APPLIED METHODS FOR SPARSE SAMPLING OF HEAD-RELATED TRANSFER FUNCTIONS

Lior Arbel[1] *Zamir Ben-Hur*[2] *David Lou Alon*[2] *Boaz Rafaely*[1]

[1] Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

[2] Facebook Reality Labs Research, Facebook, 1 Hacker Way, Menlo Park, CA 94025, USA

ABSTRACT

Production of high fidelity spatial audio applications requires individual head-related transfer functions (HRTFs). As the acquisition of HRTF is an elaborate process, interest lies in interpolating full length HRTF from sparse samples. Ear-alignment is a recently developed pre-processing technique, shown to reduce an HRTF's spherical harmonics order, thus permitting sparse sampling over fewer directions. This paper describes the application of two methods for ear-aligned HRTF interpolation by sparse sampling: Orthogonal Matching Pursuit and Principal Component Analysis. These methods consist of generating unique vector sets for HRTF representation. The methods were tested over an HRTF dataset, indicating that interpolation errors using small sampling schemes may be further reduced by up to 5 dB in comparison with spherical harmonics interpolation.

***Index Terms*—** Spatial audio, head-related transfer functions (HRTFs), spherical-harmonics, principal component analysis, orthogonal matching pursuit.

1. INTRODUCTION

Spatial audio synthesis requires an accurate head-related transfer function (HRTF) - the acoustic transfer function from a sound source to a listener's ear. Generic HRTFs such as the ones measured on dummy heads are often used. However, generic HRTFs often produce poorer results in terms of localization and externalization. Superior spatial audio experiences require individual HRTFs, specifically acquired for a single human listener over a large number of directions. A disadvantage of individually measured HRTFs is their acquisition, being an elaborate and timely process requiring an expensive measurement setup [1–3].

The acquisition of HRTFs may be simplified by obtaining a sparsely sampled HRTF over a few measurement points. The full length HRTF is then interpolated from the samples. One such prominent approach is representing the HRTF using a linear combination of spherical harmonics (SH). Most of the energy is typically contained in the lower SH orders. Thus, the representation may be truncated to include only a small number of low orders, while still maintaining low interpolation errors [4–7]. Lower SH order required for the representation, leads to fewer sparse directions required for measurement. Once a representation is obtained, the HRTF may be interpolated to any desired directions, or further estimated by different methods [8].

Various pre-processing methods were proposed for SH order reduction by concentrating most of the contained energy at low orders [9, 10]. The recently developed ear-alignment method was shown to be especially effective [11]. In this method, the HRTF is phase corrected to each of the subject's ears, rather than the center of its head, as the reference location. The effectiveness of the ear-alignment pre-processing method provides an opportunity for the additional improvement of HRTF sparse sampling.

This paper presents methods for further reducing the required sparse measurements of ear-aligned HRTFs. The first method employs an Orthogonal Matching Pursuit (OMP) algorithm over the SH domain. This method consists of representing the HRTF by any subset of the full series. The second method consists of analyzing an existing HRTF dataset by Principal Component Analysis (PCA). This analysis generates a unique set of spatial vectors which constitutes a best fit to the dataset. The vectors are then used for representation and interpolation in place of spherical harmonics. Both methods require a training phase in which the vector sets are assembled. Early application of PCA to HRTF data was performed by Kistler and Wightman, who also provide an extensive overview of the method [12]. A recent overview of HRTF interpolation by PCA is given by Xie [4]. Various studies had found PCA to be effective in HRTF representation and estimation [13–15]. Existing PCA applications to HRTF interpolation involve non ear-aligned HRTFs.

The results presented here indicate that further reduction of HRTFs sparse sample directions may be achieved. Significant reduction is obtained by PCA, while OMP does not offer an improvement in terms of sparse sampling. However, OMP was found to have potential for an efficient representation. Both methods may be integrated in existing HRTF measurement processes by performing a training phase.

