A Wide Dataset of Ear Shapes and Pinna-Related Transfer Functions Generated by Random Ear Drawings

Corentin Guezenoc, Renaud Seguier

A Wide Dataset of Ear Shapes and Pinna-Related Transfer Functions Generated by Random Ear Drawings

Corentin Guezenoc* and Renaud Séguier†
FAST Research Team
IETR (CNRS UMR 6164)
CentraleSupélec
Avenue de la Boulaie, 35510 Cesson-Sévigné, France

Head-related transfer functions (HRTFs) individualization is a key matter in binaural synthesis. However, currently available databases are limited in size compared to the high dimensionality of the data. Hereby, we present the process of generating a synthetic dataset of 1000 ear shapes and matching sets of pinna-related transfer functions (PRTFs), named WiDESPREaD (wide dataset of ear shapes and pinna-related transfer functions obtained by random ear drawings) and made freely available to other researchers. Contributions in this article are three-fold. First, from a proprietary dataset of 119 three-dimensional left-ear scans, we build a matching dataset of PRTFs by performing fast-multipole boundary element method (FM-BEM) calculations. Second, we investigate the underlying geometry of each type of high-dimensional data using principal component analysis (PCA). We find that this linear machine learning technique performs better at modeling and reducing data dimensionality on ear shapes than on matching PRTF sets. Third, based on these findings, we devise a method to generate an arbitrarily large synthetic database of PRTF sets that relies on the random drawing of ear shapes and subsequent FM-BEM computations.

I. INTRODUCTION

In daily life we unconsciously capture the spatial characteristics of the acoustic scene around us thanks to auditory cues such as sound level, time-of-arrival and spectrum. Such cues derive from the alterations of sound on its acoustic path to our eardrums, which depend not only on the room and the position of the acoustic source, but also on the listener's morphology. Their mathematical description in free-field is called head-related transfer functions (HRTFs) in the frequency domain and head-related impulse responses (HRIRs) in the time domain [30]. They are the cornerstone of a technique called binaural synthesis that allows the creation of a virtual auditory environment through headphones: by convolving a given sound sample with the right pair of HRIRs before presenting it to the listener, the sound sample is perceived at the desired location.

The use of a non-individual HRTF set in binaural synthesis is known to cause discrepancies such as wrong perception of elevation, weak externalization and front-back inversions [37]. Thus, a lot of work has been done for the past decades towards user-friendly HRTF individualization, among which four categories can be identified [10]. Acoustical measurement [38] is the state-of-the-art method and relies on a heavy measurement apparatus and is time-intensive. Numerical simulation allows the simulation of HRTFs from a 3D scan of a listener's morphology [18]. The associated individual measurement phase is much less troublesome in terms of equipment (a portable light-based scanner can be used for instance), nevertheless the approach is time-intensive, particularly so during the simulation step. Finally, the two latter families of approaches aim at providing somewhat low-cost but real-time solutions to the matter. They are usually based either on anthropometric measurements [13, 14, 27, 44] or on perceptual feed-back [15, 29, 34, 40] and often rely heavily on HRTF databases.

However, currently available measured HRTF databases [1, 5–7, 26, 36] are small compared to the dimensionality of the data. Indeed, the largest that we know of, the ARI (acoustics research institute) database [26], features 120 subjects, while the dimension of a typical high-resolution HRIR set [5] is about $1.2 \cdot 10^6$ (256 time-domain samples $\times$ 2300 directions $\times$ 2 ears). While work has been done towards combining existing databases [2, 35], such composite databases can hardly attain the same level of homogeneity as a database made in a single campaign. Furthermore, the total number of subjects would amount to a few hundreds at best. Synthetic datasets have also been built by numerically simulating HRTF sets from scans of listener morphology [6, 16, 19, 33]. However, to the best of our knowledge, only the HUTUBS database [6] is fully public. Moreover, such datasets are not larger than acoustically measured ones. Indeed, the largest that we know of, HUTUBS [6], features 96 subjects which is less than the ARI one. This can be explained by the fact that, although less tedious than acoustic measurements, the acquisition of morphological scans for a large number of human subjects is far from trivial.

