max} \geq m_{\max}. \tag{14}$$

These maximum values of $n_{\max}$, $l_{\max}$, and $m_{\max}$ should be determined with respect to the sampling theorem and depend on both spatial and frequency sampling. However, HRTFs do not need to be reconstructed with the highest available accuracy, since fine details are psychoacoustically irrelevant. Thus, it is important to find the minimum values of the limits for which the approximation error is negligible.

The spatial dependence is described by SHs and thus the limits for spatial frequency ($l_{\max}$ and $m_{\max}$) can be determined basing on SH-related research. ROMIGH *et al.* (2015) noticed that some of the localization abilities occur for HRTFs approximated by SHs of the maximum order as low as 2. Approximations using SHs of the maximum order 6 are already statistically indistinguishable from the unprocessed HRTFs and the localization does not improve with further increasing the maximum order. It is important to notice that ROMIGH *et al.* (2015) used the same preprocessing as proposed within this paper, i.e., the least-squares fitting based on magnitudes in a logarithmic scale. These results suggest that setting $l_{\max}$ and $m_{\max}$ to 6 should be sufficient to prevail the localization properties of a set of HRTFs.

Similar research was conducted regarding sensitivity to spectral details of HRTFs. KULKARNI and COLBURN (1998) checked how much smoothing in frequency can be applied to HRTFs without a significant impact on the localization abilities. For this reason, they prepared stimuli by truncating the Fourier series representing the logarithmic spectra of empirical HRTFs. The listeners that took part in the experiment were unable to discriminate between the real and smoothed virtual sound sources in most of the setups for HRTFs reconstructed from only 16 coefficients. For 32 coefficients and above, all of the listeners performed at chance for all the tested directions. Even though the frequency angle in HSHs is not represented by the Fourier series, the same spatial frequency along that angle can be chosen by setting appropriate $n_{\max}$. Since

for chosen frequency mapping the spectra are supposed to fit on only half the hypersphere, $n_{\max}$ must be doubled to exhibit a desired resolution over the effectively used range of ψ. Thus, to match the spatial frequency of 32 spectrum coefficients from the described experiment, $n_{\max}$ should be set to 64.

Furthermore, as stated in Subsec. 3.1.1, effectively utilizing only half of the hypersphere allows to ignore non-ψ-symmetric HSHs. Following the HSH and Gegenbauer polynomials definitions given in Eqs. (2) and (5), it can be noted that for Z_{nl}^{m} to be ψ-symmetric, the difference between n and l has to be even. All the parameter configurations where this difference is odd should be thus disregarded when summing over l in Eq. (13).

3.2.2. Computations

Assuming sampling at K 4D directions $\Omega_k \equiv (\varphi_k, \theta_k, \psi_k)$, the HSH coefficients can be determined by solving the following matrix equation:

$$\begin{bmatrix} Z_{00}^{0}(\Omega_1) & \dots & Z_{n_{\max}l_{\max}}^{m_{\max}}(\Omega_1) \\ \vdots & \ddots & \vdots \\ Z_{00}^{0}(\Omega_K) & \dots & Z_{n_{\max}l_{\max}}^{m_{\max}}(\Omega_K) \end{bmatrix} \begin{bmatrix} \alpha_{00}^{0} \\ \vdots \\ \alpha_{n_{\max}l_{\max}}^{m_{\max}} \end{bmatrix} = \begin{bmatrix} H(\Omega_1) \\ \vdots \\ H(\Omega_K) \end{bmatrix}, \tag{15}$$

where $H(\Omega)$ are given in dBs to minimize errors in logarithmic scale.

Since the system is overdetermined, it is usually impossible to find coefficients which would perfectly satisfy this equation. However, it can be solved in the least-squares sense, minimizing the error of approximation. This approach is commonly embraced in research concerning the SH approximation (ZHANG *et al.*, 2012; ROMIGH *et al.*, 2015; PASQUAL *e al.*, 2014; ALON *et al.*, 2018). A few existing 4D directivity models featured separate computation of spatial and frequency dependencies, using least-squares fitting for space and then direct integration for frequency (ZHANG *et al.*, 2010; 2015). Including both frequency and spatial dependent functions in a single least-squares minimization to the best of my knowledge has not been yet applied for these kind of data, making it another novel element of the paper. While such an approach demands more computational resources, it provides better fitting and acknowledges coupling of space and frequency in directivity functions. It is also worth noting that the increased computational complexity mainly concerns determining the HSH coefficients and is caused by replacing multiple smaller matrix equations by one large. In real-time, with proper optimization, rendering binaural sound should be comparably fast for both SH and HSH HRTF representations.