2. HRTF DECOMPOSITION AND INTERPOLATION

This section describes the interpolation of an HRTF using an arbitrary vector basis. Consider an HRTF, $H(k, \Omega)$, where k is the wave number, $\Omega = (\theta, \phi)$ is the angular direction, θ is the elevation angle and ϕ is the azimuth angle. The HRTF may be represented by a basis of complex spherical functions $\{v_1(\Omega), v_2(\Omega), \ldots\}$. It is now assumed that the HRTF is of a limited order over the basis, and is densely sampled at Q directions. The representation therefore consists of a sum of L basis elements and is formulated in matrix form:

$$\mathbf{h} = \mathbf{V}\mathbf{w}, \tag{1}$$

where $\mathbf{h} = [H(k, \Omega_1), \ldots, H(k, \Omega_Q)]^T$ is a $Q \times 1$ vector of HRTF samples over Q directions, $\mathbf{w}$ is the $L \times 1$ HRTF coefficients vector and $\mathbf{V}$ is a $Q \times L$ matrix consisting of the first L basis elements evaluated at Q directions, defined by its qth row as $\{\mathbf{V}\}_q = [v_1(\Omega_q), \ldots, v_L(\Omega_q)]$.

Provided that Q satisfies $Q \geq L$, the coefficients $\mathbf{w}$ may be calculated from the samples by multiplying Eq. (1) by the pseudo-inverse of matrix $\mathbf{V}$, defined as $\mathbf{V}^\dagger = (\mathbf{V}^H\mathbf{V})^{-1}\mathbf{V}^H$, where $(\cdot)^H$ denotes the Hermitian operator:

$$\mathbf{w} = \mathbf{V}^\dagger\mathbf{h}. \tag{2}$$

For a sparsely sampled HRTF at $\tilde{Q}$ directions, $\tilde{\mathbf{h}}$, assuming $\tilde{Q} \geq L$, each of the column vectors in $\mathbf{V}$ are sparse sampled to obtain $\tilde{\mathbf{V}}$, a $\tilde{Q} \times L$ sparse matrix. The coefficients $\hat{\mathbf{w}}$ can be computed from the sparse samples:

$$\hat{\mathbf{w}} = \tilde{\mathbf{V}}^\dagger\tilde{\mathbf{h}}. \tag{3}$$

The HRTF may be interpolated back to Q directions by reformulating Eq. (1):

$$\hat{\mathbf{h}} = \mathbf{V}\hat{\mathbf{w}}, \tag{4}$$

where $\hat{\mathbf{h}}$ is the interpolated HRTF. The interpolation error per frequency is defined as:

$$\epsilon(f) = 10\log_{10}\frac{\|\mathbf{h} - \hat{\mathbf{h}}\|^2}{\|\mathbf{h}\|^2}. \tag{5}$$

The process of interpolation refers to the estimation of the densely sampled $\mathbf{h}$, given sparse samples $\tilde{\mathbf{h}}$ and basis $\mathbf{V}$. Note that the interpolation may suffer from errors if the number of basis vectors, L, or the number of sparse sample, $\tilde{Q}$, is too small.

3. METHODS FOR HRTF INTERPOLATION

The following subsections describe methods for sparse HRTF interpolation using different vector sets. Each method provides an alternative definition of matrix $\mathbf{V}$ for Eqs. (2)–(4).

3.1. Spherical Harmonics Truncation

Spherical harmonics are commonly used for HRTF representation. An HRTF of limited order N is represented by a finite SH expansion. The representation is performed by Eq. (2). The matrix $\mathbf{V}$ consists of $L = (N+1)^2$ vectors, defined by its qth row as: $\{\mathbf{V}\}_q = [Y_0^0(\Omega_q), Y_1^{-1}(\Omega_q) \ldots Y_N^N(\Omega_q)]$, where $Y_n^m(\Omega)$ is the SH function of order n and degree m.

Typically, $Q \geq (N+1)^2$ may no longer be satisfied at high frequencies. In such cases, the HRTF's representation is truncated. The sparse representation is obtained by Eq. (3).

