Efficient Representation and Sparse Sampling of Head-Related Transfer Functions Using Phase-Correction Based on Ear Alignment

Zamir Ben-Hur, David Lou Alon, Ravish Mehra, and Boaz Rafaely, *Senior Member, IEEE*

***Abstract*—With the proliferation of high quality virtual reality systems, the demand for high fidelity spatial audio reproduction has grown. This requires individual head-related transfer functions (HRTFs) with high spatial resolution. Acquiring such HRTFs is not always possible, which motivates the need for sparsely sampled HRTFs. Additionally, real-time applications require compact representation of HRTFs. Recently, spherical-harmonics (SH) has been suggested for efficient interpolation and representation of HRTFs. However, representation of sparse HRTFs with a limited SH order may introduce spatial aliasing and truncation errors, which have a detrimental effect on the reproduced spatial audio. This is because the HRTF is inherently of a high spatial order. One approach to overcome this limitation is to pre-process the HRTF, with the aim of reducing its effective SH order. A recent study showed that order-reduction can be achieved by time-alignment of HRTFs, through numerical estimation of the time delays of the HRTFs. In this paper, a new method for pre-processing HRTFs in order to reduce their effective order is presented. The method uses phase-correction based on ear alignment, by exploiting the dual-centering nature of HRTF measurements. In contrast to time-alignment, the phase-correction is performed parametrically, making it more robust to measurement noise. The SH order reduction and ensuing interpolation errors due to sparse sampling were analyzed for these two methods. Results indicate significant reduction in the effective SH order, where only 100 measurements and order 6 are required to achieve a normalized mean square error below −10 dB compared to a fully-sampled, high-order HRTF.**

***Index Terms*—Spatial audio, spherical-harmonics, head-related transfer functions (HRTFs).**

I. Introduction

SPATIAL audio interpolation plays an increasingly important role in applications such as virtual and augmented reality, spatial music, multimedia and gaming [1], [2]. Spatial audio gives the listener the sensation that sound sources are positioned in 3D space and helps to create immersive virtual sound scenes [3], [4]. A key component in spatial audio rendering through headphones is the head-related transfer function (HRTF) [5], which is the acoustic transfer function from a sound source to a listener's eardrum [6]. The HRTF is a complex-valued function of direction and frequency. At each frequency, it represents both the magnitude and phase shifts in the transformation of the sound-pressure waveform from the source to the eardrum.

For high quality spatial audio rendering, an individual HRTF is required [7], with high resolution in both the space and the frequency domains. Measurement of an individual HRTF at high spatial resolution is challenging and requires special and expensive equipment [8]–[10]. Therefore, using sparsely measured HRTFs is of great interest. However, the use of a sparse HRTF in head-tracked virtual audio systems necessitates some type of interpolation and real-time filter updating in order to create an accurate virtual sound environment.

A considerable amount of work has been done to find an adequate representation in the spatial domain that will facilitate efficient spatial interpolation, and that can be used in real-time applications. However, current methods of interpolation, such as bilinear interpolation or cubic spline interpolation, do not provide sufficiently accurate or high quality HRTFs from sparse measurements, due to the high spatial complexity of the HRTF, especially at high frequencies [5], [11].

Recently, the spherical-harmonics (SH) representation has become key to a very common interpolation approach [12]–[16], taking advantage of the spatial continuity and completeness properties of the SH basis functions over the sphere. However, SH representation of sparse HRTFs may lead to sparsity error, which comprises both truncation and aliasing errors [17], [18]. Truncation error is caused by the limited SH order [19], due to the limited number of spatial samples, or due to real-time requirements, where a small number of SH coefficients leads to a lower computational cost. The truncation error has been shown to have a detrimental effect on the reproduced spatial audio [19], [20]. Spatial aliasing error is caused by the limited number of spatial samples, where the high order SH coefficients of the sampled function are aliased into the low order coefficients [21], [22]. HRTFs can be considered to be order limited functions, where their order increases with frequency, and has been theoretically shown to be more than $N = 40$ at a frequency of 20 kHz [15]. This means that at least $(N + 1)^2 = 1681$ measurement directions are needed for accurate SH representation

Manuscript received March 31, 2019; revised June 27, 2019, August 7, 2019, and August 26, 2019; accepted September 24, 2019. Date of publication October 4, 2019; date of current version November 26, 2019. This work was supported by Facebook Reality Labs. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Federico Fontana. *(Corresponding author: Zamir Ben-Hur.)*

Z. Ben-Hur and B. Rafaely are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: zami@post.bgu.ac.il; br@bgu.ac.il).

D. L. Alon and R. Mehra are with the Facebook Reality Labs, Menlo Park, CA 94025 USA (e-mail: davidalon@fb.com; ravish.mehra@oculus.com).

Digital Object Identifier 10.1109/TASLP.2019.2945479

of the HRTF up to 20 kHz. Therefore, it is of great importance to find ways to reproduce high quality HRTFs from a sparse set of measurements.

One approach for reducing the required number of measurements is to pre-process the HRTF in order to reduce its effective SH order. Several pre-processing methods have been previously suggested in the context of SH representation of HRTFs [12], [16], [23]–[25]. A recent study by Brinkmann and Weinzierl [26] compared these methods, using a simulated HRTF, in terms of SH energy distribution and binaural models for source localization, coloration and correlation. They showed that complex frequency representation of the HRTF, which seems to be the most common method [12], [15], [19], [27]–[29], requires the highest SH order and leads to the largest errors for a given SH order. The time-alignment method, which was first suggested by Evans *et al*. [12], where the Head-Related Impulse Responses (HRIRs, which are the time domain representation of the HRTFs) are aligned in the time domain by removing their delays, requires the lowest SH order. They also tested several mixed methods, i.e. magnitude-phase representations, where the magnitude and phase responses were interpolated separately (using frequency unwrapped phase [12], spatial unwrapped phase [23], logarithmic magnitude [16] and smoothed magnitude [24]). In a recent study by Zaunschirm *et al*. [25], another method that uses a frequency-dependent time-alignment of the HRTF and results in similar SH order reduction to that of the original time-alignment method was suggested. The time-alignment method requires estimation of the time delay of each HRIR. However, this approach might be sensitive to errors, mainly due to low SNR values and the multiple peaks in the HRIR on the contralateral side [30], [31].

In this paper, we present a new pre-processing method that exploits the fact that the HRTF is defined as a ratio between two transfer functions, and performs phase-correction accordingly. The correction of the phase of a measured HRTF is performed by translating the origin of the free-field component from the center of the head to the position of the ear. The advantage of using such a correction is that it can be computed parametrically, without the need for numerical estimations, which makes it more robust to measurement noise. The effect of the phase-correction on the SH representation of sparse HRTFs is presented theoretically (Sections III, IV and V), and demonstrated numerically in terms of SH order reduction and interpolation errors using a rigid sphere approximation of the HRTF, and manikin and human HRTFs (Sections VI, VII and VIII). Results indicate that the SH order required to achieve a normalized Mean Square Error (MSE) of less than -10 dB, compared to a high-order HRTF, is reduced to 9 or less. This leads to sparse HRTF sampling with only 100 measurement directions, which is less than 10% of the number of measurements required to fully sample an HRTF of order around 30. Section II reviews the basic theory of SH representation of HRTFs and Section IX outlines the conclusions of the research.

II. SH REPRESENTATION OF HRTFS

This section presents the relevant background for SH representation of HRTFs. The HRTF, which describes the filtering effect due to the head, torso and ears of a human, is introduced as an acoustic transfer function. Consider a far-field arbitrary source from direction, $\Omega \equiv (\theta, \phi) \in \mathbb{S}^2$. A pair of HRTFs, H^l and H^r, for this source and for the left and right ears, respectively, is defined as [6]:

$$H^{l\backslash r}(\Omega, k) = \frac{p^{l\backslash r}(\Omega, k)}{p_0(\Omega, k)}, \tag{1}$$

where p^l and p^r represent the complex-valued sound pressure in the frequency domain at the left and right ears, respectively, p_0 represents the complex-valued free-field sound pressure in the frequency domain at the center of the head with the head absent, $k = \frac{2\pi f}{c}$ is the wave number, f is the frequency, and c is the speed of sound. Note that defining the HRTF as a ratio between sound pressures, instead of between transfer functions, is derived from the assumption that both pressures are measured with the same sound source.

SH representation of HRTFs has become widely used in recent research [12], [15], [16], [28], [32]. This representation is beneficial for many applications of spatial audio reproduction, such as spatial coding [12], efficient interpolation [14] and rendering of spatial audio from microphone array recordings [13], [29].

The SH decomposition, also referred to as the Inverse Spherical Fourier Transform (ISFT) of the HRTF is given by:

$$H^{l\backslash r}(\Omega, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} h_{nm}^{l\backslash r}(k) Y_n^m(\Omega), \tag{2}$$

where $Y_n^m(\Omega)$ is the complex SH basis function of order n and degree m [33], and $h_{nm}^{l\backslash r}(k)$ are the SH coefficients, which can be derived from $H^{l\backslash r}(\Omega, k)$ by the Spherical Fourier Transform (SFT):

$$h_{nm}^{l\backslash r}(k) = \int_{\Omega \in \mathbb{S}^2} H^{l\backslash r}(\Omega, k) [Y_n^m(\Omega)]^* \, \mathrm{d}\Omega, \tag{3}$$

where $\int_{\Omega \in \mathbb{S}^2} (\cdot) \mathrm{d}\Omega \equiv \int_0^{2\pi} \int_0^{\pi} (\cdot) \sin(\theta) \mathrm{d}\theta \mathrm{d}\phi$.

It is now assumed that the HRTF is order limited to order N, and that it is sampled at Q directions. The infinite summation in Eq. (2) can then be truncated and reformulated in matrix form as

$$\mathbf{h} = \mathbf{Y} \mathbf{h_{nm}}, \tag{4}$$

where the $Q \times 1$ vector $\mathbf{h} = [H(\Omega_1, k), \ldots, H(\Omega_Q, k)]^T$ holds the HRTF measurements over Q directions (i.e. in the space domain). The $l\backslash r$ notation has been removed for brevity. The $Q \times (N+1)^2$ SH transformation matrix, $\mathbf{Y}$, is given by

$$\mathbf{Y} = \begin{bmatrix} Y_0^0(\Omega_1) & Y_1^{-1}(\Omega_1) & \cdots & Y_N^N(\Omega_1) \\ Y_0^0(\Omega_2) & Y_1^{-1}(\Omega_2) & \cdots & Y_N^N(\Omega_2) \\ \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\Omega_Q) & Y_1^{-1}(\Omega_Q) & \cdots & Y_N^N(\Omega_Q) \end{bmatrix}, \tag{5}$$

and $\mathbf{h_{nm}} = [h_{00}(k), h_{0(-1)}(k), \ldots, h_{NN}(k)]^T$ is an $(N+1)^2 \times 1$ vector of the HRTF SH coefficients.