HRTF sets usually lack data for low elevation angles ($\theta \to \pi$) because of the measurement setup restrictions. This can lead to some irrational values in the

---

[5]It is possible to obtain a higher spatial frequency, e.g., along θ than along ψ by using only a selection of HSHs for higher values of the corresponding parameters. However, this case is not applicable within this paper and thus it is not further discussed.

unsampled region, since the least-squares solver minimizes an error only at the points where data is available. One way of dealing with this effect is to apply a proper regularization (ZHANG *et al.*, 2015; ZOTKIN *et al.*, 2009). AHRENS *et al.* (2012) proposed another, even more efficient solution based on filling the missing region by means of the low-order SH approximation. However, since in the HSH definition the direction dependence is described by SHs, it is logical to assume that these issues can be handled for the HSH approximation in the same manner. For the comparison between HSHs and SHs, applying any of these methods should not impact the analysis results (both approximations are expected to be impacted in the same way). Thus, simple, non-regularized least-squares fitting was used for both bases.

The computations and analysis were performed in MATLAB using Objective-Oriented Directivity (SZWAJCOWSKI, 2021). Both the classes used in this research as well as precomputed objects containing raw and approximated data can be found in its database.

4. Exemplary approximations

To showcase efficiency of the HSH approximation, the method was tested on exemplary data. The chosen set of HRTFs comes from the original measurements of Knowles Electronics Manikin for Acoustic Research (KEMAR) with large pinnae performed at Massachusetts Institute of Technology (GARDNER, MARTIN, 1995) and is a typical HRTF set for evaluating different models (ZHANG *et al.*, 2009; 2010; 2015; SHEKARCHI *et al.*, 2013). The set contains 710 HRIRs for each ear, measured at different directions and each consisting of 512 samples recorded with the sampling frequency of 44.1 kHz. The HRIRs were converted to HCS as described in Subsec. 3.1. Only data for the left ear was used.

Although, according to Subsec. 3.2.1, the parameter limits of $\{l_{\max}, m_{\max}\} = 6$ and $n_{\max} = 64$ should be sufficient, it is important to remember that the SH-focused research used exact frequency spectra, and the spectral-focused one used exact spatial representation. Smoothing in both space and frequency can thus have more impact on the localization abilities. For this reason, the limits used in the exemplary approximation were set to slightly higher values of $\{l_{\max}, m_{\max}\} = 8$ and $n_{\max} = 80$.

For the comparison, the SH approximation was also performed on the same exemplary data. The procedure was exactly the same as for the HSH approximation, but the computations were carried out separately for each frequency bin. Then, approximated values for given directions were put together to retrieve discrete spectra. The maximum order and degree of SHs used for approximation was set to 8, to match the limiting parameters of SHs embedded in the HSHs. Exemplary

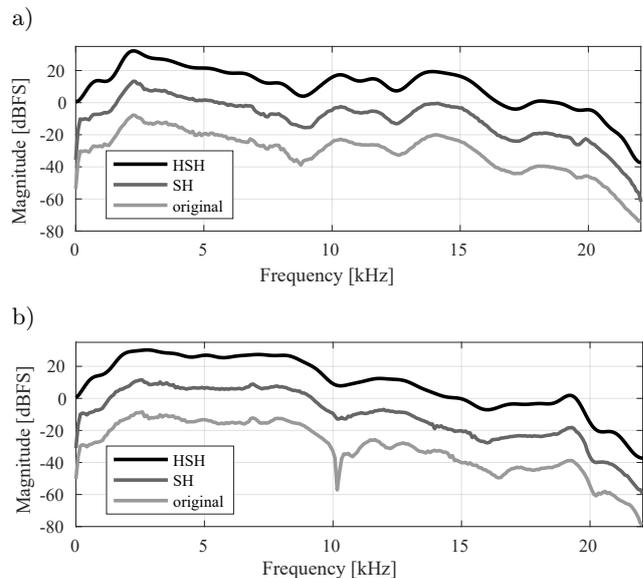raw spectra and their SH and HSH approximations are shown in Fig. 2.

Fig. 2. Raw and approximated spectra of KEMAR HRTFs for two exemplary directions: a) azimuth: 0°, elevation: 0°, far-field; b) azimuth: 30°, elevation: 30°, far-field. The spectra are offset by 20 dB for clarity. See text for more details.

The HSH approximation retrieves the general shape very well, but the resulting spectra are smoother than both the ones computed from the raw data and the ones obtained by means of the SH approximation. The most problematic details to represent in the HSH domain are deep notches, e.g., around 10 kHz in Fig. 2b. However, this is not exclusive to HSH approximation but comes from the spatial dependence represented by SHs, as the SH approximation features the same issues. The notches are caused by rapid changes in phase. ZAGALA and ZOTTER (2019) developed a method to deal with such a problem for the low-order SHs. Their algorithm could be extended and applied to HSHs if necessary.