3.2. Orthogonal Matching Pursuit in the Spherical Harmonics Domain

Orthogonal Matching Pursuit (OMP) is an iterative algorithm for the sparse approximation of a signal by a set of vectors [16]. Each iteration consists of selecting a new vector from the set. This vector is appended to all previously selected vectors, termed "*dictionary*", which are used to approximate the signal. With each successive iteration, the dictionary expands and the approximation error decreases.

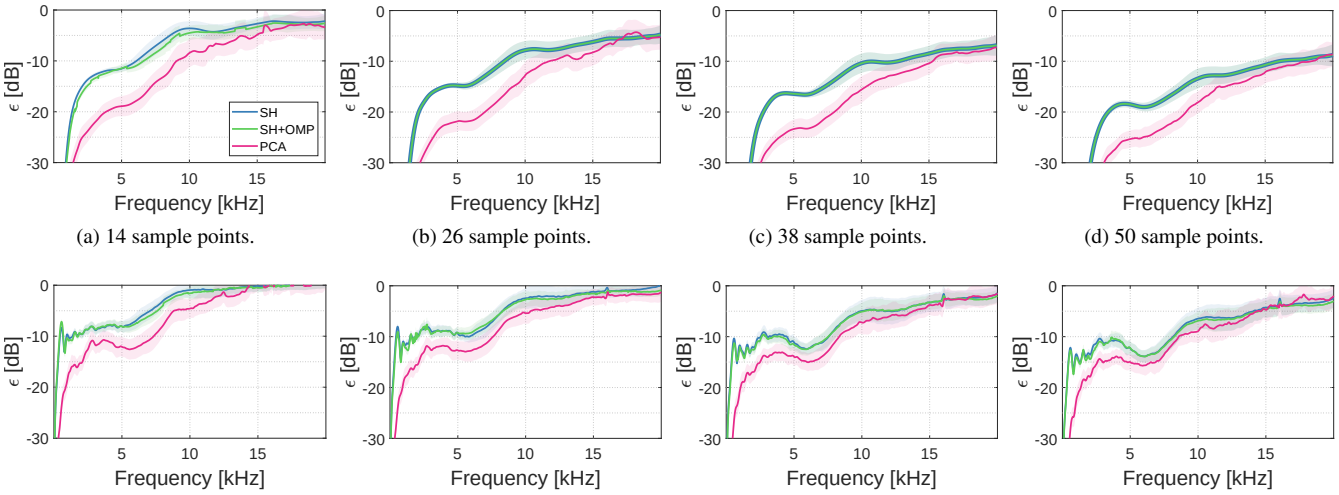
The OMP algorithm is used here to find the most efficient representation of an HRTF by a limited number of SH. OMP cherry-picks the SH of various non-consecutive orders and degrees which best represent the signal. This is opposed to SH truncation, described in Sec. 3.1, in which SH of all orders and degrees are used up to a predefined order. The coefficients are obtained by Eq. (2) using the matrix $\mathbf{V}$ defined by its qth row as $\{\mathbf{V}\}_q = [Y_{n_1}^{m_1}(\Omega_q), Y_{n_2}^{m_2}(\Omega_q) \ldots Y_{n_L}^{m_L}(\Omega_q)]$, where $(n_1, m_1), \ldots, (n_L, m_L)$ are the SH coefficients indices selected by the algorithm.

3.3. Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method for the efficient representation of a data set assumed to be correlated [17]. The method generates a vector set in which each consecutive element contributes the most to the remaining data fit, while being orthogonal to all preceding elements.

Let the $P \times Q$ matrix $\mathbf{D}$ be the data matrix, consisting of a set of P HRTFs of length Q. The matrix is decomposed by SVD as $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where the set of orthogonal column vectors composing $\mathbf{V}$ are the PCA *principal directions* spanning the rows of $\mathbf{D}$, and $\mathbf{U}\mathbf{\Sigma}$ are the *principal components*. Note that often $Q \gg P$, thus the set does not span the entire $\mathbb{C}^Q$ space.