Given a set of HRTF measurements over a sufficient number of directions $Q \geq (N+1)^2$, the HRTF coefficients in the SH

domain can be calculated from the HRTF measurements, by using the discrete representation of the SFT [34],

$$\mathbf{h_{nm}} = \mathbf{Y}^{\dagger}\mathbf{h}, \tag{6}$$

where $\mathbf{Y}^{\dagger} = (\mathbf{Y}^H\mathbf{Y})^{-1}\mathbf{Y}^H$ is the pseudo inverse of the SH transformation matrix. Such representation in the SH domain allows for interpolation, i.e. for the calculation of the HRTF at any L desired directions using the discrete ISFT,

$$\widehat{\mathbf{h}}_L = \mathbf{Y}_L\mathbf{h_{nm}}, \tag{7}$$

where $\mathbf{Y}_L$ is the SH transformation matrix, as given in Eq. (5), calculated at the L desired directions, and $\widehat{\mathbf{h}}_L = [\widehat{H}(\Omega_1, k), \ldots, \widehat{H}(\Omega_L, k)]^T$ is the interpolated HRTF at the L desired directions.

In practice, the number of HRTF measurement directions is limited, and the SH order of the HRTF increases with frequency [15], [29]. Thus, if the relation $Q \geq (N+1)^2$ no longer holds, a sparsity error is introduced, which comprises both aliasing and truncation errors [17], [21].

III. The Effect of Dual-Centering of HRTF Measurements

In this section, the effect of defining the HRTF as a ratio of two transfer functions measured at different positions, as given in Eq. (1), on the SH representation of the HRTF is presented.

While HRTFs are measured with a microphone located at the ear, the origin of the spherical coordinate system is at the center of the head. In order to gain insight into the potential effect of this dual-centering measurement process, the simple case of a "free-field HRTF" is analyzed. This simplification is chosen because real HRTFs can only be analyzed numerically, while the "free-field HRTF" can be presented analytically.

Consider a single plane-wave of unit amplitude in free-field arriving from direction Ω with wave number k. The sound pressure at position (Ω_0, r), can be written as [35]:

$$\begin{aligned} p_0(\Omega, k, \Omega_0, r) &= e^{ikr\cos\Theta} \\ &= \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n j_n(kr)[Y_n^m(\Omega)]^* Y_n^m(\Omega_0), \end{aligned} \tag{8}$$

where Θ is the angle between Ω and Ω_0, and $j_n(\cdot)$ is the spherical Bessel function.

Defining the position of the ear to be at $(\Omega_{l\backslash r}, r_a)$, the "free-field HRTF" is defined by substituting Eq. (8) in Eq. (1):

$$H_0^{l\backslash r}(\Omega, k) = \frac{p_0(\Omega, k, \Omega_{l\backslash r}, r_a)}{p_0(\Omega, k, \Omega_0, 0)} = p_0(\Omega, k, \Omega_{l\backslash r}, r_a), \tag{9}$$

where $p_0(\Omega, k, \Omega_0, 0) = 1$ for all (Ω, k). Thus, for a sound field composed of plane-waves from directions $\Omega \in \mathbb{S}^2$ the "free-field HRTF" can be written as:

$$H_0^{l\backslash r}(\Omega, k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} 4\pi i^n j_n(kr_a)[Y_n^m(\Omega)]^* Y_n^m(\Omega_{l\backslash r}). \tag{10}$$

This equation is similar to the ISFT presented in Eq. (2). However, some mathematical manipulation is needed in order to derive the SH coefficients of the "free-field HRTF," as presented in Appendix A. The resulting coefficients are given by:

$$h_{nm\,0}^{l\backslash r}(k) = 4\pi i^n j_n(kr_a)[Y_n^m(\Omega_{l\backslash r})]^*. \tag{11}$$

Theoretically, these coefficients have some energy at every order n, which means that the HRTF is of infinite SH order. Nonetheless, due to the behavior of the spherical Bessel function, which has a negligible magnitude for $kr >> n$ [35], the "free-field HRTF" can be considered to be order limited by $N = \lceil kr_a \rceil$ (Ward & Abhayapala [36] showed that $N = \lceil kr \rceil$ gives an interpolation error of around -14 dB).

Thus, while the possible order of a "free-field HRTF" measured at the position of the ear and normalized by the center pressure is $N = \lceil kr_a \rceil$, looking at Eq. (9) it is clear that if the pressure at the position of the ear and the center pressure were measured at the same position, the "free-field HRTF" would be a constant 1, which is of SH order zero. This case demonstrates how sound pressure at the ear, which is measured at a distance r_a from the origin (center of the head), when normalized by a sound pressure at the origin, can lead to an increase in the SH order of the HRTF by approximately $N = \lceil kr_a \rceil$. Fig. 1 shows an example of this added order as a function of frequency, demonstrating the possible large effect at high frequencies where the SH order increases up to 32. Note the similarity of the orders in Fig. 1 to the actual order of HRTFs [15], which suggests that the translation of the origin may explain the high orders, as will be investigated in the remainder of this paper.

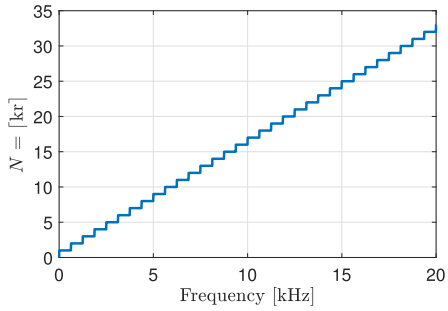


Fig. 1. Added SH order due to normalization of the ear pressure by center pressure, $N = \lceil kr \rceil$, as function of frequency. Computed for $r = 8.75$ cm and $c = 343.5$ m/s.

Although the theoretical explanation presented in this section may not be accurate for real measured HRTFs, it gives an insight with regard to the possible increase in SH order due to the fact that the ear microphone is located away from the origin.

IV. Phase-Correction by Ear Alignment

To counteract the effect described in the previous section, and to possibly reduce the effective SH order of the HRTF, a phase-correction method based on ear alignment is suggested.

Considering the effect of normalizing the ear pressure by the center pressure, which possibly increases the SH order of the HRTF, an ear-aligned HRTF could potentially be of lower SH order than the original HRTF. The ear-aligned HRTF can be defined as the ratio of the pressure at the ear and the free-field pressure at the position of the ear. A measured HRTF can be

aligned by translating the free-field pressure from the center of the head to the position of the ear. This can be formulated using Eq. (1) as:

$$H_a^{l\backslash r}(\Omega,k) = H^{l\backslash r}(\Omega,k)\cdot\frac{p_0(\Omega,k)}{p_0^{l\backslash r}(\Omega,k)} = \frac{p^{l\backslash r}(\Omega,k)}{p_0^{l\backslash r}(\Omega,k)}, \quad (12)$$

where H_a is the ear-aligned HRTF, and $p_0^{l\backslash r}$ is the free-field pressure at the position of the left or right ear. Assuming a far-field HRTF, the sound source can be a plane-wave, and the pressure in free-field can be computed using an exponential formulation as given in Eq. (8), which leads to the phase-correction formulation:

$$\begin{aligned} H_a^{l\backslash r}(\Omega,k) &= H^{l\backslash r}(\Omega,k)\frac{p_0(\Omega,k,\Omega_0,0)}{p_0(\Omega,k,\Omega_{l\backslash r},r_a)} \\ &= H^{l\backslash r}(\Omega,k)e^{-ikr_a\cos\Theta_{l\backslash r}}, \end{aligned} \quad (13)$$

where r_a is the radius of the head, $\Theta_{l\backslash r}$ is the angle between the direction of the source, Ω, and the direction of the ear, $\Omega_{l\backslash r}$, and $\cos\Theta_{l\backslash r} = \cos\theta\cos\theta_{l\backslash r} + \cos(\phi - \phi_{l\backslash r})\sin\theta\sin\theta_{l\backslash r}$. Note that, unlike other previously suggested methods for efficient HRTF representation (e.g. minimum phase [16]), this phase-correction process is invertible, which means that going from $H^{l\backslash r}$ to $H_a^{l\backslash r}$ and back can be performed without any loss of information.

The proposed phase-correction may seem similar to translating the HRTF in space to be centered at the position of the ear. In fact, in the context of surrounding spherical microphone arrays, acoustic centering has been studied [37], and it was found that the location of a source inside a spherical microphone array influences the energy distribution of the SH coefficients. It was concluded that acoustic centering of the sound source achieved the most compact SH representation. Richter *et al*. [27] suggested performing HRTF acoustic centering in the SH domain in post-processing, in order to achieve a compact transformation for real-time auralization. However, in this work, in contrast to acoustic centering, the proposed phase-correction can be performed in pre-processing, before performing the SFT, which means that a lower number of spatial sampling points of the HRTF may be required. Furthermore, the approach in [27] may suffer from spatial aliasing and truncation errors, due the the post-processing in the SH domain, while in the proposed method herein these errors are absent.

Additionally, the assumed SH order reduction of the ear-aligned HRTF can explain the findings of Brinkmann and Weinzierl [26] regarding the reduced order of time-aligned HRIRs. Phase-correction by ear alignment can be interpreted as "virtually" removing the inherent delay in an HRIR caused by normalizing the pressure at the ear by the center pressure. This is evident from Eq. (13), where the phase in the exponential represents a delay from the center to the ear due to a source at Ω. The difference between these two methods (time-alignment and phase-correction) is that while performing time-alignment requires numerically estimating the time delays, phase-correction can be performed parametrically with the parameters r_a and $\Omega_{l\backslash r}$. However, estimation of the time delays may be challenging and its accuracy depends on the HRTF direction and on the quality of the measurements [30], [31]. On the other hand, using the phase-correction with fixed parameters makes it data-independent (except for the choice of $\Omega_{l\backslash r}$ and r_a), which can potentially improve its robustness to measurement noises. In subsequent sections we will compare the two pre-processing methods, in the context of SH order reduction (Section VII) and sparse sampled HRTFs (Section VIII).

V. Sparse Sampling With Phase Correction

The previous section presented the phase-correction method for the pre-processing of HRTFs. This section will show the use of this method for sparse sampling of HRTFs.

