Audio Engineering Society

Convention Paper 9892

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

A method for efficiently calculating head-related transfer functions directly from head scan point clouds

Rahulram Sridhar and Edgar Y. Choueiri

3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, NJ, 08544, USA

Correspondence should be addressed to Rahulram Sridhar (rahulram@princeton.edu)

ABSTRACT

A method is developed for efficiently calculating head-related transfer functions (HRTFs) directly from head scan point clouds of a subject using a database of HRTFs, and corresponding head scans, of many subjects. Consumer applications require HRTFs be estimated accurately and efficiently, but existing methods do not simultaneously meet these requirements. The presented method uses efficient matrix multiplications to compute HRTFs from spherical harmonic representations of head scan point clouds that may be obtained from consumer-grade cameras. The method was applied to a database of only 23 subjects, and while calculated interaural time difference errors are found to be above estimated perceptual thresholds for some spatial directions, HRTF spectral distortions up to 6 kHz fall below perceptual thresholds for most directions.

1 INTRODUCTION

One of the challenges associated with spatial sound reproduction using synthesized binaural signals is acquisition of a listener's head-related transfer functions (HRTFs) accurately, efficiently, and conveniently. Several methods exist for estimating individualized HRTFs, since direct measurement in an anechoic chamber is infeasible for commercial implementation [1, Ch. 3]. For a comprehensive review, see Blauert [1], Xie [2], and references therein.

Amongst the methods, those that use 3D morphological scans of a listener to estimate HRTFs can be accurate, but suffer from computational inefficiency and implementation complexity. The most common approach to estimate HRTFs from morphological scans is by using the boundary element method to numerically solve the wave equation on a meshed surface of the scan [2, Ch. 4]. This method is limited by its high computational cost and need for an accurate 3D mesh, which requires careful capture and extensive post-processing of raw scan data [3, 4]. Time-domain techniques such as the finite-difference time-domain method [5] and adaptive rectangular decomposition [6] have also been used. However, these techniques also either suffer from computational inefficiency [4, Table I], or estimation inaccuracy (see, for example, Fig. 2 in Meshram et al. [6] where important spectral features are sometimes not captured, even at low frequencies).

Tao et al. [7] use 11^{th}-degree spherical harmonics to represent the shape of the KEMAR dummy head and

a boundary element method to compute HRTFs for frequencies below 3 kHz. This method is prone to inefficiency at higher frequencies, and still requires a meshed scan. More recently, Politis et al. [8] assign the same set of measured interaural time differences (ITDs) to listeners with similar head shapes that are identified by one of three types of spherical transformations of meshed scans. However, while this method can be extended to obtain HRTF estimates, it remains inaccurate for listeners who do not have head shapes that are similar to those of other listeners.

To address these issues, we present a method that allows computation of HRTFs directly from spherical harmonic representations of head scan point clouds that may be conveniently obtained from consumer-grade cameras. Our method does not require sophisticated mesh generation techniques and is computationally efficient, requiring only matrix multiplications.

In the rest of the paper, we formulate our method in Sec. 2, apply and validate it in Sec. 3, and conclude in Sec. 4.

2 METHOD FORMULATION

2.1 Definitions and conventions

To represent ITDs, HRTF magnitudes, and head scan point clouds, we use real-valued spherical harmonics of degree $n \geq 0$ and order $m \in [-n, n]$ given by

$$Y_n^m(\theta,\phi) = N_n^{|m|} P_n^{|m|}(\sin\phi) \times \begin{cases} \cos m\theta & \text{for } m \geq 0, \\ \sin |m|\theta & \text{for } m < 0, \end{cases}$$

where $P_n^{|m|}$ is the associated Legendre polynomial of degree n and order $|m|$, and N_n^m is a normalization term given by

$$N_n^m = (-1)^m \sqrt{\frac{(2n+1)(2-\delta_m)}{4\pi} \frac{(n-m)!}{(n+m)!}},$$

where δ_m is the Kronecker delta. Furthermore, $\theta \in [0^\circ, 360^\circ)$ denotes azimuth, and $\phi \in [-90^\circ, 90^\circ]$ elevation, following the coordinate systems described in AES69-2015 [9], and illustrated in Fig. 1. The radial distance, r, from the origin is given by $r = \sqrt{x^2+y^2+z^2}$.

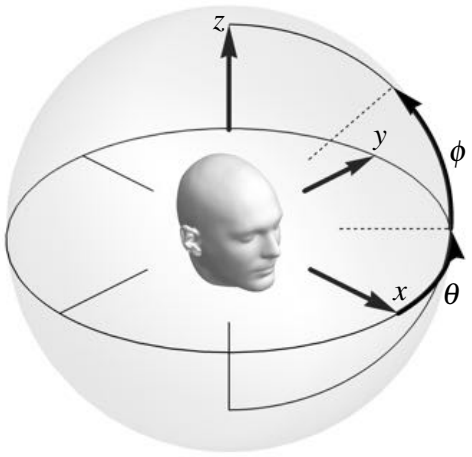

Fig. 1: Coordinate systems used here, and specified by AES69-2015 [9].