By the properties of ψ-symmetric HSHs, their first derivative with respect to ψ has to be 0 at $\pi/2$, which causes minor discrepancies at the highest frequencies. However, since they lie beyond the human hearing range, these discrepancies are irrelevant from the perceptual point of view. Another problematic frequency band is the very lowest one. In this case, the SH approximation is much more precise than the HSH one; for low frequencies, the directivity takes very simple shapes (mostly omnidirectional), which are easy to express in the SH domain. One of the reasons for the inaccuracy of the HSH approximation in this band are rapid changes of magnitude levels within first few frequency bins, which come from the audio chain used in the HRTF measurement rather than the physical properties of HRTF themselves. This issue is investigated in depth in the following subsection.

4.1. Frequency weighting

Since some frequency bands lie beyond human hearing range, their impact on the approximation results should not be as high as the impact of the audible part of the spectra. This variation in relevance can be achieved by applying weighting to the least-squares minimization. Equation (15) then takes the form of:

$$(\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z})\boldsymbol{\alpha} = \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{H}, \qquad (16)$$

where $\mathbf{Z}$, $\boldsymbol{\alpha}$, and $\mathbf{H}$ denote the respective matrices from Eq. (15), $^{\mathrm{T}}$ denotes matrix transposition, and $\mathbf{W}$ is a diagonal matrix with weights for consecutive angle triplets:

$$\mathbf{W} = \begin{bmatrix} w(\Omega_1) & 0 & \dots & 0 \\ 0 & w(\Omega_2) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & w(\Omega_K) \end{bmatrix}. \qquad (17)$$

One of the regions where fitting could be improved by applying proper weights is the frequency band represented by the first few frequency bins. For all the directions, a drop-off can be noticed as the frequency approaches 0, which could be caused by a low-frequency limit of sound sources used in the HRTF measurement or a high-pass filter embedded in the audio chain. For the lowest frequencies, where the corresponding sound waves are much longer than the size of a human head, the HRTFs can be considered omnidirectional (see Fig. 1). This omnidirectionality mirrors the convergence of HCS for the frequency angle $\psi \to 0$. Since the problematic first bins do not hold any important information, they can be completely removed at the stage of computing the HSH approximation (i.e., have weight equal to 0). However, one needs to be wary that, under some circumstances, removing data from certain regions can lead to overfitting.

The other perceptually irrelevant part of the spectra are the highest frequencies. In this case, there are more than one frequency bins lying beyond the human hearing range. Their weights were arbitrarily chosen to start decreasing above 20 kHz and reach 0 at the Nyquist frequency following the shape of cosine function.

The HSH approximation was performed once again, applying the weights as described above. For most parts of the frequency spectrum, no noticeable changes occurred in regard to the previously performed the HSH computation on the complete data. Since there is on average more weight on the audible frequency band, the resulting approximation should be closer to the original data, but the improvements were minuscule in most of the frequency range. However, the fitting for the lowest frequencies has been notably improved, except for the region corresponding to the first

bins, which were effectively removed from the least-squares minimization (Fig. 3). Even though the resulting approximation is less accurate with regard to the measured data (large errors at first two bins), it is likely more accurate with regard to the factual physical HRTFs and improves fitting at several following bins.
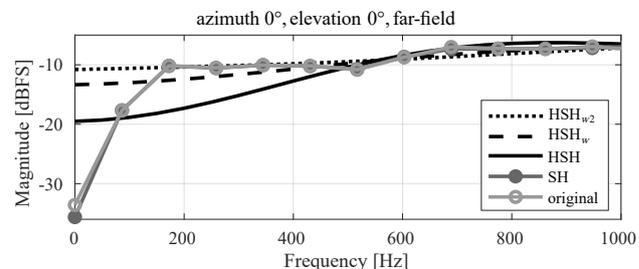

Fig. 3. Raw and approximated spectra of KEMAR HRTF for the direction straight ahead zoomed on the lowest frequencies. HSH_w and HSH_{w2} denote the HSH approximation with weights applied ignoring first or first two frequency bins, respectively (see text for more details).

From the perceptual point of view, the fitting in regions outside of the hearing range is irrelevant, and so the reduced impact of extreme frequency bins should improve the overall accuracy, even if only slightly. Thus, the following sections consider only this new approximation.

4.2. Approximation error analysis

Since 4D data is difficult to display efficiently as a 2D image, it is required to introduce numerical measures to compare accuracy of the SH and HSH approximations over the whole sphere. The key difference between these two approaches is the way of handling the frequency dependence and thus, for the comparison, the approximation error should be presented in the frequency domain. In similar research, various measures of approximation error were used. Currently, one of the most popular approaches is to evaluate it as the root-mean-square (RMS) error of the differences between logarithmic spectra (e.g., (Liu *et al.*, 2019; Romigh *et al.*, 2015; Nishino *et al.*, 1999; Li *et al.*, 2021)). A common alternative is to evaluate the error based on the linear magnitude (e.g., (Hu *et al.*, 2019; Zhang *et al.*, 2015)). Since this article has been written, there is no clear indication which of these approaches surpasses others, including the authors' speculations. However, the measure based on dB is more intuitive to interpret and the computations were performed on the logarithmic data, so it seems appropriate to use the logarithmic error as well. RMS errors for the HSH approximation[6] were computed by averaging