Given a sampled HRTF over Q directions, $\{\Omega_q\}_{q=1}^{Q}$, its SH coefficients can be calculated using the SFT, as in Eq. (6). Assuming that the SH order of the HRTF is N, it will require $Q \geq (N+1)^2$ measurement directions for aliasing-free sampling [21]. Now, assuming that the phase-correction can provide an HRTF with a lower SH order $\widetilde{N} < N$ (this assumption is validated later in Sections VII and VIII), which requires $\widetilde{Q}$ measurements, then for aliasing-free sampling we get $\widetilde{Q} < Q$. For the case where $\widetilde{Q} \ll Q$, sparse sampling can be performed, while preserving the ability to reproduce high quality HRTFs.

In order to interpolate an HRTF from its sparse samples, first, prior to the SFT, a phase-correction is performed directly on the measured HRTF as given in Eq. (13):

$$\mathbf{h}_a = \mathbf{C}_{\widetilde{Q}}\mathbf{h}, \quad (14)$$

where $\mathbf{h}_a = [H_a(\Omega_1,k),\ldots,H_a(\Omega_{\widetilde{Q}},k)]^T$, $\mathbf{C}_{\widetilde{Q}} = \text{diag}[e^{-ikr_a\cos\Theta_1},\ldots,e^{-ikr_a\cos\Theta_{\widetilde{Q}}}]$, and Θ_q is the angle between the measured direction Ω_q and the direction of the ear $\Omega_{l\backslash r}$. Assuming that the phase-correction results in an ear-aligned HRTF, $\mathbf{h}_a$, of SH order $\widetilde{N}$, the SH coefficients of $\mathbf{h}_a$ can be calculated without error:

$$\mathbf{h}_{\mathbf{nm}}^{\mathbf{a}} = \widetilde{\mathbf{Y}}^{\dagger}\mathbf{h}_a, \quad (15)$$

where $\widetilde{\mathbf{Y}}$ is the order $\widetilde{N}$ SH transformation matrix of size $\widetilde{Q}\times(\widetilde{N}+1)^2$ defined similarly to in Eq. (5). From here, the HRTF can be interpolated to any desired direction, using an ISFT of order $\widetilde{N}$, as given in Eq. (7), $\widehat{\mathbf{h}}_{a\,L} = \widetilde{\mathbf{Y}}_L\mathbf{h}_{\mathbf{nm}}^{\mathbf{a}}$, and the inverse phase-correction can then be applied:

$$\widehat{\mathbf{h}}_L = \mathbf{C}_L^H\widehat{\mathbf{h}}_{a\,L}, \quad (16)$$

where $\widehat{\mathbf{h}}_L$ is the interpolated HRTF at the L desired directions, $\mathbf{C}_L = \text{diag}[e^{-ikr_a\cos\Theta_1},\ldots,e^{-ikr_a\cos\Theta_L}]$ and Θ_L is the angle between the desired direction Ω_l and the direction of the ear $\Omega_{l\backslash r}$.

The overall process of interpolating an HRTF at any desired L directions from its sparse samples, can be formulated as:

$$\widehat{\mathbf{h}}_L = \mathbf{C}_L^H\widetilde{\mathbf{Y}}_L\widetilde{\mathbf{Y}}^{\dagger}\mathbf{C}_{\widetilde{Q}}\mathbf{h}. \quad (17)$$

VI. Measures for Effective SH Order

To quantify the effect of the phase-correction on SH representation of HRTFs, several measures will be used. This section presents these measures. First, the energy distribution of the SH

coefficients is analyzed by calculating the SH spectrum and the parameter b_X, which is the lowest order that contains at least $X\%$ of the energy. The SH spectrum, which is the energy of the SH coefficients at every order n, is defined as:

$$E_n(f) = \sum_{m=-n}^{n} |h_{nm}(f)|^2. \tag{18}$$

Note that the measures in this section are presented as a function of frequency, f, instead of wave number, k. This change is done in order to be consistent with the presentation of the measures as a function of frequency in Sections VII and VIII.

The parameter b_X is then calculated as:

$$b_X\{\mathbf{h_{nm}}\}(f) = \left\{ \min N_b \in \mathbb{Z} : \sum_{n=0}^{N_b} E_n(f) \geq \frac{X}{100} \sum_{n=0}^{N} E_n(f) \right\}. \tag{19}$$

To evaluate the effect of the phase-correction on the interpolated HRTF, an interpolation error is calculated as a normalized MSE:

$$\epsilon_L(f) = 10 \log_{10} \frac{\|\mathbf{h}_L - \widehat{\mathbf{h}}_L\|^2}{\|\mathbf{h}_L\|^2}, \tag{20}$$

where $\mathbf{h}_L$ is the original HRTF, and $\widehat{\mathbf{h}}_L$ is the interpolated HRTF. Furthermore, magnitude and phase errors are defined as:

$$\epsilon_L^{\text{mag}}(f) = 10 \log_{10} \frac{\|\mathbf{h}_L^{\text{mag}} - \widehat{\mathbf{h}}_L^{\text{mag}}\|^2}{\|\mathbf{h}_L^{\text{mag}}\|^2}, \tag{21}$$

$$\epsilon_L^{\text{phase}}(f) = 10 \log_{10} \frac{\|\mathbf{h}_L^{\text{phase}} - \widehat{\mathbf{h}}_L^{\text{phase}}\|^2}{\|\mathbf{h}_L^{\text{phase}}\|^2}, \tag{22}$$

where $\mathbf{h}_L^{\text{mag}} = \left[|H(\Omega_1, f)|, \ldots, |H(\Omega_L, f)|\right]^T$ and $\widehat{\mathbf{h}}_L^{\text{mag}} = \left[|\widehat{H}(\Omega_1, f)|, \ldots, |\widehat{H}(\Omega_L, f)|\right]^T$ are vectors of the magnitude values of the original and interpolated HRTFs, respectively, and $\mathbf{h}_L^{\text{phase}}$ and $\widehat{\mathbf{h}}_L^{\text{phase}}$ are vectors of the phase values of the HRTFs.

VII. Simulation Study of Rigid Sphere and Manikin HRTFs

The measures defined in the previous section were calculated for three different HRTFs based on: (i) rigid sphere approximation, (ii) simulated HRTF of KEMAR [38], (iii) measured HRTF of Neumann KU-100 [39].

1) *Rigid Sphere:* As a first evaluation, an HRTF of an ideal rigid sphere is considered. This simplification offers us the ability to mathematically analyze the response, and to evaluate numerically the theoretical solution [40].
Consider a unit amplitude incident plane-wave with wave number k arriving from Ω_k. The "left ear" HRTF of a rigid sphere of radius r_a can be written, using the SH, as [41]:

$$H^l(\Omega_k, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} b_n(kr_a)[Y_n^m(\Omega_k)]^* Y_n^m(\Omega_l), \tag{23}$$

where b_n is given for a rigid sphere as follows:

$$b_n(kr_a) = 4\pi i^n \left[j_n(kr_a) - \frac{j_n'(kr_a)}{h_n'(kr_a)} h_n(kr_a) \right], \tag{24}$$

where j_n and h_n are the spherical Bessel and Hankel functions, and j_n' and h_n' are their derivatives.
Using Eq. (13) to correct the phase will give:

$$H_a^l(\Omega_k, k) = H^l(\Omega_k, k) e^{-ikr_a \cos\Theta_l}. \tag{25}$$

The rigid sphere HRTF was computed for a sphere with radius $r_a = 8.75$ cm, and position of the ear at $(90^\circ, 90^\circ)$, for a total of $Q = 2702$ directions in accordance with a Lebedev sampling scheme [42], which can provide an HRTF up to a spatial order of 44.

2) *KEMAR:* In order to study the effect of phase-correction on an HRTF, and to be able to compare the results with previously suggested pre-processing methods [26], a simulated HRTF of KEMAR that was used in [26] was analyzed. The HRTF was simulated using the Boundary Element Method (BEM) based on a 3D scan of the head of a KEMAR. The edge lengths of the mesh were graded from 1 mm at the left ear to 10 mm at the right ear (only the left ear HRTF was computed). The simulated frequencies were in the range of 100 Hz to 22 kHz with a resolution of 100 Hz. HRIRs were obtained by inverse Fourier transform, and shortened to 256 samples at a sampling rate of 44.1 kHz by applying squared sine fade-ins and fade-outs of 10 samples. A total of $Q = 4334$ directions were simulated in accordance with a Lebedev sampling scheme $(N = 56)$. For detailed information about the simulation process see Refs. [26], [43].
3) *KU-100:* The HRTF of the Neumann KU-100 dummy head was used to demonstrate the effect of phase-correction on measured HRTF data. The measurement contains a full sphere far field HRTF with a total of $Q = 2702$ directions in accordance with a Lebedev sampling scheme [39]. The measurement was conducted in the anechoic chamber at Cologne University of Applied Sciences. The chamber has a lower frequency boundary of around 200 Hz. Measured HRIRs were obtained with 128 samples at a sampling rate of 48 kHz. For detailed information about the measurement process see Ref. [39].

As a first step, for all three HRTFs, the phase-correction was applied using nominal parameters of $r_a = 8.75$ cm and assumed position of the left ear at $(\theta^l, \phi^l) = (90^\circ, 90^\circ)$. These parameters were chosen in accordance with the spherical head model [40]. A sensitivity analysis for these parameters is presented in Section VIII, together with an optimization study. In order to compare between the phase-correction and the time-alignment methods, a time-aligned HRTF, $H_{ta}(\Omega, k)$, was calculated as in Ref. [26], using an onset detection with a threshold of -20 dB and a ten times upsampled and low-passed HRIR (8th order Butterworth, $f_c = 3$ kHz) [31], [44]. For the interpolation of the time-aligned HRTFs, the delays were also interpolated using the SFT and were added to the interpolated time-aligned HRTFs.