The decomposition coefficients of a sparse sampled HRTF $\tilde{\mathbf{h}}$, are obtained by the truncated and sparse sampled $\mathbf{V}$, as described in Eq. (3). Interpolation to Q directions is then performed by Eq. (4). Note that as common in PCA analysis, matrix $\mathbf{D}$ is mean-centered, and so is $\hat{\mathbf{h}}$. Therefore, the mean vector should be added in Eq. (4).



(a) 14 sample points. (b) 26 sample points. (c) 38 sample points. (d) 50 sample points.

(e) 14 sample points. (f) 26 sample points. (g) 38 sample points. (h) 50 sample points.

Fig. 1: Interpolation error averaged over 36 test set participants compared between the three methods, for both simulated (top) and measured (bottom) HRTFs. Standard deviations are shown in shaded areas. The interpolation used Lebedev sampling schemes of different sizes. Note that in some plots, SH truncation and OMP traces overlap.

4. SPARSE INTERPOLATION - PERFORMANCE EVALUATION

This section presents a performance evaluation of interpolation methods based on sparse sampling, compared to the benchmark method of SH truncation.

4.1. Methodology

The evaluation was performed using simulated and measured HRTFs of the HUTUBS dataset, using 90 human subjects for whom anthropometric data was supplied. The simulated HRTFs consist of a $Q = 1730$ Lebedev grid and the measured HRTFs consist of a $Q = 440$ non-standard grid as described in [18]. The entire dataset was ear-aligned by preprocessing [11], using the head width provided per subject. The 90 subjects were divided to two groups: the *training set*, containing 54 random subjects (60% of the entire dataset), and the *test set*, containing the remaining 36 subjects.

The training set was used by OMP and PCA algorithms, per frequency, for generating unique vector sets as described in Secs. 3.2 and 3.3. In OMP, sets were created by obtaining the coefficients using Eq. (2), averaging over the training set, and selecting the SH with the largest average coefficient per iteration. In PCA, the elements were generated and ordered by the corresponding eigenvalue magnitude. For simple SH truncation, no processing of the training set was required, as the representation vector set is the standard SH basis.

Interpolation errors were calculated on the test set. Each HRTF was down sampled from the original directions using sparse Lebedev schemes of sizes $\tilde{Q} = 14 - 50$. The coefficients $\hat{\mathbf{w}}$ were estimated for each vector set using Eq. (3), and the HRTFs were interpolated to the original directions using Eq. (4). The interpolation error may also be affected by the number of elements, or basis vectors used. The number that generated the lowest error was selected per each method and sampling scheme.

4.2. Results

Average interpolation errors over the test set by all three methods are shown in Fig. 1. The total number of PCA elements is limited by the training set size (here, 54), while the SH basis is infinite. Thus, denser sampling schemes would have permitted more SH elements but would not have affected PCA.

Fig. 1 shows that PCA interpolation consistently obtained lower errors than SH truncation: the average PCA interpolation errors over the $4 - 12$ kHz frequency range are $5 - 6$ dB (simulations) and $2 - 4$ dB (measurements) lower. These differences diminish at higher frequencies, and with denser sampling schemes. Contrarily, OMP interpolation produces only a negligible difference. In some cases of SH representation, the standard sequential SH basis is actually the optimal, and so SH truncation and OMP produce identical results.

4.3. Discussion

The results demonstrate that PCA may provide a more effective representation than SH, even with ear-alignment. Typically, each consecutive PCA element is of a lesser importance, as opposed to SH representation, where significant energy can be found at high orders. Subsequently, PCA's advantage over SH truncation may diminish as more sample points and elements are used, as shown for the measured data in Fig. 1.