---

[6]From this section on, the HSH approximation is the one with first two frequency bins removed and weights applied for the high frequencies (earlier denoted as HSH_{w2}).

over all 710 directions available in the measurement data using the following formula:

$$\text{RMS}_{\widehat{H}}(f) = \sqrt{\frac{a^*}{\sum_{k=1}^{K} \delta_{\psi_k \psi_f}}}, \qquad (18)$$

where

$$a^* = \sum_{k=1}^{K} \left(\widehat{H}(\varphi_k, \theta_k, \psi_k) - H(\varphi_k, \theta_k, \psi_k) \right)^2 \delta_{\psi_k \psi_f}$$

and ψ_f is the frequency angle corresponding to the frequency f, derived from Eq. (9). The RMS error for the SH approximation was computed in an analogous way. In addition, for both methods and for each frequency bin, the 95th percentile (P_{95}) was determined for absolute values of $\widehat{H}(\Omega) - H(\Omega)$. Errors in the form of RMS an P_{95} values for both the SH and HSH approximations are plotted in Fig. 4, while Fig. 5 shows difference between the RMS plots from Fig. 4 as well as 5th and 95th percentile determined on the sets of differences between absolute errors in dB for these two methods:

$$P_{x\,\text{HSH-SH}}(f) = P_x(\{|\widehat{\mathbf{H}}_{\text{HSH}}^f - \mathbf{H}^f| - |\widehat{\mathbf{H}}_{\text{SH}}^f - \mathbf{H}^f|\}), \ (19)$$

where $\mathbf{H}^f$, $\widehat{\mathbf{H}}_{\text{SH}}^f$, and $\widehat{\mathbf{H}}_{\text{HSH}}^f$ are sub-vectors of the original data, SH and HSH approximations, respectively, which contain only values for a given frequency f.
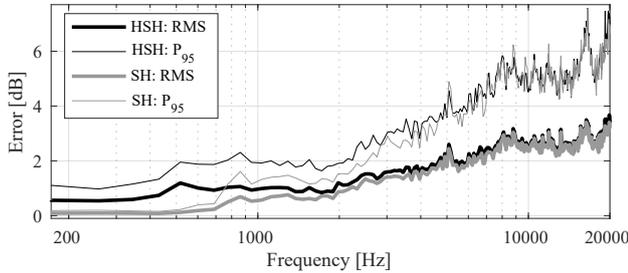

Fig. 4. RMS and P_{95} of the HSH and SH approximation errors in reference to the discrete data.
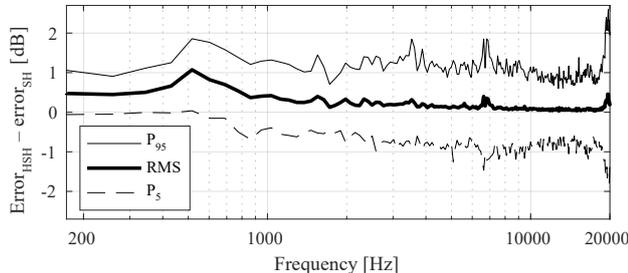

Fig. 5. Difference in RMS errors and 5th, and 95th percentile of differences between the HSH and SH approximation errors.

For the vast majority of the hearing range, the difference in accuracy between the HSH and SH approximations is small, reaching about 1 dB around 500 Hz

and becoming less and less significant as the frequency rises. The largest discrepancies occur for the lower frequency regions (below 1 kHz), where the SH approximation provides near perfect reconstruction of the original data, but even there the absolute error is still overall low (around 0.5–1 dB). The values of P_5 indicate that in some cases the HSH approximation yields lower errors than the SH one (especially for higher frequencies). The RMS error, however, is lower for the SH approximation for each frequency bin, even if only to a small extent.

4.2.1. Single-term error

To improve the comparability of the HSH approximation accuracy, the error needs to be provided as a single value averaged over all directions and frequencies. For logarithmic scale, an error measure called spectral distortion (SD) is usually employed. Its definition resembles the RMS definition in Eq. (18) but without the Kronecker deltas, so that the averaging is performed over both space and frequency:

$$\text{SD}_{\widehat{H}} = \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(\widehat{H}(\Omega_k) - H(\Omega_k) \right)^2}. \qquad (20)$$

SD values were computed for both SH and HSH approximations in the range 100 Hz to 20 kHz (the lower limit was imposed to exclude the first two frequency bins). The results were $\text{SD}_{\text{SH}} = 2.32$ dB and $\text{SD}_{\text{HSH}} = 2.44$ dB. Since SD requires a ground truth data, it cannot accurately capture the magnitude of differences between two approximations and so the difference between SD values is more representative than the SD of the differences between approximated data (just like the RMS difference is more accurate than the RMS differences in Fig. 5).