In this paper, we aim at alleviating the lack of large-scale datasets. First, in Section III, we supplement a dataset of 119 3D human left-ear scans with the corresponding 119 simulated PRTF sets. Then, in Sections IV, V and VI, we investigate the underlying geometry and the potential for dimensionality reduction of both types of

*Electronic address: corentin.guezenoc@centralesupelec.fr
†Electronic address: renaud.seguier@centralesupelec.fr

data by performing principal component analysis (PCA) on each dataset. Although it is a coarse machine learning technique whose limitations include linearity, PCA[1] is a good starting point thanks to its algorithmic simplicity and high interpretability. Finally, based on our findings, we present in Section VII a method to generate an arbitrarily large synthetic database of PRTF sets, which relies on random ear shape drawings and numerical acoustic simulations.

Let us note that, while we focus here on ear shapes and matching PRTFs, the information contained in PRTFs is key to the matter of HRTF individualization. Indeed, pinnae have a vast influence on the spectral features involved in perceptual discrepancies due to a lack of individualization [3]. Furthermore, pinnae constitute the most complex component of HRTF-impacting morphology, in terms of shape, inter-individual variability and in terms of how small physical changes can have a strong influence on the resulting filters.

II. ORIGINAL EAR SHAPE DATASET

Work presented in this article is based on a proprietary dataset of left ear 3-D scans of $n = 119$ human subjects. The dataset was constituted in previous work by Ghorbal, Séguier and Bonjour [9]. The pinna meshes were acquired using a commercial structured-light-based scanner. They were then normalized in size and rigidly aligned. Finally, they were registered: the point clouds were re-sampled so that they were in semantic correspondence with each other, thus sharing an identical number of vertices $n_v = 17176$. 35750 triangular faces were defined based on the indices of the n_v vertices: the definition of the faces is identical from one mesh to the other.

In the following, we denote by $E = \{\mathbf{e}_1, \dots \mathbf{e}_n\}$ the set of n ear point clouds whose x, y and z coordinates were concatenated into row vectors $\mathbf{e}_1, \dots \mathbf{e}_n \in \mathbb{R}^{3n_v}$, with $3n_v = 54528$. As mentioned above, the only change from one mesh to the other resides in the coordinates of the n_v vertices. Therefore, the term ‘ear shape’ is hereon meant as ear point cloud.

III. NUMERICAL SIMULATIONS OF PRTFS

For all ear shape $\mathbf{e}_i$ in E, we simulated numerically the corresponding PRTF set $\mathbf{p}_i \in \mathbb{C}^{n_f \times n_d}$, where n_f and n_d denote respectively the number of frequency bins and the number of directions of measurements. Simulations were carried out using the fast-multipole boundary element method (FM-BEM) [11], thanks to the mesh2hrtf software developed by the ARI team [41, 43].

We denote $\varphi : \mathbb{R}^{3n_v} \longrightarrow \mathbb{C}^{n_f \times n_d}$ the process of going from a registered n_v-vertex ear point cloud to the corresponding simulated PRTF set, which is described in the rest of the subsection.

Simulations were made for $n_f = 160$ frequencies from 0.1 to 16 kHz, regularly spaced with a step of 100 Hz. The frequency resolution was chosen so that it was finer than the equivalent rectangular bandwidth (ERB)-based frequency scale in most of the frequency range. Indeed, the ERB scale is appropriate for HRTFs according to [22] and the 100-Hz-spaced linear scale is finer than the ERB scale for frequencies above 700 Hz, which is more than sufficient in the case of PRTFs, who include little spectral variations below 4-5 kHz.

A. Mesh closing and grading

First, we derived the ear mesh from the ear point cloud by incorporating the 35750 triangular faces defined by the indices of the n_v vertice, as explained in Section II.