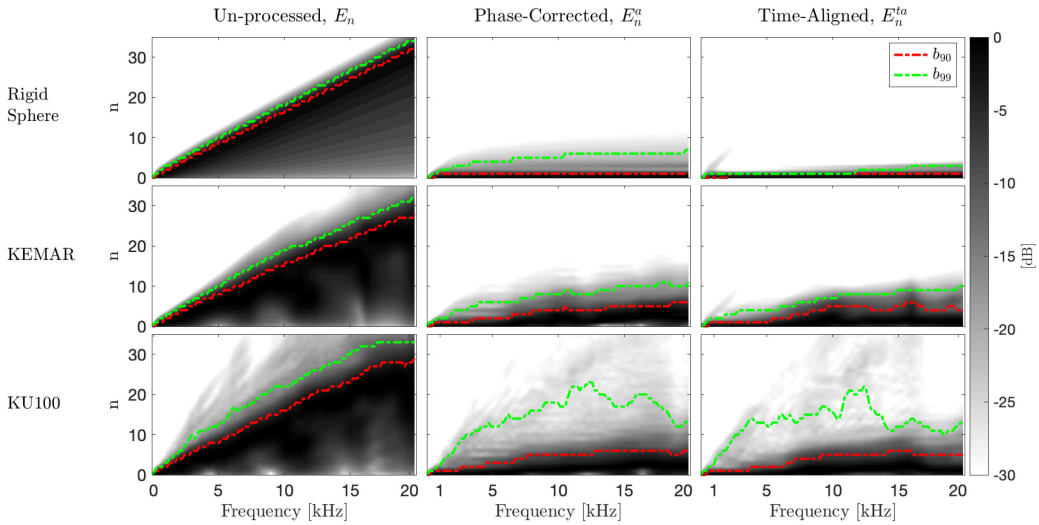


Fig. 2. Normalized SH spectra, E_n, of different HRTFs (rows) and different pre-processing methods (columns), computed according to Eq. (18). The dashed red and green lines represent the parameters b_{90} and b_{99}, respectively, where b_X is the lowest order at which at least X% of the total energy is contained at order b_X and below, computed according to Eq. (19) with $N = 35$ and $X = 90, 99$.

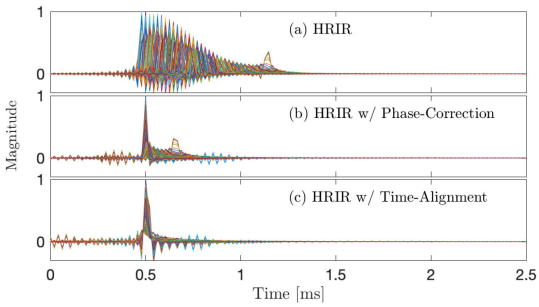


Fig. 3. HRIRs of a rigid sphere for all directions, computed using inverse Fourier transform of Eq. (23) (a), after phase-correction (b), and after time-alignment (c).

A. SH Spectrum Analysis

Fig. 2 presents the SH spectrum and the parameters b_{90} and b_{99} as a function of frequency and SH order. Each row in the figure presents a different HRTF, and each column presents a different pre-processing method. The SH spectra were normalized by the maximum value for each frequency. The figure shows how the energy of the high order SH coefficients of the phase-corrected and time-aligned signals are significantly reduced compared to the un-processed signal. This verifies the proposition presented to explain Fig. 1, in which the high orders of the original HRTF actually originate from the translation of the origin. In particular, the b_{90} line, which presents the order at which 90% of the energy is maintained, is reduced to be below order 10 for all frequencies and all HRTFs. While the rigid sphere and simulated KEMAR HRTFs have negligible energy above order 20 for all frequencies, the KU-100 HRTF still has energy up to order 35. This is also evident from the b_{99} line. This behavior can be explained by the fact that the KU-100 HRTF is measured and contains measurement artifacts that may lead to added spatial complexity, as is also evident from the energy distribution of the un-processed HRTF, which has higher energy at high SH orders compared to the simulated KEMAR HRTF.

Note that for the rigid sphere HRTF, time-alignment seems to reduce the SH energy slightly more than phase-correction. This can be explained by the fact that for this ideal case, the time-alignment has an advantage due to its ability to completely align the HRIR, while the phase-correction only corrects the free-field component. This means that the HRIR from the contralateral direction will still have some delay compared to the HRIR from the ipsilateral direction (due to diffraction). This behavior is demonstrated in Fig. 3, where the HRIRs of a rigid sphere after phase-correction and after time-alignment are presented, compared to the original HRIR. It can be seen that the time-aligned HRIR results in ideally aligned impulse responses, while the phase-corrected HRIR shows some misalignment, due to the signals from the contralateral directions.

B. Interpolation Error Analysis

For each HRTF, the interpolation error, $\epsilon_L(f)$, was computed as in accordance with Eq. (20) with $\mathbf{h}_L$ that contains 240 directions that were chosen from the original HRTF (closest to nearly-uniform distribution). The interpolated HRTF, $\widehat{\mathbf{h}}_L$, was computed by performing SFT with the remaining directions up to order 35 using Eq. (6) (without any processing, with phase-correction or with time-alignment). Finally, the HRTF coefficients were truncated in the SH domain to various orders, and then interpolated into the original 240 directions (using Eq. (7)).

Fig. 4 shows the minimum SH order in the truncation process needed to guarantee an interpolation error of less than a specific value of [−20 dB, −15 dB −10 dB]. This is shown for the three different HRTFs (rows) and pre-processing methods (columns).

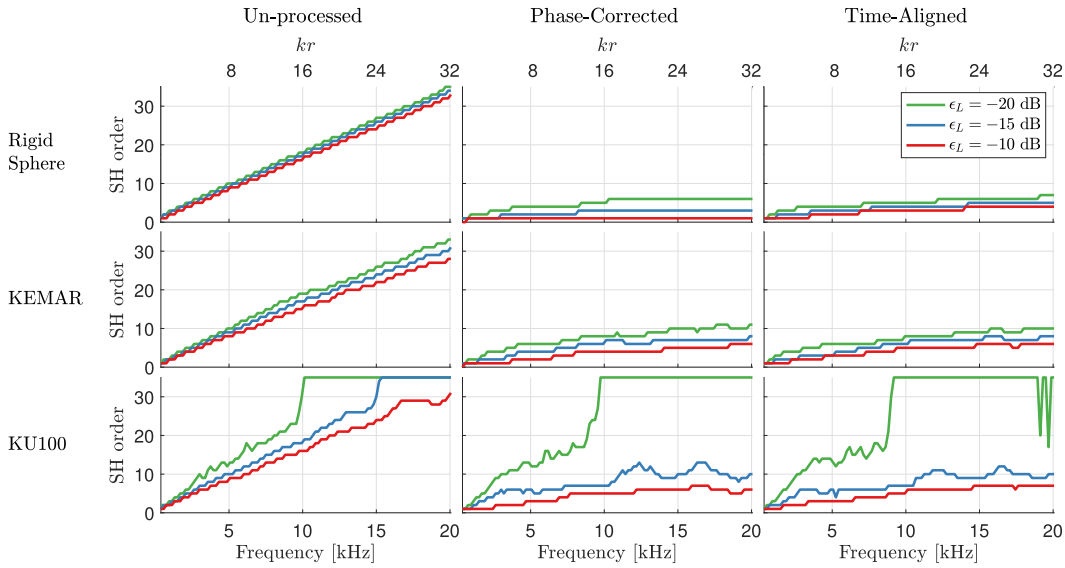


Fig. 4. The required SH order to achieve Normalized MSEs which are lower than a selected value [$\epsilon_L = -20$ dB, -15 dB, -10 dB]. The normalized error is computed as in Eq. (20), and presented as a function of frequency and kr, for different HRTFs (rows) and different pre-processing methods (columns).

The figure demonstrates the reduction in the required SH order for achieving reasonable HRTF interpolation over all frequencies, where the required SH order is significantly reduced when using phase-correction or time-alignment. In order to achieve an interpolation error of less than -10 dB, the required order is reduced to less than 8 for the entire bandwidth for all three tested HRTFs. For a -15 dB error, order 13 is needed.

As presented in Section IV, the phase-correction was shown to theoretically reduce the SH order by $N = \lceil kr \rceil$ (for an error of around -14 dB). The figure shows that this is a good approximation in the case of a rigid sphere, and also a very close approximation for KEMAR and KU100 HRTF sets.

In order to achieve a -20 dB error, which can be translated to a 1% error, the required order is still reduced significantly for the rigid sphere and KEMAR HRTFs. However, for the measured KU-100 HRTF the pre-processing methods are able to reduce the required order only up to around 8 kHz. This can be considered to be the limitation of these methods for a measured HRTF, as will be further investigated in Section VIII.

Fig. 5 shows the interpolation errors for the KU-100 HRTF as a function of SH order, averaged across frequencies from 0.3 to 20 kHz. Fig. 5(a) shows the complex error, ϵ_L, computed by Eq. (20), as presented at the beginning of the section, where $\widehat{\mathbf{h}}_L$ is defined as an interpolated version of $\mathbf{h}_L$ after order-truncated SFT, with or without pre-processing. The effect of the phase-correction and the time-alignment are clearly seen as a large reduction in the interpolation error over all orders. It seems that the phase-correction has some advantage at low SH orders. Fig. 5(b) shows the magnitude error, ϵ_L^{mag}, computed by Eq. (21), where $\mathbf{h}_L^{\text{mag}}$ and $\widehat{\mathbf{h}}_L^{\text{mag}}$ are the vectors of magnitude values of $\mathbf{h}_L$ and $\widehat{\mathbf{h}}_L$, respectively. In addition to the computation of the magnitude error by using the magnitude values of the original and interpolated HRTFs (i.e. using complex response interpolation), the error was computed for interpolating only

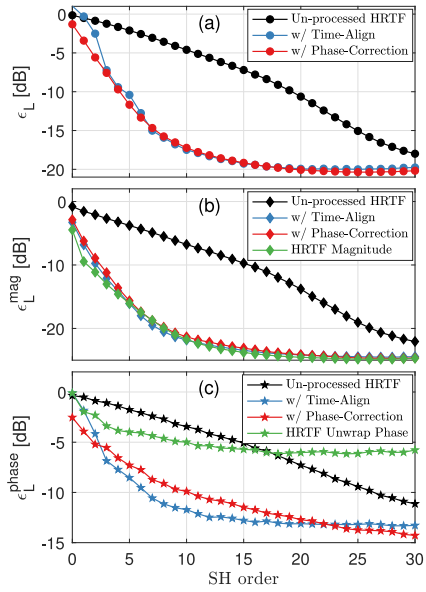


Fig. 5. Normalized MSE errors, ϵ_L (a), ϵ_L^{mag} (b), $\epsilon_L^{\text{phase}}$ (c), computed as in Eqs. (20), (21) and (22), respectively, for KU-100 HRTF, as a function of SH order, averaged across frequencies from 300 Hz up to 20 kHz. "HRTF Magnitude" was computed by interpolating only the magnitude of the original HRTF, and comparing it to the original HRTF magnitude. "HRTF Unwrap Phase" was computed by interpolating the unwrapped phase of the original HRTF, and comparing it to the phase of the original HRTF.