5. Discussion

In general, the HSH and SH approximations provide very similar accuracy within the hearing range. However, it is important to note that the minuscule difference in the approximations' errors for middle and audible high frequencies does not mean that there are no errors within these bands. In fact, the errors in both cases are lower for the lowest frequencies, but the SH approximation is very accurate in this region and thus the difference is bigger than for the middle and high frequencies. Such characteristics suggest that the lower accuracy of HSH approximation for higher frequencies comes from inability of low-order SHs to properly express the spatial variability of HRTFs in these bands. However, the maximum order of SHs was chosen based on perceptual studies and so it indicates that these magnitudes of errors are imperceptible. It is notable, that the approximations are not identical even when the RMS error is close to 0; the differences are small

(mostly below 1 dB) and are distributed relatively evenly between being in favor of either SH or HSH approximation. Such small differences in middle and high frequencies have been proven to be perceptually insignificant by many researchers (e.g., (BREEBAART, KOHLRAUSCH, 2001; XIE, ZHANG, 2010; HUOPANIEMI *et al.*, 1999)).

At low frequencies, the accuracy of the HSH approximation improved significantly after removing the first two frequency bins from the least-squares minimization. However, the SH approximation still outperforms the HSH one, providing near perfect reconstruction of the original spectrum at low frequencies. The largest differences occur around 500 Hz, which might be specific for the dataset used for evaluation. The way accuracy depends on frequency can be impacted by frequency mapping. Linear mapping treats all frequencies as equally important in the least-squares fitting, while it may be beneficial to employ a frequency scale more aligned with how the human hearing system works, e.g., the logarithmic or mel scale. On the other hand, it is known that spectral localization cues are dependent mostly on the shape of pinna, which impacts only higher frequencies (ALGAZI *et al.*, 2001; LANGENDIJK, BRONKHORST, 2002). Alternatively, the importance of accuracy in certain frequency bands can be manipulated by the frequency weighing if needed.

To further put the magnitude of errors in perspective, the approximation errors can be put against the measurement errors. Some research has been presented on the impact of different variables on the spectral variations between HRTF measurements (ANDREOPOULOU *et al.*, 2013; 2015); the HSH approximation error is on par with (or lower than) the human-related variations in repeated individual HRTF measurements or the differences between HRTF sets of the same artificial head but obtained in different laboratories. What is more, the difference between SH and HSH approximations is lower than the differences between the measurements performed in the same laboratory within a week (0.12 dB versus 0.17 dB). The frequency smoothing introduced by the HSH approximation, although changing the original values, might also be perceived as a positive effect, since some of the distortion in the raw data might be caused by noise and imperfections of the measurement setup.

Generally, the application of HSHs instead of SHs to approximate a magnitude of HRTFs yields almost the same accuracy while exhibiting some additional advantages. First of all, the resulting representation is fully continuous, not only in space but also in frequency. This continuity enables retrieving discrete HRTFs of any frequency resolution without any additional processing. The most prominent advantage of application of SHs to directivity functions was thus extended to another dimension. An exemplary benefit of using the HRTF model of infinite frequency resolution is the control of balance between precision and latency of binaural rendering by reading discrete HRTFs of any desired resolution.

What is more, even less amount of data is required to store a HRTF set in the HSH domain; in the analyzed example, the HSH approximation was described by 3081 coefficients, while the SH one required 81 for each of 257 frequency bins (255 excluding the first two bins that were ignored in the HSH approximation either way). The total number of SH coefficients was thus 20817 (20655), which is almost seven times more than in the HSH approximation, while retaining similar amount of perceptually relevant information. Even if SH approximation was determined at only 40 frequency bands to match the reduced frequency resolution of HSHs for $n_{max} = 80$, the total number of SH coefficients would be 3240 which is slightly more than the number of HSH coefficients (3081). It is also worth noting that 40-bin HRTFs would be computationally inefficient and would need to be resampled while no such operation is needed in the HSH representation, where spectra of any resolution can be read without any extra processing. Furthermore, comparing the number of HSH coefficients (3081) to the number of raw data samples (182470), the reduction of data size is almost 60-fold with SD below 2.5 dB, while the best configuration of the HRTF compression method proposed by SHEKARCHI *et al.* (2013) reached the compression ratio of only 40 for the same accuracy (to get the compression ratio of 60, SD increases to about 2.8 dB). Furthermore, their method is discrete in both space and frequency while the HSH representation is fully continuous. HSHs can be thus useful in designing a compression format that would include all the variability of HRTFs within a single set of coefficients, allowing them to be expressed holistically and compactly.