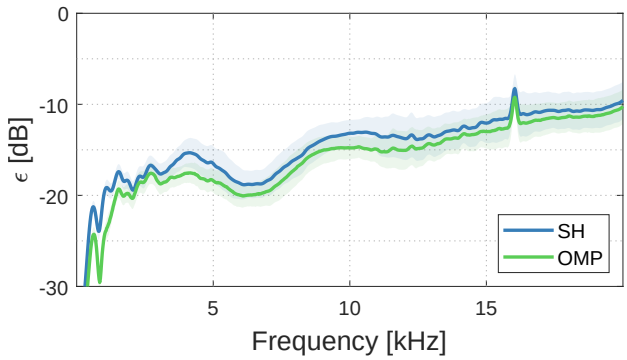
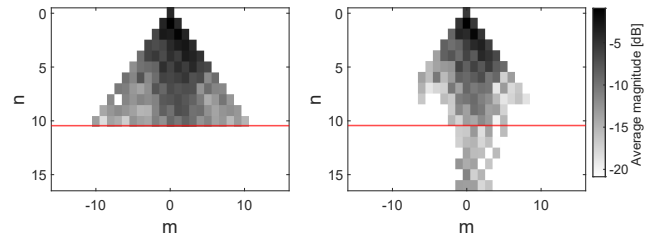


Fig. 2: Interpolation error averaged over the test set, calculated with a 302 points Lebedev grid and limited to 121 reconstruction basis vectors. Standard deviations are shown in shaded areas. OMP was trained up to SH order 16.



(a) SH Truncation. (b) OMP over SH.

Fig. 3: Average spherical harmonics spectrum for the measured test set at 10 kHz, SH truncation (left) and SH selected by OMP (right). The representation was limited to 121 elements. SH are arranged in a 2D grid, with Y_0^0 at center top. The standard representation consists of all SH up to order 10 (red line). In OMP representation, higher SH orders were selected in place of lower order SH.

Rather than using more elements, PCA performance may be further improved using larger training sets. Additional interpolation error reductions were observed in preliminary experiments with training sets expanded from 50 to 80 participants.

Despite obtaining lower interpolation errors, PCA has several disadvantages. The interpolation is limited to the original directions acquired for the training set, as these are the locations in which the PCA elements are evaluated. This is opposed to SH interpolation, where the elements may be evaluated numerically in any direction. Theoretically, this limitation may be overcome by interpolating the elements to additional directions, an approach not attempted here. Furthermore, PCA requires a training phase, as opposed to SH truncation. From a computational perspective, training the PCA's algorithm is negligible. However, several dozens of full length HRTFs must be acquired for the training. Therefore, integrating PCA in a measurement process is worthwhile only if a large number of HRTFs is planned to be acquired. In addition, the compatibility of PCA components to different datasets was not explored. In case PCA elements may not be used across different sets, a separate training process is required for each. Lastly, the standard deviation of the PCA interpolation error was $\sim 0.5 - 1$ dB higher than that of SH truncation. This is to be expected, as per definition PCA is most effective for correlated data, and less effective for outliers. Therefore, PCA may not be suitable for measurement processes where great variability among subjects is expected.

Note that this study only applied PCA across listeners in the frequency domain. PCA application to ear-aligned HRTFs across frequencies and across directions may also be worth exploring.

The interpolation errors obtained by OMP using sparse sampling schemes do not offer a significant improvement over SH truncation. It is only with much larger sampling schemes that an improvement is apparent. Interpolation errors obtained by OMP and SH truncation using a 302 points Lebedev sampling scheme are shown in Fig. 2. The average OMP error is 1.5 dB lower over the $4 - 12$ kHz frequency range. This reduction is achieved due to OMP's ability to select specific elements best representing the data. Fig. 3 shows the selection of 121 SH for the representation of the entire measured test set at 10 kHz. The SH truncation representation (left) uses all SH up to order 10. In contrast, OMP (right) skips some low order elements, mostly in the more extreme degrees. These elements are substituted by SH of higher orders, with degrees closer to $m = 0$.

The results indicate that OMP is most likely ineffective for sparse interpolation purposes. However, OMP may be of interest in applications such as HRTF prediction, which require representation with a reduced number of elements.