.

2.2 Spherical harmonic representation of functions sampled on a sphere

Given the column vector,

$$\mathbf{s} = [S(\theta_1,\phi_1), S(\theta_2,\phi_2), \cdots, S(\theta_V,\phi_V)]^T,$$

of V spatial-samples of a band-limited function, S, defined on the surface of a sphere, and the vector of spherical harmonics,

$$\mathbf{y}_n^m = [Y_n^m(\theta_1,\phi_1), Y_n^m(\theta_2,\phi_2), \cdots, Y_n^m(\theta_V,\phi_V)]^T, \quad (1)$$

we can determine the projection, $\mathbf{s}_P$, of $\mathbf{s}$ onto the column space of $\mathbf{Y} := [\mathbf{y}_0^0, \mathbf{y}_1^{-1}, \cdots, \mathbf{y}_n^n]$ as

$$\mathbf{s}_P = \mathbf{Y}\left(\mathbf{Y}^T\mathbf{Y}\right)^{-1}\mathbf{Y}^T\mathbf{s} = \mathbf{Y}\,\mathbf{c}, \quad (2)$$

where $\mathbf{c}$ is a vector of so-called spherical harmonic coefficients, and $\mathbf{Y}^T$ denotes the transpose of $\mathbf{Y}$. $\mathbf{s}_P$ is the spherical harmonic representation of $\mathbf{s}$.

2.2.1 HRTF magnitude and ITD

Romigh et al. [10] show that, compared to complex-valued HRTFs or their absolute magnitudes, HRTF magnitudes in dB are most efficiently represented using spherical harmonics. Therefore, in this work, we use Eq. (2) with $\mathbf{s}$ consisting of HRTF magnitudes in dB for each frequency, and $n = n_H$, where n_H is the

largest spherical harmonic degree used for representing these magnitudes, to compute $\mathbf{c} = \mathbf{c}_{H_{l/r}}[k]$, where the subscript is used to identify either the left- or right-ear HRTF magnitude. $k \in [0, N_f - 1]$ is the frequency-index with N_f being the number of frequency samples in each HRTF magnitude spectrum, and is proportional to frequency, f, given by $f = kF_s/N_f$, where F_s is the sampling rate. A similar calculation is performed by replacing HRTF magnitudes with measured ITD values to compute $\mathbf{c} = \mathbf{c}_I$.

2.2.2 Head scan point clouds

To represent head scan point clouds in terms of spherical harmonics, we choose $\mathbf{s}$ to be the vector of Euclidean distances, r, of each of the points (each associated with a unique spatial direction) from the coordinate system origin. We then use a similar procedure, as described in Sec. 2.2.1, to compute the vector, $\mathbf{c}_S$, of point cloud spherical harmonic coefficients.

2.3 Computation of HRTFs from scans

Let $u = 1, 2, \ldots, U$ identify each of U subjects, and let $\mathbf{c}^{(u)}$ denote the vector $\mathbf{c}$ associated with subject u. Define matrices $\mathbf{D}_S$, $\mathbf{D}_{H_{l/r}}[k]$, and $\mathbf{D}_I$ of spherical harmonic coefficients such that

$$\mathbf{D}_S = \left[\mathbf{c}_S^{(1)}, \ldots, \mathbf{c}_S^{(U)}\right]^T,$$
$$\mathbf{D}_{H_{l/r}}[k] = \left[\left(\mathbf{c}_{H_{l/r}}^{(1)}[k]\right), \ldots, \left(\mathbf{c}_{H_{l/r}}^{(U)}[k]\right)\right]^T,$$
$$\mathbf{D}_I = \left[\mathbf{c}_I^{(1)}, \ldots, \mathbf{c}_I^{(U)}\right]^T,$$

where $\mathbf{c}_S$, $\mathbf{c}_{H_{l/r}}[k]$, and $\mathbf{c}_I$ for all U subjects are obtained as described in Secs. 2.2.1 and 2.2.2.

We seek a matrix $\mathbf{X}_{H_{l/r}}[k]$ whose columns are the projections of the corresponding columns of $\mathbf{D}_{H_{l/r}}[k]$ on the column space of $\mathbf{D}_S$, and a column vector $\mathbf{X}_I$ which is the projection of $\mathbf{D}_I$, also on this space. $\mathbf{X}_{H_{l/r}}[k]$ is given by

$$\mathbf{X}_{H_{l/r}}[k] = \left(\mathbf{D}_S^T \mathbf{D}_S\right)^{-1} \mathbf{D}_S^T \mathbf{D}_{H_{l/r}}[k], \tag{3}$$

and $\mathbf{X}_I$ may be computed similarly using $\mathbf{D}_I$ instead of $\mathbf{D}_{H_{l/r}}[k]$.