the magnitude response of the HRTF, which has been shown to improve the magnitude interpolation error [12], [16]. This is marked as "HRTF magnitude" in Fig. 5(b). In this case, the interpolation is performed directly on the magnitude response of the HRTF, which means that the pre-processing methods presented in this paper cannot affect it. The figure shows that using phase-correction or time-alignment and interpolating the complex response, result in similar performance to the case of

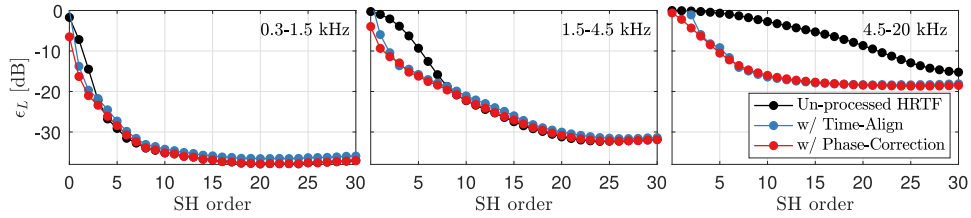


Fig. 6. Normalized MSE, ϵ_L, computed as in Eq. (20) for KU-100 HRTF, as a function of SH order, averaged across different frequency bands.

interpolating only the magnitude response. Fig. 5(c) shows the results for the phase error, $\epsilon_L^{\text{phase}}$, computed by Eq. (22), where $\mathbf{h}_L^{\text{phase}}$ and $\widehat{\mathbf{h}}_L^{\text{phase}}$ are the vectors of phase values of $\mathbf{h}_L$ and $\widehat{\mathbf{h}}_L$, respectively. Additionally, the phase error was computed by interpolating directly the unwrapped phase, which is different to interpolating the complex response and then computing the phase. This is marked "HRTF Unwrap Phase" in Fig. 5(c). Interestingly, the phase-correction and the time-alignment achieved better results than interpolating directly the unwrapped phase.

Fig. 6 shows the interpolation error, ϵ_L, for different frequency bands, demonstrating the significant contribution of high frequencies (above 4.5 kHz) to the interpolation error, and the improvement in the error at these frequencies obtained by using phase-correction or time-alignment.

Similar behavior of the interpolation error was obtained by using the rigid sphere and KEMAR HRTFs (with even lower errors when using the pre-processing methods, as can be expected from Figs. 2 and 4). These results are not presented here due to space constraints.

VIII. Experimental Investigation of Sparsely Sampled Human HRTFs

The previous section analyzed the effect of phase-correction and time-alignment on SH representation of HRTFs using densely sampled HRTFs, where only truncation error is introduced. This section analyzes the effect on sparsely sampled HRTFs, where sparsity error is introduced, by investigating human HRTFs obtained both from simulations and measurements.

A. Experimental Setup

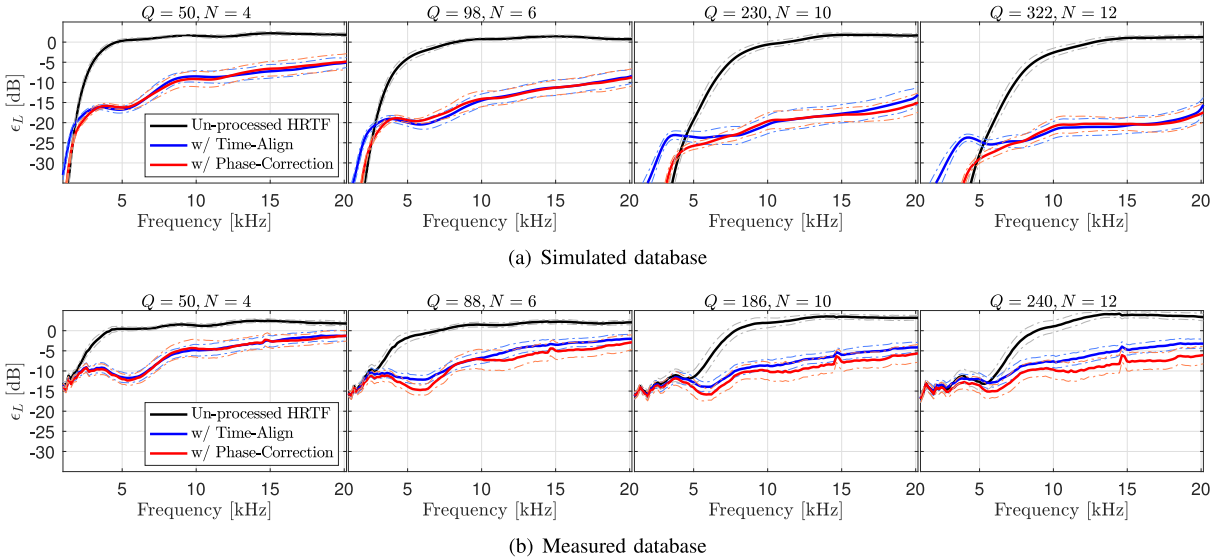
The HTRFs were taken from the HUTUBS database [45], [46]. This database contains numerical simulations and acoustic measurements of full-spherical HRTFs of 94 human subjects. The simulated HRTFs were computed using the BEM method for frequencies between 100 Hz and 22 kHz in steps of 100 Hz, generating a $Q = 1730$ point Lebedev grid. The edge length of the meshes was gradually increased from 1 mm at the simulated ear to 10 mm at the opposite ear (HRTFs were simulated separately for the left and right ear). HRIRs were obtained by inverse Fourier transform, and shortened to 256 samples at a sampling rate of 44.1 kHz by applying squared sine fade-ins of 10 samples and fade-outs of 20 samples. The acoustically measured HRTFs were measured in the anechoic chamber of the Technical University of Berlin with a sampling rate of 44.1 kHz, in a frequency range between 200 Hz and 20 kHz. Final HRIRs were truncated to 256 samples. The measurement regime comprised an equal elevation grid with 10° resolution and equal azimuth grids of different resolutions at each elevation (10° resolution at the horizontal plane decreasing towards the poles, to yield an almost constant great circle distance between neighboring points of the same elevation). This results in a full-spherical sampling grid with $Q = 440$ points, which ideally allows a SH transform of order $N = 16$. For detailed information about the simulation and measurement processes see Ref. [45].

B. Performance Analysis With Nominal Parameters

Interpolation errors were computed for each HRTF from the tested database by first taking out 121 directions from the HRTF, and then, using sparse subsets from the remaining directions and interpolating these into the original 121 directions. The 121 directions were chosen as the closest to a 10th order Extremal sampling scheme [47]. The sparse subsets were taken by choosing the Q directions that are the closest to an Nth order Gaussian sampling scheme [34]. The normalized MSE was calculated for the left ear HRTF of each subject, with phase-correction, with time-alignment and without any processing. The phase-correction was performed using the same nominal parameters as in Section VII, for all subjects (head radius of $r_a = 8.75$ cm, and left ear position of $(90^\circ, 90^\circ)$). Fig. 7 shows the mean and STD of the error across subjects, as a function of frequency, for different subsets.

Fig. 7(a) presents the results for the simulated HRTFs. The figure shows the significant effect of the phase-correction and time-alignment on the interpolated HRTF when using sparse HRTFs. With only 50 measurements and SH order 4, the errors reduced to below -10 dB up to 8 kHz and below -5 dB up to 20 kHz. With $Q = 322$ and $N = 12$, the errors are below -20 dB up to 18 kHz. These results are in agreement with the results presented in Section VII, where the phase-correction and time-alignment reduced 99% of the SH energy and lowered the required SH order for achieving -20 dB error to be around $N = 12$.

Fig. 7(b) presents the results for the measured HRTFs. Note that the numbers of used measurements are in some cases different from a Gaussian sampling of the same order; this is due to the limited number of original measurements (440). The positive effect of the pre-processing methods on the measured HRTF is smaller compared to the simulated HRTFs. However, the interpolation errors are still reduced significantly, where, using only 50 measurements and SH order 4, the errors reduced to below -10 dB up to 7 kHz and below -3 dB up to 14 kHz. With $Q = 186$ and $N = 10$, the errors are below -10 dB up to



(a) Simulated database

(b) Measured database

Fig. 7. Normalized MSE, ϵ_L, as a function of frequency, computed using HRTFs of 95 subjects taken from HUTUBS database [45]. The errors are computed for each subject using different subsets from the original HRTF. The bold lines represent the means across subjects, and the dashed lines represent the STDs: (a) results for simulated HRTFs; (b) results for measured HRTFs.

12 kHz when using phase-correction, and below -5 dB up to 20 kHz. Interestingly, no improvement is visible above $Q = 186$; this is also evident in the analysis of the measured KU-100 HRTF in Section VII. This behavior can be explained by the fact that the phase at high frequencies in acoustic measurements is very sensitive to measurement noise, which introduces errors that the pre-processing methods cannot overcome.

The results of the measured HRTFs demonstrate the advantage of the phase-correction compared to the time-alignment method. As the number of measurements increases, the differences between the two pre-processing methods also increases, where the largest difference is around 4 dB. The higher error of the time-alignment method can be explained by the fact that it is based on delay estimation, which cannot be robustly performed for signals with a low SNR. Such signals exist, for example, at low elevation directions due to the measurement system.[1] This result may suggest that the fact that the phase-correction is done parametrically and is data-independent makes it more robust to measurement noise, such as low SNR. To demonstrate this, the interpolation errors of the simulated HRTFs were computed again, with added noise. Fig. 8 shows the error as presented in Fig. 7(a) with $Q = 322, N = 12$ for different SNR values [30 dB, 35 dB, 40 dB, 50 dB]. The SNR was computed as the ratio between the energies of the left ear HRTF at direction $(90°, 90°)$ and the added Gaussian noise. The figure demonstrates the robustness of the phase-correction method to noise, where the interpolation error is almost the same for all SNR values, while the error when using time-alignment increases significantly when the SNR decreases.

[1]Note that several techniques and parameters for estimating the delays have been investigated by the authors. None of them achieved better results than the ones presented in the paper.