The downside of the HSH representation is that it requires more computational resources to determine the HSH coefficients via least-squares fitting. However, with the continuous development of processing units, this issue becomes less and less relevant.

The mathematical structure of the HSH representation allows it to be utilized not only for HRTFs but for any type of directivity function, e.g., directivity of electroacoustic devices, whether they are sound sources or receivers. These types of directivity functions are easier to measure and so applying HSHs as approximation tools might be of lesser value, but they can still be useful for such objects in any machine learning problems requiring proper directivity parametrization. HSHs, capturing holistic nature of the far-field directivity, seem to be better fitted for such tasks than SHs.

One of the interesting questions regarding the presented method is the impact of the choice of limiting parameters n_{max}, l_{max}, and m_{max} on the approximation accuracy. Increasing these limits will improve the

accuracy, but the quality and quantity of the improvement remains to be determined. This is intrinsically linked with the question of what is the minimum required accuracy from the perceptual standpoint and what is the best way to achieve it. However, the presented subject is very broad and requires a thorough analysis to provide reliable answers. For this reason, it is decided to be out of scope of the present paper, but will be considered in future studies.

6. Conclusions

Within this paper, the theory of HCS, real HSHs, and their application to express frequency-dependent directivity data such as HRTFs was presented. The entire computational process was described, starting from discrete HRIRs to determining HSH coefficients. The special focus was put on efficient mapping of frequency scale to an angle, which was required to express the directivity data in HCS. The location of the hyperspherical poles was leveraged to match the omnidirectionality of directivity functions such as HRTFs for low frequencies, making the HSH basis surprisingly well-suited to express such functions, without physical motivation.

The exemplary HSH representation was determined for KEMAR HRTFs using perceptually-motivated number of HSHs. The HSH approximation yields similar levels of accuracy to the corresponding SH one, while providing continuity in frequency and a significant reduction of required amount of data. Thanks to the continuity over both space and frequency, the HSH representation captures the holistic nature of far-field HRTF characteristics. Thus, the HSH representation can be considered an upgrade over the currently popular SH-based approach in practical applications such as binaural rendering, parametrization or data storage. However, there are still many aspects to investigate, such as, e.g., impact of the limiting parameters on the accuracy, exploring different frequency mapping or weighing, perceptual tests, etc.

The method described within this paper not only is a value by itself, but also presents wider possibilities of perceiving directivity functions by modeling frequency as an extra dimension, introducing coupling between their spectral and spatial properties. The HSH model can serve as a base for development of similar representations applying different 4D functions, e.g., created by merging basis functions of lower dimensions. Furthermore, this paper focuses only on the magnitude of far-field HRTFs, which can be in future complemented by the distance and phase dependencies if needed.

Acknowledgments

References

1. AES (2015), *AES69-2015. AES standard for file exchange – Spatial acoustic data file format*, Audio Engineering Society.

2. Agterberg M.J.H, Snik A.F.M, Hol M.K.S, Van Wanrooij, M.M., Van Opstal A.J. (2012), Contribution of monaural and binaural cues to sound localization in listeners with acquired unilateral conductive hearing loss: Improved directional hearing with a bone-conduction device, *Hearing Research*, **286**(1–2): 9–18, doi: 10.1016/j.heares.2012.02.012.

3. Ahrens J., Thomas M.R.P., Tashev I. (2012), HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data, [in:] *Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (IEEE)*.

4. Algazi V.R., Avendano C., Duda R.O. (2001), Elevation localization and head-related transfer function analysis at low frequencies, *The Journal of the Acoustical Society of America*, **109**(3): 1110–1122, doi: 10.1121/1.1349185.

5. Alon D.L., Ben-Hur Z., Rafaely B., Mehra R. (2018), Sparse head-related transfer function representation with spatial aliasing cancellation, [in:] *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6792–6796, doi: 10.1109/ICASSP.2018.8462101.

6. Andreopoulou A., Begault D.R., Katz B.F. (2015), Inter-Laboratory round robin HRTF measurement comparison, *IEEE Journal on Selected Topics in Signal Processing*, **9**(5): 895–906, doi: 10.1109/JSTSP.2015.2400417.

7. Andreopoulou A., Rogińska A., Mohanraj H. (2013), Analysis of the spectral variations in repeated head-related transfer function measurements, [in:] *International Conference on Auditory Display*, pp. 213–218.

8. Bates A.P., Khalid Z., Kennedy R.A. (2015), On the use of Slepian functions for the reconstruction of the head-related transfer function on the sphere, [in:] *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*.

9. Begault D.R., Wenzel E.M., Anderson M.R. (2001), Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source, *Journal of the Audio Engineering Society*, **49**(10): 904–916.

10. Ben-Hur Z., Alon D.L., Rafaely B., Mehra R. (2019), Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions, *EURASIP Journal on Audio and Music Processing*, **2019**: 5, doi: 10.1186/s13636-019-0148-x.