The matrix $\mathbf{X}_{H_{l/r}}[k]$ and vector $\mathbf{X}_I$ may then be used to compute $\mathbf{c}_{H_{l/r}}[k]$ and $\mathbf{c}_I$, from $\mathbf{c}_S$, using the relations

$$\mathbf{c}_{H_{l/r}}[k] = \mathbf{X}_{H_{l/r}}^T[k]\mathbf{c}_S \quad \text{and} \quad \mathbf{c}_I = \mathbf{X}_I^T \mathbf{c}_S, \tag{4}$$

respectively, for any subject.

Finally, HRTF magnitudes for a set of directions (θ_i, ϕ_i), $i = 1, 2, \ldots, V$ may then be computed using Eq. (2) with $\mathbf{c} = \mathbf{c}_{H_{l/r}}[k]$ from Eq. (4), and with $n = n_H$. The resulting $\mathbf{s}_P$ is the vector of desired magnitudes at a given frequency. These calculations may be repeated for each frequency to compute complete HRTF magnitude spectra. ITDs may be computed similarly by using $\mathbf{c}_I$ instead of $\mathbf{c}_{H_{l/r}}[k]$ and setting $n = n_I$. Complete HRTFs may be derived by determining the minimum-phase HRIRs from the HRTF magnitude spectra, and introducing the appropriate computed ITDs.

2.3.1 Comparison with anthropometry-based regression methods

Like other anthropometry-based methods, we assume relationships exist between a listener's HRTFs and anthropometric features. However, we further assume that these relationships may be found by representing both HRTF and anthropometric data using the same kind of basis functions (spherical harmonics in this case). This assumption distinguishes our method from other anthropometry-based methods (see Nishino et al. [11], for example), where such relationships are sought between measured anthropometric features represented in Euclidean space with naive Cartesian basis, and HRTF features represented in some arbitrarily chosen, typically principal component, space. Since our method, like other anthropometry-based methods, is data-driven, we assume that convergence to the aforementioned relationships will be achieved with a sufficiently large amount of data. However, unlike our method, those that represent HRTF features on a principal component space, for example, rely on convergence not only to these relationships, but also to a set of basis functions, since the principal component space is typically a function of the data used to develop the method. Furthermore, while previous anthropometry-based HRTF methods are based on the difficult task of first identifying and measuring specific anthropometric features of a subject [2, Ch. 7], our method, in comparison, does not require such features to be explicitly identified.

3 APPLICATION AND VALIDATION

3.1 Data acquisition and pre-processing

To apply and validate our method, we use the measured HRTFs and head scans of 25 subjects from the

RIEC[1] [12] database. For each subject, we normalize HRTFs such that the largest magnitude at 468.75 Hz, considering HRTFs for all spatial directions, is 0 dB. Following Romigh et al. [10], we make all HRIRs minimum-phase and truncate each to a duration of 5.8 ms. Following Andreopoulou and Katz [13], we compute ITDs by low-pass filtering the original, non-minimum-phase HRIRs using a 3 kHz cut-off frequency, identifying left- and right-HRIR onsets using a 30% threshold[2], and subtracting right-HRIR onsets from corresponding left-HRIR onsets.

We resample all head scan point cloud data to contain 5000 to 6000 roughly evenly distributed points, and align each scan such that the y-axis (see Fig. 1) is also approximately the interaural axis, with the coordinate system origin located approximately halfway between the entrances to the two ear canals. We do not close any holes or do any processing on the scan besides alignment, basic noise reduction (outlier removal), and resampling.

3.2 Application

To apply our method, we compute $\mathbf{X}_{H_{l/r}}[k]$ and $\mathbf{X}_I$ using 23 "training" subjects. Due to the limited training data, we are restricted to $n_S \leq 3$ to ensure that the matrix $\mathbf{D}_S$ is not "fat" (i.e. one with fewer rows than columns). We choose $n_S = 2$ to compute HRTF magnitudes since we found this produces the best results for the amount of training data available. We choose $n_S = 1$ to compute ITD because this roughly corresponds to an ellipsoidal-approximation to the subject's head shape, and simple geometrical methods of the head have previously been successfully used to compute ITD [14, 15].

Figure 2 shows average RMS error, $\xi_l[k]$, computed here in dB as a function of n_H and f. $\xi_l[k]$ is given by

$$\xi_l[k] = \frac{1}{U}\sum_{u=1}^{U}\sqrt{\frac{1}{V}\sum_{v=1}^{V}\left|\xi_{v_l}^{(u)}[k]\right|^2}, \tag{5}$$

where

$$\xi_{v_l}^{(u)}[k] = \left|20\log_{10}\frac{\left|H_{v_l}^{(u)}[k]\right|}{\left|\hat{H}_{v_l}^{(u)}[k]\right|}\right| \tag{6}$$

[1] http://www.riec.tohoku.ac.jp/pub/hrtf/index.html

[2] Although Andreopoulou and Katz [13] recommend using 3.16%, we use 30% because many of the HRIRs contain undesirable pre-responses that result in incorrectly identified onsets when using 3.16%.