Second, we closed the ear mesh by filling the canal hole based on our prior knowledge of the boundary’s vertex indices, and then by stitching the resulting mesh onto a cylindrical base mesh. Using such a small base mesh instead of one of a head and torso has consequences: spectral features that are usually found in HRTFs are altered (head shadowing effect is reduced to a smaller angular zone and shifted to higher frequencies) or absent (ripples due to the torso). However, as we did not have at our disposal a dataset of individual 3-D head and torso scans, in the latter case we would only have been able to use a generic head and torso mesh, which would have mixed non-individual spectral features with the individual pinna-related ones, at the cost of a great increase in required computing resources. These steps were scripted in Blender[2] Python and performed automatically using various Blender built-in mesh treatments.

Third, a re-sampling (also called grading) of the mesh was performed. This step is a pre-requirement to any boundary element simulation: the mesh ought to be as regular as possible and sampled finely enough with regard to the maximum simulated frequency. According to Gumerov, O’Donovan, Duraiswami and Zotkin [12], for instance, the mesh should present a uniform vertex distribution, equilateral triangles and at least five elements per wavelength. In our case, we used the progressive grading approach proposed by Ziegelwanger, Kreuzer and Majdak in [42] and made available on-line as an OpenFlipper[3] plug-in, which makes the mesh finer near the ear canal (where the sound source is positioned) and progressively coarser elsewhere. This considerably decreases the computing cost of the FM-BEM simulation compared

[1] PCA is closely related to the Karhunen-Loève transform, widely used in the field of information theory.

[2] https://www.blender.org/

[3] http://www.openflipper.org/

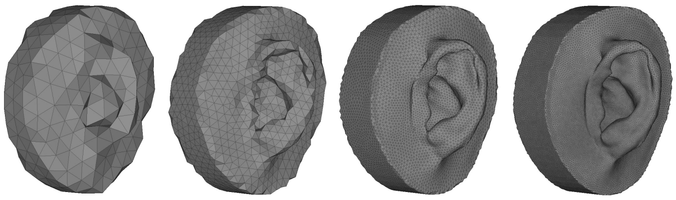


FIG. 1: Simulation-ready meshes derived from ear point cloud $\mathbf{e}_1$ for four mesh grading configurations, each corresponding to a frequency band. Left to right: [0.1, 0.4 kHz], [0.5, 2.0 kHz], [2.1, 3.5 kHz] and [3.6, 16 kHz].

to uniform re-sampling, while maintaining numerical accuracy.

Additionally, in order to further reduce the computational cost, we adapted the mesh grading step to each of four different frequency bands. At low frequencies, a uniform re-sampling was enough due to the low number of required elements. It was performed with target edge lengths of 10 and 5 mm, in the frequency bands [0.1, 0.4 kHz] and [0.5, 2.0 kHz], respectively. At higher frequencies, the re-sampling was progressive, with target minimum and maximum edge lengths of 2 and 5 mm, and 0.7 and 5 mm, in the frequency bands [2.1, 3.5 kHz] and [3.6, 16 kHz], respectively. An example of simulation-ready meshes (each corresponding to a mesh grading configuration) is displayed in Figure 1.

B. Simulation settings

According to the reciprocity principle [45], a sound source was placed at the entry of the filled ear canal, while virtual microphones were disposed on a spherical grid centered on the pinna: a 2-meter-radius icosahedral geodesic polyhedron of frequency 256 ($n_d = 2562$ directions), displayed in Figure 2. Although not relevant to the rest of this article, PRTFs were computed for other virtual microphone grids as well, which are included in WiDESPREaD.