5. CONCLUSION

This paper investigated two methods for interpolation of ear-aligned HRTFs from sparse samples using vector sets selected by Orthogonal Matching Pursuit or generated by Principal Component Analysis. The methods were evaluated by the representation and interpolation of simulated and measured HRTF datasets. PCA representation showed a significant reduction of interpolation errors over spherical harmonics truncation with sparse sampling schemes. OMP showed no improvement with sparse sampling schemes, and a mild improvement with denser sampling schemes. Each method may be incorporated into existing HRTF measurement configurations by acquiring a set of full length HRTFs and training the algorithms. While the results presented here rely only on objective comparisons, subjective evaluations are necessary for verifying the results, and are suggested as future work. In addition, PCA's observed advantages merit further investigation, examining the optimal size of the training set, and exploring methods tailored for spatial data such as GLRAM and T-SVD.

6. REFERENCES

[1] Ralph V. Algazi, Richard O. Duda, Dennis M. Thompson, and Carlos Avendano, "The cipic hrtf database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*. IEEE, 2001, pp. 99–102.

[2] Craig T. Jin, Pierre Guillon, Nicolas Epain, Reza Zolfaghari, André van Schaik, Anthony I. Tew, Carl Hetherington, and Jonathan Thorpe, "Creating the sydney york morphological and acoustic recordings of ears database," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 37–46, 2014.

[3] Fabian Brinkmann, Alexander Lindau, Stefan Weinzierl, Gunnar Geissler, and Steven van de Par, "A high resolution head-related transfer function database including different orientations of head above the torso," 2013.

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

[5] Michael J. Evans, James A.S. Angus, and Anthony I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2400–2411, 1998.

[6] Nail A. Gumerov, Adam E. O'Donovan, Ramani Duraiswami, and Dmitry N. Zotkin, "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*, vol. 127, no. 1, pp. 370–386, 2010.

[7] Griffin D. Romigh, Douglas S. Brungart, Richard M. Stern, and Brian D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 921–930, 2015.

[8] David Lou Alon, Zamir Ben-Hur, Boaz Rafaely, and Ravish Mehra, "Sparse head-related transfer function representation with spatial aliasing cancellation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6792–6796.

[9] Fabian Brinkmann and Stefan Weinzierl, "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition," in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

[10] Jianjun He, Woon-Seng Gan, and Ee-Leng Tan, "On the preprocessing and postprocessing of hrtf individualization based on sparse representation of anthropometric features," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 639–643.

[11] Zamir Ben-Hur, David Lou Alon, Ravish Mehra, and Boaz Rafaely, "Efficient representation and sparse sampling of head-related transfer functions using phase-correction based on ear alignment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 2249–2262, 2019.

[12] Doris J. Kistler and Frederic L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.

[13] Mengfan Zhang, Zhongshu Ge, Tiejun Liu, Xihong Wu, and Tianshu Qu, "Modeling of individual hrtfs based on spatial principal component analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 785–797, 2020.

[14] Parham Mokhtari, Hiroaki Kato, Hironori Takemoto, Ryouichi Nishimura, Seigo Enomoto, Seiji Adachi, and Tatsuya Kitamura, "Further observations on a principal components analysis of head-related transfer functions," *Scientific reports*, vol. 9, no. 1, pp. 1–7, 2019.

[15] Shouichi Takane, "Spatial principal component analysis of head-related transfer functions using their complex logarithm with unwrapping of phase," in *19th International Congress on Acoustics*, 2019.

[16] Yagyensh Chandra Pati, Ramin Rezaiifar, and Perinkulam Sambamurthy Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proceedings of 27th Asilomar conference on signals, systems and computers*. IEEE, 1993, pp. 40–44.

[17] Svante Wold, Kim Esbensen, and Paul Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[18] Fabian Brinkmann, Manoj Dinakaran, Robert Pelzer, Peter Grosche, Daniel Voss, and Stefan Weinzierl, "A cross-evaluated database of measured and simulated hrtfs including 3d head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, 2019.