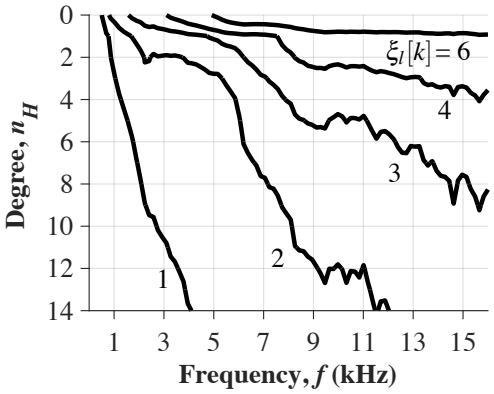

Fig. 2: Contour plot of $\xi_l[k]$, in dB, as a function of n_H and f.

denotes absolute spectral distortion in dB, with $|\cdot|$ denoting absolute value. $H_{v_l}^{(u)}[k]$ and $\hat{H}_{v_l}^{(u)}[k]$ denote, respectively, the measured and projected, left-ear HRTFs for subject u and direction v, and RMS averaging is performed over $U = 25$ subjects and all $V = 865$ spatial directions.

We see that $\xi_l[k]$ for $n_H = 6$ is similar to RMS errors shown in Fig. 5 published by Romigh et al. [10] for a "truncation order" of four, the minimum for imperceptible spherical harmonic HRTF representations [10]. Therefore, we choose $n_H = 6$ for computing HRTF magnitudes, and also to make inferences about the perceptibility of errors induced by our method.

Figure 3 shows a plot of average absolute ITD error, ε, as a function of n_I when averaging over all subjects, and either over all 865 directions or just the 72 directions on the horizontal plane (i.e., $\phi = 0°$). ε is given by

$$\varepsilon = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}\varepsilon_v^{(u)}, \tag{7}$$

where

$$\varepsilon_v^{(u)} = \left|\tau_v^{(u)} - \hat{\tau}_v^{(u)}\right|, \tag{8}$$

is the absolute ITD error, and $\tau_v^{(u)}$ and $\hat{\tau}_v^{(u)}$ denote, respectively, the measured and projected ITDs for subject u and direction v.

We see that ε decreases significantly between $n_I = 0$ and $n_I = 1$, whereas beyond $n_I = 3$, the decrease is slow. Therefore, we choose $n_I = 3$ for computing ITDs.

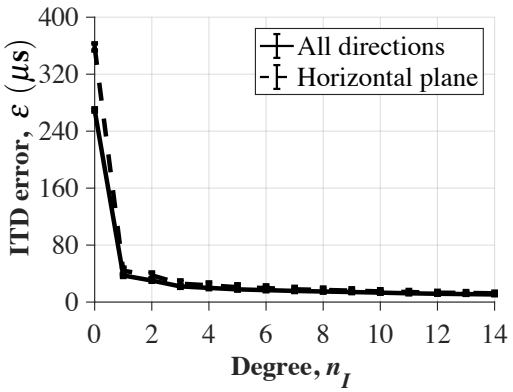

Fig. 3: Plot of ε as a function of n_I, averaged over all 25 subjects, and either over all 865 spatial directions (solid line), or just the 72 directions on the horizontal plane (dashed line). Error bars (too small to be visible) represent standard error of the mean when averaging over subjects.

3.3 Validation

We validate our method using two "test" subjects from the database whose data were not used to compute $\mathbf{X}_{H_{l/r}}[k]$ and $\mathbf{X}_I$.

To evaluate the accuracy of the computed HRTF magnitudes for these subjects, we compute, for the left-ear HRTF magnitudes of each subject, log-weighted average spectral distortion, $\xi_{v_l}^{(u)}$, in dB, given by

$$\xi_{v_l}^{(u)} = \frac{\sum_{k=K_1}^{K_2} \chi[k] \cdot \xi_{v_l}^{(u)}[k]}{\sum_{k=K_1}^{K_2} \chi[k]}, \qquad (9)$$

where the weights, $\chi[k]$, are given by

$$\chi[k] = \log \frac{k+0.5}{k-0.5}.$$

$\xi_{v_l}^{(u)}[k]$ is computed as shown in Eq. (6) with $\hat{H}_{v_l}^{(u)}[k]$ now denoting the HRTFs computed using our method. K_1 and K_2 are frequency indices corresponding to lower- and upper-frequency bounds over which averaging is performed. We average over each of three frequency bands given by: (i) $0 < f < 2$ kHz, (ii) $2 < f < 8$ kHz, and (iii) $8 < f < 16$ kHz. We also compute RMS values of $\xi_{v_l}^{(u)}$, averaging over all spatial directions, in each frequency band, and compare these values with the RMS error, $\xi_l[k]$, shown in Fig. 2, to approximately determine the perceptibility of $\xi_{v_l}^{(u)}$.