11. BLOMMER M.A., WAKEFIELD G.H. (1997), Pole-zero approximations for head-related transfer functions using a logarithmic error criterion, *IEEE Transactions on Speech and Audio Processing*, **5**(3): 278–287, doi: 10.1109/89.568734.

12. BONVALLET B., GRIFFIN N., LI J. (2007), A 3D shape descriptor: 4D hyperspherical harmonics an exploration into the fourth dimension, [in:] *Proceedings of IASTED International Conference on Graphics and Visualization in Engineering*, pp. 113–116.

13. BREEBAART J., KOHLRAUSCH A. (2001), Perceptual (ir)relevance of HRTF magnitude and phase spectra, [in:] *Proceedings of 110th AES Convention.*

14. BRINKMANN F., WEINZIERL S. (2018), Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition, [in:] *2018 AES International Conference on Audio for Virtual and Augmented Reality.*

15. CHEN J., VAN VEEN B.D., HECOX K.E. (1995), A spatial feature extraction and regularization model for the head-related transfer function, *The Journal of the Acoustical Society of America*, **97**(1): 439–452, doi: 10.1121/1.413110.

16. DOMOKOS G. (1967), Four-dimensional symmetry, *Physical Review*, **159**(5): 1387–1403, doi: 10.1103/PhysRev.159.1387.

17. EVANS M.J., ANGUS J.A.S., TEW A.I. (1998), Analyzing head-related transfer function measurements using surface spherical harmonics, *The Journal of the Acoustical Society of America*, **104**(4): 2400–2411, doi: 10.1121/1.423749.

18. GARDNER W.G., MARTIN K.D. (1995), HRTF measurements of a KEMAR, *The Journal of the Acoustical Society of America*, **97**(6): 3907–3908, doi: 10.1121/1.412407.

19. HARTUNG K., BRAASCH J., STERBING S.J. (1999), Comparison of different methods for the interpolation of head-related transfer functions, [in:] *16th International Conference: Spatial Sound Reproduction*, pp. 319–3129.

20. HU S., TREVINO J., SALVADOR C., SAKAMOTO S., SUZUKI Y. (2019), Modeling head-related transfer functions with spherical wavelets, *Applied Acoustics*, **146**: 81–88, doi: 10.1016/j.apacoust.2018.10.026.

21. HUOPANIEMI J., ZACHAROV N., KARJALAINEN M. (1999), Objective and subjective evaluation of head-related transfer function filter design, *Journal of the Audio Engineering Society*, **47**(4): 218–239.

22. KISTLER D.J., WIGHTMAN F.L. (1992), A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction, *The Journal of the Acoustical Society of America*, **91**(3): 1637–1647, doi: 10.1121/1.402444.

23. KULKARNI A., COLBURN H.S. (1995), Efficient finite-impulse-response filter models of the head-related transfer function, *The Journal of the Acoustical Society of America*, **97**(5): 3278–3278, doi: 10.1121/1.411579.

24. KULKARNI A., COLBURN H.S. (1998), Role of spectral detail in sound-source localization, *Nature*, **396**(6713): 747–749, doi: 10.1038/25526.

25. KULKARNI A., COLBURN H.S. (2004), Infinite-impulse-response models of the head-related transfer function, *The Journal of the Acoustical Society of America*, **115**(4): 1714–1728, doi: 10.1121/1.1650332.

26. KULKARNI A., ISABELLE S.K., COLBURN H.S. (1999), Sensitivity of human subjects to head-related transfer-function phase spectra, *The Journal of the Acoustical Society of America*, **105**(5): 2821–2840, doi: 10.1121/1.426898.

27. LANGENDIJK E.H.A., BRONKHORST A.W. (2002), Contribution of spectral cues to human sound localization, *The Journal of the Acoustical Society of America*, **112**(4): 1583–1596, doi: 10.1121/1.1501901.

28. LI J., WU B., YAO D., YAN Y. (2021), A mixed-order modeling approach for head-related transfer function in the spherical harmonic domain, *Applied Acoustics*, **176**: 107828, doi: 10.1016/j.apacoust.2020.107828.

29. LIU H., FANG Y., HUANG Q. (2019), Efficient representation of head-related transfer functions with combination of spherical harmonics and spherical wavelets, *IEEE Access*, **7**: 78214–78222, doi: 10.1109/ACCESS.2019.2921388.

30. MACPHERSON E.A., MIDDLEBROOKS J.C. (2002), Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited, *The Journal of the Acoustical Society of America*, **111**(5): 2219–2236, doi: 10.1121/1.1471898.

31. MAJDAK P. *et al.* (2013), Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions, [in:] *Proceedings of 134th AES Convention.*

32. NISHINO T., KAJITA S., TAKEDA K., ITAKURA F. (1999), Interpolating head related transfer functions in the median plane, [in:] *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 167–170.