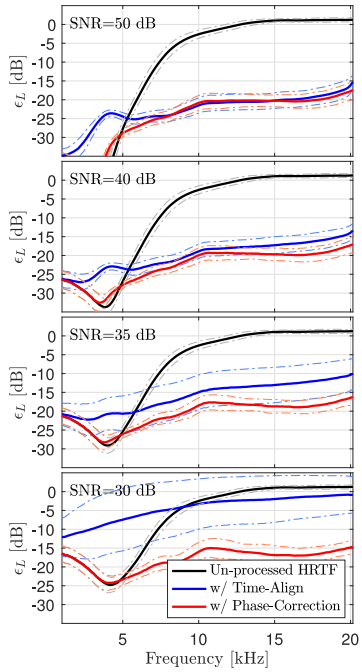


Fig. 8. Normalized MSE, ϵ_L, as a function of frequency, computed using simulated HRTFs, with added Gaussian noise, of 95 subjects taken from HUTUBS database [45]. The interpolated HTRFs were computed for each subject using a subset of $Q = 322$ directions from the original HRTF with SH order 12. The bold lines represent the means across subjects, and the dashed lines represent the STDs. Results are presented for different SNR values [30 dB, 35 dB, 40 dB, 50 dB].

In order to qualitatively evaluate how the interpolation error varies across directional regions of the HRTF, the interpolation errors were calculated for different angles. Fig. 9 shows the interpolation errors as a function of azimuth and elevation angles, averaged across all 94 subjects. The interpolated HRTFs

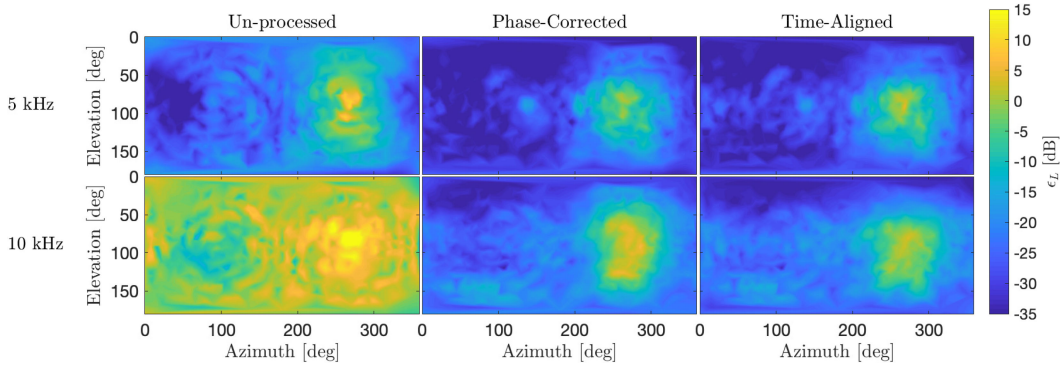


Fig. 9. Normalized MSE, ϵ_L, as a function of direction (azimuth, elevation), for two frequencies (rows) and un-processed and the two studied pre-processing methods (columns), averaged across 91 simulated HRTFs. The HRTFs were interpolated from a subset of $Q = 230$ directions with SH order $N = 10$.

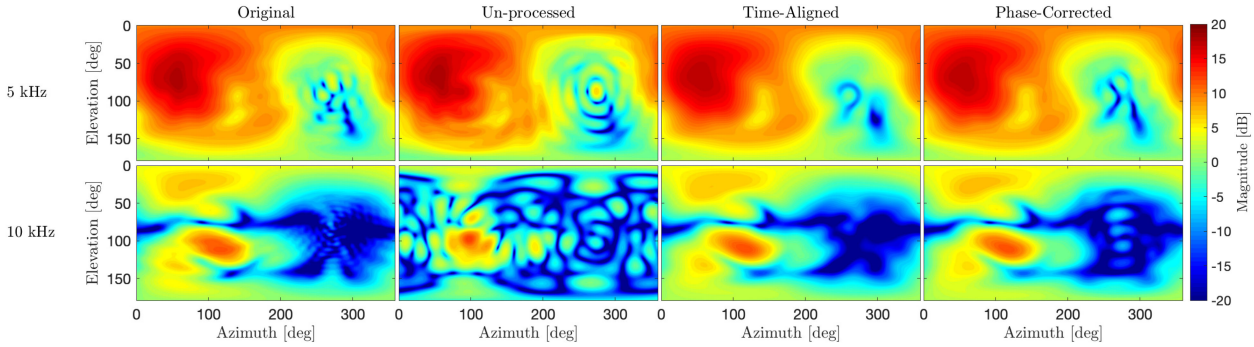
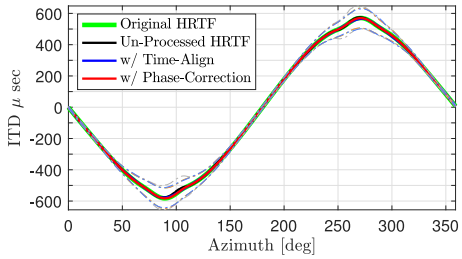


Fig. 10. Magnitude of the left ear HRTF of FABIAN head-and-torso-simulator, as a function of direction (azimuth, elevation), for two different frequencies (rows) and different pre-processing methods (columns). The HRTFs were interpolated from a subset of $Q = 230$ directions with SH order $N = 10$. The original HRTF (left column) was calculated using the original $Q = 1730$ directions using SH order $N = 35$.

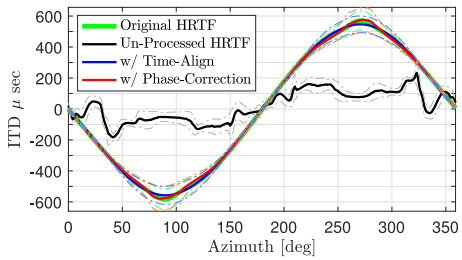
were calculated by taking the subset of $Q = 230, N = 10$ of the simulated HRTF and interpolating it into a full sphere with 1 degree resolution, with and without pre-processing. The figure shows the errors at 5 kHz and 10 kHz, demonstrating the reduction in the interpolation error when pre-processing is applied. It is interesting to note that relatively large errors appear at the contralateral directions. However, these error may not be as important as the figure may suggest. First, at these directions the overall energy of the HRTF is expected to be relatively low (see Fig. 10 for example). Second, as shown by psychoacoustic experiments, the information at high-frequencies at the contralateral directions are less important for sound localization [48].

To demonstrate how the interpolation error affects the magnitude of an interpolated HRTF as a function of direction, a simulated HRTF from the HUTUBS database of the FABIAN head-and-torso-simulator [49] was analyzed. As a reference, the original HRTF was also interpolated into the same grid using SH interpolation of order 35. Fig. 10 shows the magnitude of the interpolated left ear HRTF as a function of direction for the same two analyzed frequencies, 5 kHz and 10 kHz. The figure illustrates the errors introduced to sparse measurements, which can be seen as a highly distorted magnitude pattern for the un-processed HRTF, especially at 10 kHz. The pre-processed HRTFs, with phase-correction and time-alignment, are much more similar to the reference, while larger differences are still visible at the contralateral directions.

In addition to the magnitude errors, the effect of the different pre-processing methods on the Interaural Time Difference (ITD) is of great interest for spatial audio reproduction. To analyze this effect, the ITDs were computed for the simulated HRTFs of the 94 subjects using the subsets of $Q = 230, N = 10$ and $Q = 50, N = 4$, which were then interpolated into the horizontal plane with 1 degree resolution, with and without pre-processing. As a reference, the original HRTF was also interpolated into the same grid using SH interpolation of order 35. Then, the ITD was computed using the threshold detection method, with a threshold of -30 dB applied to a 3 kHz low-pass filtered version of the HRIRs. This method has been shown by Andreopoulou and Katz [31] to be the most perceptually relevant procedure for ITD estimation. Fig. 11 presents the mean and STD of the ITD values for the two subsets across subjects. As expected from the behavior of the ITD, which is mostly relevant at frequencies up to 1.5 kHz [5], [6], the ITD values for the high order subset ($Q = 230, N = 10$) are accurate even with the un-processed HRTF. However, with the low order subset of $Q = 50, N = 4$, significant errors in the ITD are introduced with the un-processed HRTF, while pre-processing the HRTFs with both methods achieved significant improvement. As reported by Andreopoulou and Katz [31], the Just Noticeable Difference



(a) $Q = 230, N = 10$



(b) $Q = 50, N = 4$

Fig. 11. ITD over the horizontal plane, computed using simulated HRTFs of 95 subjects taken from HUTUBS database [45]. The interpolated HTRFs were computed for each subject using two subsets: (a) $Q = 230, N = 10$, (b) $Q = 50, N = 4$. The bold lines represent the means across subjects, and the dashed lines represent the STDs.

(JND) for the ITD is subject and angle dependent. The lower JND is in the frontal direction ($\phi = 0°$), with an average value of about 40 μs, and the highest is at the lateral directions (ϕ between $80°$ to $100°$), with an average value of about 100 μs. To assess the errors in ITD between the original HRTF and the pre-processed HRTFs with comparison to the JND, the average errors across directions were calculated for each subject. The maximum of these errors were found to be 6 μs and 12.7 μs for the time-aligned and phase-corrected HRTFs, respectively, under both sampling conditions. As these errors are much smaller than the JND, it is expected that pre-processing may distort the ITDs but in a level that is not noticeable in most directions.

C. Sensitivity Analysis For Parameter Values

The results presented above for the phase-correction method were obtained using nominal subject-independent parameters (head radius r_a, left ear azimuth ϕ^l and elevation θ^l). However, it is known that actual human heads are neither spherical, single sized, nor have their ears in a centered position. In order to study the sensitivity of the phase-correction method to the applied parameters, the interpolation errors were computed again for 91 simulated HRTFs (the HRTFs with anthropometric metrics given in the database). The analysis in Fig. 7 was repeated here with the $Q = 230, N = 10$ subset, using the phase-correction method with different parameter values (subject-independent). The error was calculated for the 121 directions selected from the original HRTF, for all 91 subjects. Fig. 12(a) shows the mean and STD values of the interpolation error over all directions and subjects using phase-correction with the left ear position at $(90°, 90°)$ and different values of head radius. Figures 12(b) and 12(c) show the errors when using $r_a = 8.75$ cm and different values of left ear elevation and azimuth positions, respectively. The different parameter values were chosen to cover a reasonable range of deviations from the nominal parameters. The figures show that small deviations from the nominal parameters ($7\,\text{cm} \le r_a \le 9\,\text{cm}$, $85° \le \theta^l \le 92°$ and $85° \le \phi^l \le 95°$) have a moderate effect on the interpolation error, mostly at very high frequencies (above 15 kHz), where a maximum of 2–3 dB differences in the errors appear in this parameter range. However, higher deviation from the nominal parameters, for example $r_a = 6$ cm, $\theta^l = 100°$ or $\phi^l = 80°$, lead to larger errors. These results suggest that the selection of values for the phase-correction parameters can be made in practice without the need for great accuracy, as long as the deviations from the nominal values as employed here are reasonably small.