To evaluate the accuracy of computed ITDs, we compute $\varepsilon_v^{(u)}$ as a function of θ and ϕ using Eq. (8), with $\hat{\tau}_v^{(u)}$ now denoting the ITDs computed using our method. To determine the perceptibility of $\varepsilon_v^{(u)}$, we use a threshold ITD error of approximately 30 μs, which we compute using the Woodworth and Schlosberg formula [16] for estimating ITD for a spherical-head with radius 0.0875 m, using a localization blur angle of 3.2° [17, Table 2.1]. Although this 30 μs threshold is only applicable for broadband white noise stimuli, and may not be applicable to all spatial directions, we use it as an *approximate* threshold for the perceptibility of $\varepsilon_v^{(u)}$.

Figure 4 shows matrix plots of $\xi_{v_l}^{(u)}$ as a function of θ and ϕ for each of the test subjects. Also shown on these plots are RMS values of $\xi_{v_l}^{(u)}$ for each frequency band, when averaging over all spatial directions. We see that $\xi_{v_l}^{(u)} < 2$ dB for all spatial directions when averaging over $0 < f < 2$ kHz, and $\xi_{v_l}^{(u)} > 4$ dB between 2 and 8 kHz for $\phi < 30°$ and $|\theta - 270°| < 30°$ for both subjects. Similar calculations (not shown here) for the right-ear HRTFs result in similar errors for $|\theta - 90°| < 30°$. This suggests that the current implementation of our method is not accurate for computing contralateral HRTFs for some spatial directions in this frequency range. Furthermore, the similarity in results across subjects for $f < 8$ kHz indicates that the performance of our method for these frequencies is reliable, and we expect similar performance for different test subjects. Additional calculations (not shown here) made by varying the choice of training and test subjects confirm this observation. Finally, when averaging over $8 < f < 16$ kHz, $\xi_{v_l}^{(u)} > 7$ dB for most spatial directions. This shows that the current implementation of our method is not accurate when computing HRTF spectral features in this frequency range. This is discussed further in Sec. 4.

From a perceptual standpoint, comparing the RMS values of $\xi_{v_l}^{(u)}$ to those plotted in Fig. 2 for $n_H = 6$, we see that, for the range $0 < f < 8$ kHz, the RMS errors in the computed HRTFs are similar to those for sixth-degree spherical harmonic representations of measured HRTFs. Based on the discussion in Sec. 3.2, we may