33. PASHA HOSSEINBOR A. *et al.* (2015), 4D hyperspherical harmonic (HyperSPHARM) representation of surface anatomy: A holistic treatment of multiple disconnected anatomical structures, *Medical Image Analysis*, **22**(1): 89–101, doi: 10.1016/j.media.2015.02.004.

34. PASQUAL A.M. (2014), Spherical harmonic analysis of the sound radiation from omnidirectional loudspeaker arrays, *Journal of Sound and Vibration*, **333**(20): 4930–4941, doi: 10.1016/j.jsv.2014.05.006.

35. RASUMOW E. et al. (2014), Smoothing individual head-related transfer functions in the frequency and spatial domains, *The Journal of the Acoustical Society of America*, **135**(4): 2012–2025, doi: 10.1121/1.4867372.

36. ROMIGH G.D., BRUNGART D.S., STERN R.M., SIMPSON B.D. (2015), Efficient real spherical harmonic representation of head-related transfer functions, [in:] *IEEE Journal on Selected Topics in Signal Processing*, **9**(5): 921–930, doi: 10.1109/JSTSP.2015.2421876.

37. Shabtai N.R., Behler G., Vorländer M., Weinzierl S. (2017), Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments, *The Journal of the Acoustical Society of America*, **141**(2): 1246–1256, doi: 10.1121/ 1.4976071.

38. Shekarchi S., Hallam J., Christensen-Dalsgaard J. (2013), Compression of head-related transfer function using autoregressive-moving-average models and Legendre polynomials, *The Journal of the Acoustical Society of America*, **134**(5): 3686–3696, doi: 10.1121/1.4822477.

39. *Ultraspherical polynomials* (n.d.), Encyclopedia of Mathematics, http://encyclopediaofmath.org/index.php ?title=Ultraspherical_polynomials&oldid=52128.

40. Szwajcowski A. (2021), *Objective-Oriented Directivity. MATLAB toolbox for processing directivity models* (access: 3.06.2022).

41. Szwajcowski A., Krause D., Snakowska A. (2021), Error analysis of sound source directivity interpolation based on spherical harmonics, *Archives of Acoustics*, **46**(1): 95–104, doi: 10.24425/aoa.2021.136564.

42. Varshalovich D.A., Moskalev A.N., Khersonskii V.K. (1988), *Quantum Theory of Angular Momentum*, World Scientific.

43. Wenzel E.M., Arruda M., Kistler D.J., Wightman F.L. (1993), Localization using nonindividualized head-related transfer functions, *The Journal of the Acoustical Society of America*, **94**(1): 111–123, doi: 10.1121/1.407089.

44. Xie B., Zhang T. (2010), The audibility of spectral detail of head-related transfer functions at high frequency, *Acta Acustica United with Acustica*, **96**(2): 328–339.

45. Zagala F., Zotter F. (2019), Idea for sign-change retrieval in magnitude directivity patterns, [in:] *Conference: Fortschritte der Akustik – DAGA*.

46. Zhang M., Kennedy R.A., Abhayapala T.D. (2015), Empirical determination of frequency representation in spherical harmonics-based HRTF functional modeling, [in:] *IEEE/ACM Transactions on Audio Speech and Language Processing*, **23**(2): 351–360, doi: 10.1109/TASLP.2014.2381881.

47. Zhang W., Abhayapala T.D., Kennedy R.A., Duraiswami R. (2010), Insights into head-related transfer function: Spatial dimensionality and continuous representation, *The Journal of the Acoustical Society of America*, **127**(4): 2347–2357, doi: 10.1121/1.3336399.

48. Zhang W., Kennedy R.A., Abhayapala T.D. (2009), Efficient continuous HRTF model using data independent basis functions: Experimentally guided approach, *IEEE Transactions on Audio, Speech and Language Processing*, **17**(4): 819–829, doi: 10.1109/TASL. 2009.2014265.

49. Zhang W., Zhang M., Kennedy R.A., Abhayapala T.D. (2012), On high-resolution head-related transfer function measurements: An efficient sampling scheme, *IEEE Transactions on Audio, Speech and Language Processing*, **20**(2), 575–584, doi: 10.1109/TASL. 2011.2162404.

50. Ziegler J.D., Rau M., Schilling A., Koch A. (2017), Interpolation and display of microphone directivity measurements using higher order spherical harmonics, [in:] *Proceedings of 143rd AES Convention*.

51. Zotkin D.N., Duraiswami R., Grassi E., Gumerov N.A. (2006), Fast head-related transfer function measurement via reciprocity, *The Journal of the Acoustical Society of America*, **120**(4): 2202–2215, doi: 10.1121/1.2207578.

52. Zotkin D.N., Duraiswami R., Gumerov N.A. (2009), Regularized HRTF fitting using spherical harmonics, [in:] *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 257–260.