D. Individualized Selection of Parameter Values

To further analyze the effect of the parameter values on the phase-correction method, an optimization study was performed. The optimal parameters for each subject from the database were found by minimizing the center of power of the SH coefficients of the HRTF, as defined by Ben-Hagai et al. [37]:

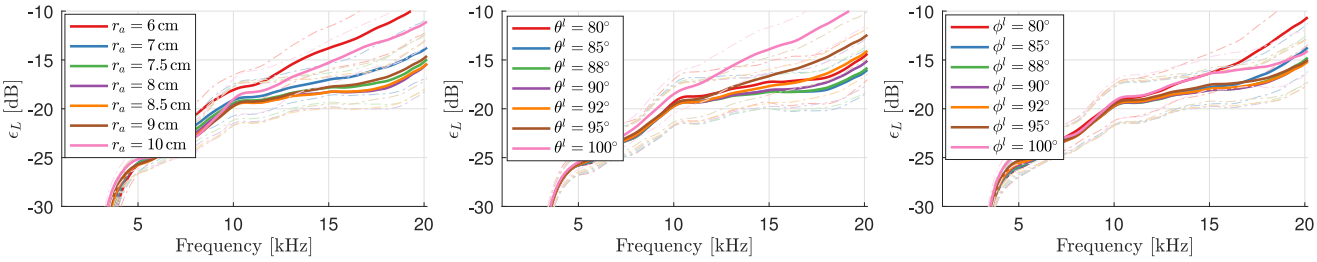
$$\mathbf{x}_{\text{opt}} = \arg\min_{\mathbf{x}} J_2, \tag{26}$$

where $\mathbf{x} = (r_a, \theta^l, \phi^l)$, $\mathbf{x}_{\text{opt}} = (r_{a\,\text{opt}}, \theta^l_{\text{opt}}, \phi^l_{\text{opt}})$ and J_2 is the center of power, defined as:

$$J_2 = \frac{\sum_{n=0}^{N} \sum_{m=-n}^{n} n|h_{nm}|^2}{\sum_{n=0}^{N} \sum_{m=-n}^{n} |h_{nm}|^2}. \tag{27}$$

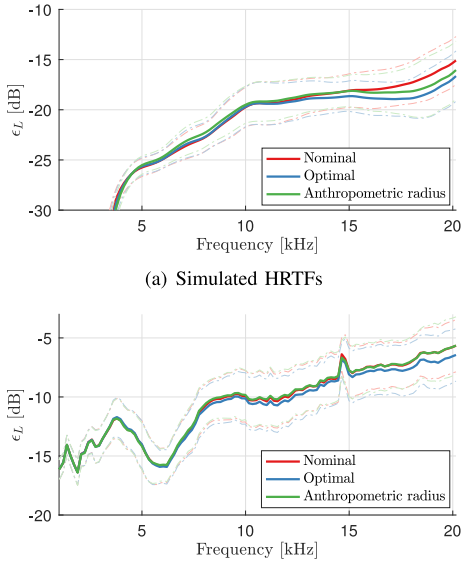
Minimization of J_2 means that the HRTF power is concentrated at the low SH orders, which leads to reduction in the effective SH order as shown in Section VII-A. An optimization solver was applied to compute $\mathbf{x}_{\text{opt}}$. Note that the optimization can be applied in practice on real data, as it does not require any additional information. The optimization problem was formulated using MATLAB version R2018b, and solved using the built-in function *fminsearch*, which implements the optimization method of the simplex search method [50].

Additionally, the subjects' head-widths were taken from the anthropometric metrics given in the database and were used for comparison with the optimal r_a. The average error between the computed optimal radius and half the measured head-width is 1.65 cm, where, in most cases, the optimal radius is larger. This can be explained by the fact that the human head is not spherical and usually the head-width (measured as the distance between the temples) is the smallest part of the head. Note that other anthropometric measures could be used for estimating the head radius, such as the head-depth and head-height. However, preliminary studies have indicated that the use of these measures, or an average of their combinations, do not yield better results than the ones presented herein. This implies that the suggested method seems to optimize for the smallest measure of the head, which is not unexpected - the smallest measure typically corresponds to the translation distance between the origin (center of the head) and the ear.



(a) Sensitivity to r_a (b) Sensitivity to θ^l (c) Sensitivity to ϕ^l

Fig. 12. Sensitivity analysis of the phase-correction parameters. Each figure presents interpolation error, ϵ_L, as a function of frequency, computed using simulated HRTFs of 91 subjects taken from HUTUBS database [45]. The errors are computed for each subject using a subset of $Q = 230$ directions from the original HRTF. The bold lines represent the means across directions and subjects, and the dashed lines represent the STDs. The interpolated HRTFs were computed using the phase-correction method with different parameters: (a) different head radius; (b) different ear elevation angle; (b) different ear azimuth angle.



(a) Simulated HRTFs

(b) Measured HRTFs

Fig. 13. Error analysis for the optimized phase-correction parameters. Each figure presents interpolation error, ϵ_L, as a function of frequency, computed using HRTFs of 91 subjects taken from HUTUBS database [45]. The bold lines represent the means across subjects, and the dashed lines represent the STDs. The interpolated HRTFs were computed using the phase-correction method with nominal, optimal and anthropometric parameters: (a) results for simulated HRTFs; (b) results for measured HRTFs.

TABLE I
MEAN AND STD ACROSS SUBJECTS OF THE OPTIMAL PARAMETERS FOR THE PHASE-CORRECTION METHOD. COMPUTED USING EQ. (26) WITH ORDER $N = 10$

	Anthropometric r_a	$r_{a\,\text{opt}}$	θ^l_{opt}	ϕ^l_{opt}
	Mean (STD)	Mean (STD)	Mean (STD)	Mean (STD)
Simulated HRTFs	7.62 (0.36) cm	8.21 (0.43) cm	88.55° (1.69°)	91.92° (2.35°)
Measured HRTFs		9.22 (0.88) cm	88.21° (3.81°)	93.73° (3.05°)

Fig. 13 presents the interpolation error, computed from the sampling subsets of $Q = 230$ and $Q = 186$ of the simulated and measured HRTFs, respectively, with SH order 10. The phase-correction was applied with: (i) the nominal parameters ($r_a = 8.75$ cm, $(\theta^l, \phi^l) = (90°, 90°)$), (ii) the optimal parameters per subject (mean and STD over subjects are presented in Table I), and (iii) the head radius taken from the anthropometric metrics (in this case, the optimal $(\theta^l_{\text{opt}}, \phi^l_{\text{opt}})$ were used). The figure shows the errors for both simulated and measured HRTFs. It can be seen that subject-dependent optimization of the parameters has a relatively small effect on the interpolation error, which is mostly present at the very high frequencies (above 15 kHz). Additionally, as expected, using the computed optimal head radius is preferable to using the measured head-width. The results presented here reinforce the conclusions from Fig. 12, that phase-correction can be used in practice by employing nominal parameter values without the need for additional subject-dependent data. Note that the presented analysis demonstrates the relative objective errors when varying the parameters values. However, while small relative errors may lead to small perceptual differences, the actual implication on perception is not clear and requires additional investigation. Overall, it seems that using nominal values for the phase-correction parameters may be useful in practice, while a further small improvement can be achieved using optimization or individualization.

The results presented in this section suggest that using the phase-correction method has practical benefit in the design of HRTF measurement systems. The proposed method enables the measurement of an HRTF at a relatively small number of directions, but with accurate interpolation with high spatial resolution. In comparison to the time-alignment method, it has been shown to be more robust to additive measurement noise, while its parametric nature makes it easier to implement in practice. Additionally, using phase-correction leads to a compact SH representation of the HRTF, which is beneficial for real-time applications. Furthermore, as was presented in [26], this reduction in the effective SH order of the HRTF is expected to lead to a smaller HRTF sampling error and, therefore, to improved perceptual performance when using these HRTFs.

IX. CONCLUSION

This paper presented a method for efficient representation of sparse sampled HRTFs using phase-correction by ear alignment. A formulation of the phase-correction method was developed by considering the dual-centering of HRTF measurements. The method was evaluated by simulation and experimental studies with simulated and measured HRTFs. Significant reduction in the effective SH order was achieved, which affords the ability to measure HRTFs with many fewer directions and still

interpolate it with high spatial resolution. A comparison with a previously suggested method of time-alignment was performed, demonstrating the robustness of the phase-correction method to additive measurement noise. Sensitivity analysis and an optimization study imply that nominal ear-alignment parameters can be used in practice and yield similar results to those obtained with optimized parameters. This paper presented an objective evaluation of the suggested method, while subjective evaluation of the effects on spatial perception, which is essential for the completeness of this study, is suggested for future work.

APPENDIX A
COMPUTATION OF "FREE-FIELD HRTF" SH COEFFICIENTS

This appendix presents the calculation stages of the SH coefficients of the "free-field HRTF," as presented in Eq. (10).

Given Eq. (10) and defining a new function $g(\Omega_k, k) \equiv [H_0^{l\backslash r}(\Omega_k, k)]^*$, then:

$$g(\Omega_k, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} [4\pi i^n j_n(kr_a)]^* Y_n^m(\Omega_k)[Y_n^m(\Omega_{l\backslash r})]^*. \tag{28}$$

Consider the definition of the ISFT (as given in Eq. (2)), the SH coefficients of $g(\Omega_k, k)$ are:

$$g_{nm}(k) = [4\pi i^n j_n(kr_a)]^*[Y_n^m(\Omega_{l\backslash r})]^*. \tag{29}$$

Now, using the *complex conjugate* property of the SFT (as in Eq. (1.46) in [34]), $g_{nm} = (-1)^m[h^{l\backslash r}_{n(-m)\,0}]^*$, where $h^{l\backslash r}_{nm\,0}$ are the SH coefficients of $H_0^{l\backslash r}$:

$$(-1)^m[h^{l\backslash r}_{n(-m)\,0}]^* = [4\pi i^n j_n(kr_a)]^*[Y_n^m(\Omega_{l\backslash r})]^*. \tag{30}$$

Applying the complex conjugate to both sides of the equation, multiplying by $(-1)^{(-m)}$ and substituting $m' = -m$:

$$h^{l\backslash r}_{nm'\,0} = 4\pi i^n j_n(kr_a)(-1)^{m'} Y_n^{(-m')}(\Omega_{l\backslash r}). \tag{31}$$

Finally, by using the *complex conjugate* property of the SH functions (as in Eq. (1.10) in [34]), $[Y_n^m(\Omega)]^* = (-1)^m Y_n^{(-m)}(\Omega)$, the SH coefficients of the free-field HRTF can be written as Eq. (11):

$$h^{l\backslash r}_{nm'\,0} = 4\pi i^n j_n(kr_a)[Y_n^{m'}(\Omega_{l\backslash r})]^*. \tag{32}$$

ACKNOWLEDGMENT

The authors would like to thank F. Brinkmann for sharing his simulated KEMAR HRTF and his time-alignment process code.