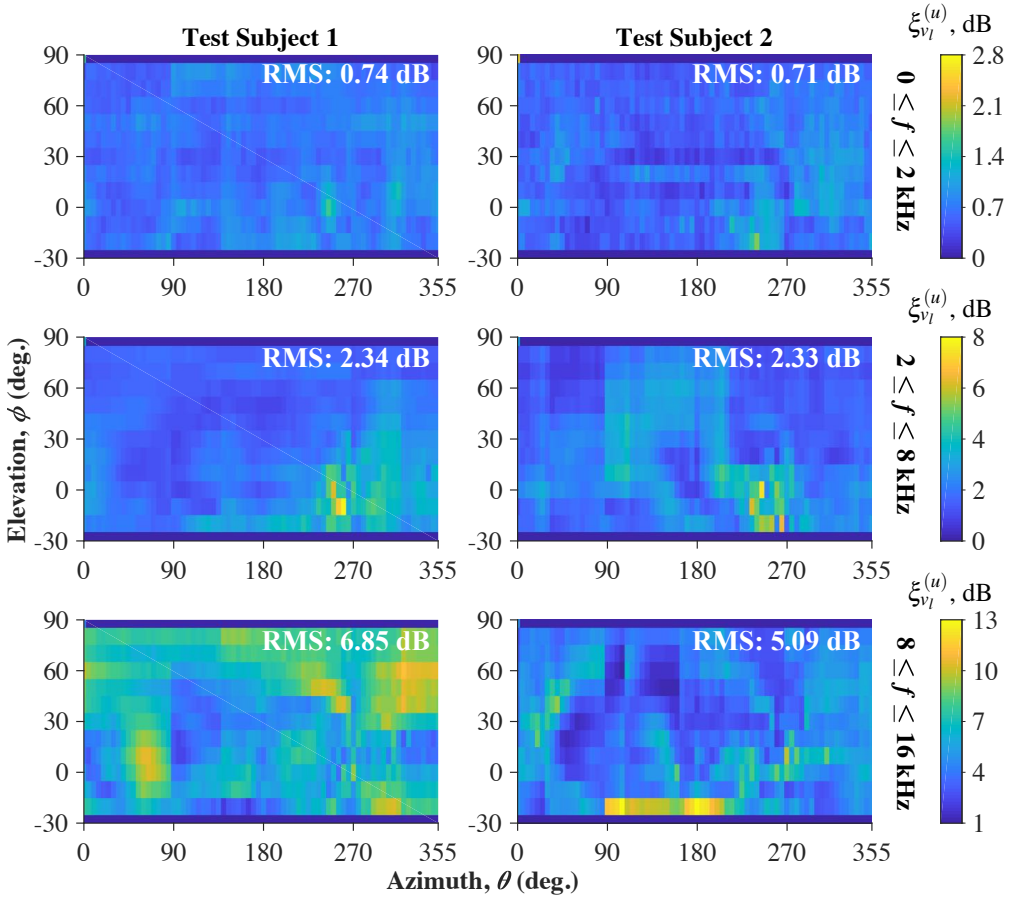

Fig. 4: Matrix plots of $\xi_{v_l}^{(u)}$ as a function of θ and ϕ for each of the test subjects. Note that values of $\xi_{v_l}^{(u)}$ for elevations $\phi = -30^\circ$ and $\phi = 90^\circ$ are not shown because sufficient data points for different θ do not exist at these elevations.

conclude that any errors introduced by our method for $f < 8$ kHz are likely imperceptible.

Figure 5 shows matrix plots of $\xi_{v_l}^{(u)}[k]$ as a function of ϕ on the median plane (i.e. $\theta = 0^\circ$ and $\theta = 180^\circ$) for each of the test subjects. For both subjects, $\xi_{v_l}^{(u)}[k] < 3$ dB for most ϕ and for $f < 6$ kHz. Also, for subject 1, $\xi_{v_l}^{(u)}[k]$ ranges from 9 to 15 dB for $7 < f < 8$ kHz and $0^\circ < \phi < 60^\circ$, suggesting that the current implementation of our method is unable to accurately capture spectral features caused by the pinnae, since these features are typically found in this frequency range [18]. Figure 6, which shows measured and computed HRTF magnitude spectra, in dB, for $\phi = 30^\circ$ and $\phi = -10^\circ$ on the median plane of subject 1 and 2, respectively, further illustrates this. We explain this discrepancy by noting that a second-degree spherical harmonic representation of the head scan is insufficient to capture shape variations in the subject's pinnae that might cause these spectral features in the HRTFs.

Finally, Fig. 7 shows a matrix plot of $\varepsilon_v^{(u)}$ as a function of θ and ϕ for each of the test subjects. We see that $\varepsilon_v^{(u)} < 30$ μs (the approximate perceptibility threshold computed earlier) for most ϕ near the median plane, except $\phi \geq 30^\circ$ behind (i.e. $\theta = 180^\circ$) subject 1. Additionally, $60 < \varepsilon_v^{(u)} < 110$ μs for $|\theta - 90^\circ| < 30^\circ$ and $|\theta - 270^\circ| < 30^\circ$, thus exceeding the perceptible threshold significantly. One possible explanation for these errors is that a large number of measured HRIRs

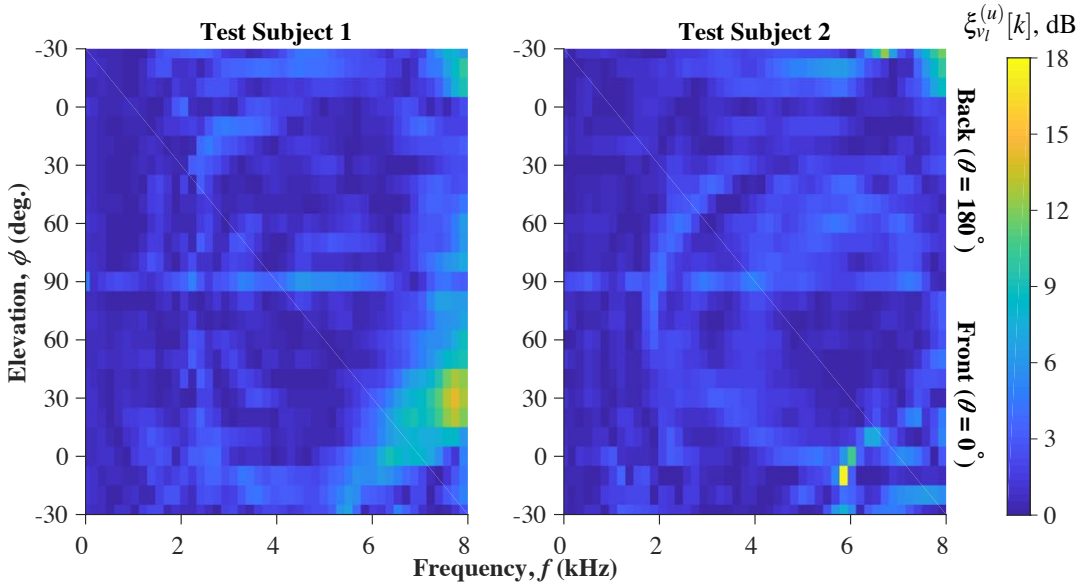

Fig. 5: Matrix plots of $\xi_{v_l}^{(u)}[k]$ as a function of ϕ on the median plane for each of the test subjects.

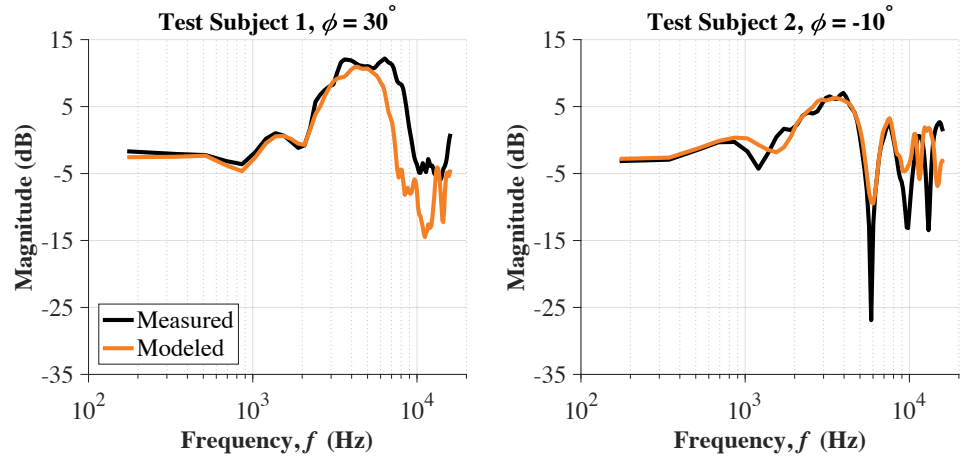

Fig. 6: Line plots of left-ear HRTF magnitude spectra in dB for $\phi = 30°$ and $\phi = -10°$ on the median plane for test subjects 1 (left) and 2 (right), respectively. Note that these plots are for the worst case, which corresponds to the region of highest error shown in Fig. 5.

from the database contained artifacts prior to the main impulse, and the thresholding method used here for estimating measured ITD, as described in Sec. 3.1, is not robust to such artifacts. This, however, needs further investigation. Currently, our method is unable to estimate ITDs accurately for some spatial directions, particularly those corresponding to large θ and ϕ.

4 CONCLUSIONS

We presented a method for computing HRTFs from low-resolution head scan point cloud data of a subject by

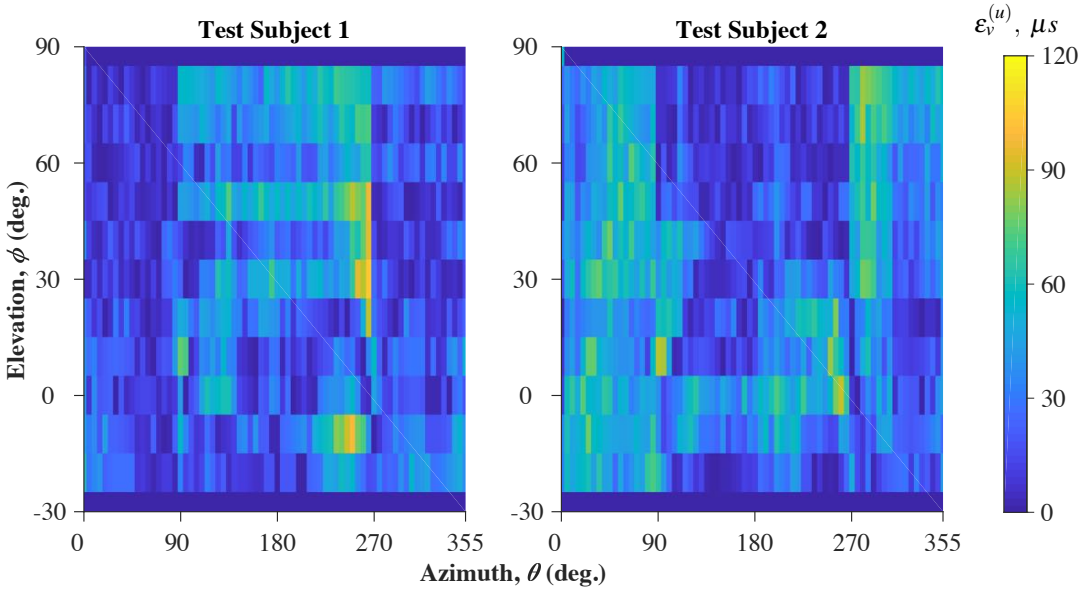

Fig. 7: Matrix plot of $\varepsilon_v^{(u)}$ as a function of θ and ϕ for each of the test subjects. Note that values of $\varepsilon_v^{(u)}$ for elevations $\phi = -30^\circ$ and $\phi = 90^\circ$ are not shown because sufficient data points for different θ do not exist at these elevations.

first projecting the point cloud data onto a vector space spanned by spherical harmonics, and then mapping the resulting spherical harmonic coefficients to ones that represent the subject's HRTFs in the same space. We defined this mapping as the projection of spherical harmonic coefficients computed from measured HRTFs for a set of "training" subjects, onto the column space of a matrix with columns containing the spherical harmonic coefficients computed from the head scan point cloud data of these subjects (see Sec. 2.3 for details). Separate mappings were derived for computing a subject's ITD in μs and HRTF magnitude spectra in dB.