REFERENCES

[1] D. R. Begault and L. J. Trejo, *3-D Sound for Virtual Reality and Multimedia*, NASA, Ames Research Center, Mountain View, CA, USA, pp. 132–136, 2000.
[2] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Berlin, Germany: Springer, 2007.
[3] H. Møller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, no. 3, pp. 171–218, 1992.
[4] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization-an overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875, 1993.
[5] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*. Plantation, FL, USA: J. Ross Publishing, 2013.
[6] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.
[7] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 111–123, 1993.
[8] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. i: Stimulus synthesis," *J. Acoust. Soc. America*, vol. 85, no. 2, pp. 858–867, 1989.
[9] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity," *J. Acoust. Soc. America*, vol. 120, no. 4, pp. 2202–2215, 2006.
[10] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *J. Audio Eng. Soc.*, vol. 55, no. 7/8, pp. 623–637, 2007.
[11] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in *Proc. Audio Eng. Soc. Conf., 16th Int. Conf.: Spatial Sound Reproduction*, 1999, pp. 319–329.
[12] M. J. Evans, J. A. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. America*, vol. 104, no. 4, pp. 2400–2411, 1998.
[13] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov, and L. S. Davis, "High order spatial audio capture and binaural head-tracked playback over headphones with HRTF cues," in *Proc. 119th Convention Audio Eng. Soc.*, 2005, pp. 1–16.
[14] W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Efficient continuous HRTF model using data independent basis functions: Experimentally guided approach," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 4, pp. 819–829, May 2009.
[15] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *J. Acoust. Soc. America*, vol. 127, no. 4, pp. 2347–2357, 2010.
[16] G. D. Romigh, D. S. Brungart, R. M. Stern, and B. D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 921–930, Aug. 2015.
[17] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, "Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions," *EURASIP J. Audio, Speech, Music Process.*, vol. 2019, no. 1, p. 5, Mar. 2019. [Online]. Available: https://doi.org/10.1186/s13636-019-0148-x
[18] D. L. Alon, Z. Ben-Hur, B. Rafaely, and R. Mehra, "Sparse head-related transfer function representation with spatial aliasing cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 6792–6796.
[19] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *J. Acoust. Soc. America*, vol. 133, no. 5, pp. 2711–2721, 2013.
[20] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *J. Acoust. Soc. America*, vol. 141, no. 6, pp. 4087–4096, 2017.
[21] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1003–1010, Mar. 2007.
[22] Z. Ben-Hur, J. Sheaffer, and B. Rafaely, "Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction," *Appl. Acoust.*, vol. 134, pp. 138–144, 2018.
[23] J. Zaar, F. Zotter, and M. Noisternig, "Phase unwrapping on the sphere for directivity functions and HRTFs," in *Proc. 38th Annu. Conv. Acoust.*, 2012, pp. 701–702.
[24] E. Rasumow *et al.*, "Smoothing individual head-related transfer functions in the frequency and spatial domains," *J. Acoust. Soc. America*, vol. 135, no. 4, pp. 2012–2025, 2014.
[25] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *J. Acoust. Soc. America*, vol. 143, no. 6, pp. 3616–3627, 2018.
[26] F. Brinkmann and S. Weinzierl, "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition," in *Proc. Audio Eng. Soc. Conf.: AES Int. Conf. Audio Virtual Augmented Reality*, 2018, pp. 1–10.

[27] J.-G. Richter, M. Pollow, F. Wefers, and J. Fels, "Spherical harmonics based HRTF datasets: Implementation and evaluation for real-time auralization," *Acta Acustica United Acustica*, vol. 100, no. 4, pp. 667–675, 2014.
[28] M. Aussal, F. Alouges, and B. Katz, "A study of spherical harmonics interpolation for HRTF exchange," in *Proc. Meetings Acoust.*, vol. 19, no. 1, 2013, Art. no. 050010.
[29] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *J. Acoust. Soc. America*, vol. 127, no. 2, pp. 823–828, 2010.
[30] B. F. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *J. Acoust. Soc. America*, vol. 135, no. 6, pp. 3530–3540, 2014.
[31] A. Andreopoulou and B. F. Katz, "Identification of perceptually relevant methods of inter-aural time difference estimation," *J. Acoust. Soc. America*, vol. 142, no. 2, pp. 588–598, 2017.
[32] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized HRTF fitting using spherical harmonics," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2009, pp. 257–260.
[33] G. Arfken, H. Weber, and F. Harris, *Mathematical Methods for Physicists: A Comprehensive Guide*. New York, NY, USA: Elsevier, 2012. [Online]. Available: https://books.google.com/books?id=qLFo_Z-PoGIC
[34] B. Rafaely, *Fundamentals of Spherical Array Processing*, vol. 8. Berlin, Germany: Springer, 2015.
[35] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. New York, NY, USA: Academic, 1999.
[36] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 697–707, Sep. 2001.
[37] I. Ben Hagai, M. Pollow, M. Vorländer, and B. Rafaely, "Acoustic centering of sources measured by surrounding spherical microphone arrays," *J. Acoust. Soc. America*, vol. 130, no. 4, pp. 2003–2015, 2011.
[38] M. Burkhard and R. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. America*, vol. 58, no. 1, pp. 214–222, 1975.
[39] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the neumann KU 100," in *Proc. 40th Italian Annu. Conf. Acoust. 39th German Annu. Conf. Acoust. Conf. Acoust.*, 2013, pp. 592–595.
[40] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. America*, vol. 104, no. 5, pp. 3048–3058, 1998.
[41] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. America*, vol. 116, no. 4,xbrk pp. 2149–2157, 2004.
[42] V. Lebedev and D. Laikov, "A quadrature formula for the sphere of the 131st algebraic order of accuracy," *Doklady. Math.*, vol. 59, no. 3, 1999, pp. 477–481.
[43] M. Dinakaran *et al.*, "Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 551–555.
[44] F. Brinkmann and S. Weinzierl, "AKtools - an open software toolbox for signal acquisition, processing, and inspection in acoustics," in *Proc. 142th Audio Engineering Society Conv.*, 2017, pp. 1–6.
[45] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," *J. Audio Eng. Soc.*, vol. 67, no. 9, pp. 705–718, 2019.
[46] B. Fabian *et al.*, "The HUTUBS head-related transfer function (HRTF) database," 2019. [Online]. Available: http://dx.doi.org/10.14279/depositonce-8487
[47] I. H. Sloan and R. S. Womersley, "Extremal systems of points and numerical integration on the sphere," *Adv. Comput. Math.*, vol. 21, no. 1/2, pp. 107–125, 2004.
[48] E. A. Macpherson and A. T. Sabin, "Binaural weighting of monaural spectral cues for sound localization," *J. Acoust. Soc. America*, vol. 121, no. 6, pp. 3677–3688, 2007.
[49] A. Lindau, T. Hohn, and S. Weinzierl, "Binaural resynthesis for comparative studies of acoustical environments," in *Proc. 122th Audio Engineering Society Conv.*, 2007, pp. 1–10.
[50] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder–mead simplex method in low dimensions," *SIAM J. Optim.*, vol. 9, no. 1, pp. 112–147, 1998.

Zamir Ben-Hur received the B.Sc. degree (*summa cum laude*) and the M.Sc. degree in electrical and computer engineering in 2015 and 2017, respectively, from Ben-Gurion University of the Negev, Beer-Sheva, Israel, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His current research interests include audio signal processing for binaural reproduction with improved spatial perception. He is a recipient of the Ben-Gurion University High-Tech fellowship.

David Lou Alon received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Ben-Gurion University, Beer-Sheva, Israel, in 2009, 2013, and 2017, respectively. He is currently working as a Research Scientist with the Facebook Reality Labs, investigating efficient representations of head-related transfer functions, and headphone equalization for spatial audio application.

Ravish Mehra received the Ph.D. degree in computer science from the University of North Carolina, Chapel Hill, NC, USA, in the field of acoustics and spatial audio. He is the Research Science Lead for the Audio Team, Facebook Reality Labs. His team is responsible for research and advanced development of new audio techniques to push the state-of-the-art for audio in VR and AR. In his doctoral work, he worked on novel physically based simulation techniques for simulating complex acoustic phenomena arising out of propagation of sound waves in large environments. His research interests span the fields of audio, acoustics, signal processing, and virtual and augmented reality. His work in acoustics and spatial audio has generated considerable interest in the audio community, and his sound propagation and spatial sound system has been integrated into virtual reality systems (Oculus HMD), with demonstrated benefits.

Boaz Rafaely (SM'01) received the B.Sc. degree (*cum laude*) in electrical engineering from Ben-Gurion University, Beer-Sheva, Israel, in 1986, the M.Sc. degree in biomedical engineering from Tel-Aviv University, Tel Aviv, Israel, in 1994, and the Ph.D. degree from the Institute of Sound and Vibration Research (ISVR), Southampton University, Southampton, U.K., in 1997. At the ISVR, he was appointed Lecturer in 1997 and Senior Lecturer in 2001, working on active control of sound and acoustic signal processing. In 2002, he spent six months as a Visiting Scientist with the Sensory Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, investigating speech enhancement for hearing aids. He then joined the Department of Electrical and Computer Engineering at Ben-Gurion University as a Senior Lecturer in 2003, and appointed Associate Professor in 2010, and Professor in 2013. He is currently heading the acoustics laboratory, investigating methods for audio signal processing and spatial audio. During 2010–2014, he has served as an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING, and during 2013–2018 as a member of the IEEE Audio and Acoustic Signal Processing Technical Committee. He also served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS during 2015–2019, *IET Signal Processing* during 2016–2019, and currently for Acta Acustica united with Acustica. During 2013–2016, he has served as the Chair of the Israeli Acoustical Association, and is currently chairing the Technical Committee on Audio Signal Processing in the European Acoustical Association. He was awarded the British Councils Clore Foundation Scholarship.