We then presented an implementation of our method using 23 training subjects to derive the mappings and validated the method using 2 "test" subjects. We validated our method in terms of spectral distortion and root-mean-square (RMS) errors of HRTF magnitudes, and also in terms of absolute ITD errors. Both these metrics were related to perceptibility criteria to approximately determine the perceptibility of the errors induced by our method.

We show that our method, when implemented using only 23 training subjects, is able to reliably compute HRTF magnitudes up to approximately 6 kHz, and that errors in this range may be imperceptible for most spatial directions. This shows that a second-degree spherical harmonic representation head scan point clouds may be sufficient to capture the necessary head shape information that influences HRTF magnitude spectra up to 6 kHz. We also conclude that for $f > 6$ kHz, the current implementation of our method does not perform reliably, and any estimation errors that may not be perceptible (for instance, see, in Fig. (5), $\xi_{v_l}^{(u)}[k] < 3$ dB between 6 to 8 kHz for median plane HRTFs for subject 2) are only due to chance. Higher-degree spherical harmonic representations of head scans are likely necessary to improve performance of the method for these frequencies. This, in turn, requires a larger training data set.

Finally, we note that the above results are achieved using unmeshed head scans with no more than 6000 data points, and with minimal pre-processing involving only resampling, basic noise reduction, and alignment. We also note that HRTF computations using our method involves only a few, computationally efficient matrix multiplications. Our method is, therefore, well-suited for use in consumer spatial audio applications.

Acknowledgements

We wish to thank Joseph Tylka for reviewing the manuscript and for some of the MATLAB code used to generate the results shown here. We also wish to thank the anonymous reviewers for their comments on the précis submission related to this paper. This work was sponsored by the Sony Corporation of America.

References

[1] Blauert, J., *The technology of binaural listening*, Springer, 2013.

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

[3] Katz, B. F., "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America*, 110(5), pp. 2440–2448, 2001.

[4] Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., and Zotkin, D. N., "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *The Journal of the Acoustical Society of America*, 127(1), pp. 370–386, 2010.

[5] Xiao, T. and Huo Liu, Q., "Finite difference computation of head-related transfer function for human hearing," *The Journal of the Acoustical Society of America*, 113(5), pp. 2434–2441, 2003.

[6] Meshram, A., Mehra, R., and Manocha, D., "Efficient HRTF computation using adaptive rectangular decomposition," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Audio Engineering Society, 2014.

[7] Tao, Y., Tew, A. I., and Porter, S. J., "A study on head-shape simplification using spherical harmonics for HRTF computation at low frequencies," *Journal of the Audio Engineering Society*, 51(9), pp. 799–805, 2003.

[8] Politis, A., Thomas, M. R., Gamper, H., and Tashev, I. J., "Applications of 3D spherical transforms to personalization of head-related transfer functions," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pp. 306–310, IEEE, 2016.

[9] AES69-2015, "AES69-2015: AES standard for file exchange - Spatial acoustic data file format," 2015.

[10] Romigh, G. D., Brungart, D. S., Stern, R. M., and Simpson, B. D., "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE Journal of Selected Topics in Signal Processing*, 9(5), pp. 921–930, 2015.

[11] Nishino, T., Inoue, N., Takeda, K., and Itakura, F., "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, 68(8), pp. 897–908, 2007.

[12] Watanabe, K., Iwaya, Y., Suzuki, Y., Takane, S., and Sato, S., "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoustical science and technology*, 35(3), pp. 159–165, 2014.

[13] Andreopoulou, A. and Katz, B. F., "Identifying a perceptually relevant estimation method of the inter-aural time delay," *The Journal of the Acoustical Society of America*, 141(5), pp. 3635–3635, 2017.

[14] Sridhar, R. and Choueiri, E., "Capturing the elevation dependence of interaural time difference with an extension of the spherical-head model," in *Audio Engineering Society Convention 139*, Audio Engineering Society, 2015.

[15] Duda, R. O., Avendano, C., and Algazi, V. R., "An adaptable ellipsoidal head model for the interaural time difference," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 2, pp. 965–968, IEEE, 1999.

[16] Woodworth, R. S. and Schlosberg, H., *Experimental psychology*, Holt, 1954.

[17] Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT press, 1997.

[18] Hebrank, J. and Wright, D., "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, 56(6), pp. 1829–1834, 1974.