The sound source was created by assigning a vibrant boundary condition to a small patch of triangular faces located on the ear canal plug. Elsewhere on the mesh, the boundary condition was set to infinitely reflective. This boundary condition setting is used in the vast majority of work involving HRTF calculations [16, 23], which makes our work more easily comparable to the literature. Furthermore, although modeling the acoustic impedance of the skin as infinite may be somewhat unrealistic, to the best of our knowledge no alternative has been proposed so far, possibly because of the limited frequency range of impedance tube measurements: up to 6.4 kHz for a standard device and up to 12.8 kHz for the experimental one proposed by Kimura, Kunio, Schumacher and Ryu [20].

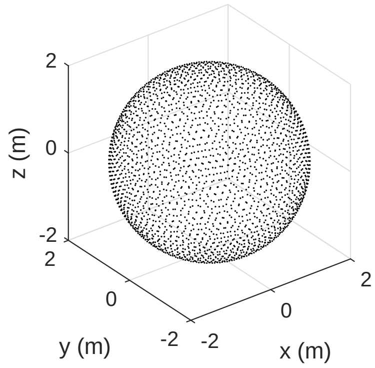


FIG. 2: Spherical grid used for PRTF simulations: 2-meter-radius icosahedral geodesic polyhedron of frequency 256 ($n_d = 2562$ vertices).

C. Post-processing

Once the simulation of a PRTF set was complete, it was post-processed as follows. Let $\mathbf{p}_{\text{raw}} \in \mathbb{C}^{n_f \times n_d}$ be a PRTF set simulated for $n_f = 160$ frequency bins that exclude the constant component and for n_d vertices of a spherical grid.

First, PRTFs were padded in frequency zero with the 100-Hz complex values.

Then, diffuse field equalization [27] was performed by removing the non-directional component, called Common Transfer Function (CTF), from the PRTFs. For all frequency bin $f = 1, \ldots n_f$ and for all direction of index $d = 1, \ldots n_d$,

$$\mathbf{p}(f, d) = \frac{\mathbf{p}_{\text{raw}}(f, d)}{\mathbf{c}(f)}, \tag{1}$$

where the CTF $\mathbf{c} \in \mathbb{C}^{n_f}$ was obtained by calculating a Voronoi-diagram-based [4] weighted average of the log-magnitude spectra of $\mathbf{p}$ over all directions $d = 1, \ldots n_d$, then by deriving the corresponding minimal phase spectrum.

IV. PCA OF EAR SHAPES

From the set of ear shapes E described in Section II, we classically constructed a statistical shape model of the ear using PCA [8]. Let there be $\mathbf{X}_E = (\mathbf{e}_1 \ \ldots \ \mathbf{e}_n)^{\text{t}} \in \mathbb{R}^{n \times 3n_v}$ the data matrix, $\bar{\mathbf{e}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}_i$ the average ear shape and $\bar{\mathbf{X}}_E = (\bar{\mathbf{e}} \ \ldots \ \bar{\mathbf{e}})^{\text{t}} \in \mathbb{R}^{n \times 3n_v}$ the matrix constituted of the average shape stacked n times. Finally, let $\mathbf{\Gamma}_E \in \mathbb{R}^{3n_v \times 3n_v}$ be the covariance matrix of $\mathbf{X}_E$:

$$\mathbf{\Gamma}_E = \frac{1}{n-1} \left(\mathbf{X}_E - \bar{\mathbf{X}}_E\right)^{\text{t}} \left(\mathbf{X}_E - \bar{\mathbf{X}}_E\right). \tag{2}$$

PCA can thus be written as

$$\mathbf{Y}_E = \left(\mathbf{X}_E - \bar{\mathbf{X}}_E\right) \mathbf{U}_E{}^{\text{t}}, \tag{3}$$

where $\mathbf{U}_E$ is obtained by diagonalizing the covariance matrix $\mathbf{\Gamma}_E$

$$\mathbf{\Gamma}_E = \mathbf{U}_E^{\mathrm{t}} \mathbf{\Sigma}_E{}^2 \mathbf{U}_E. \quad (4)$$

In the equations above, $\mathbf{\Sigma}_E{}^2 \in \mathbb{R}^{(n-1)\times(n-1)}$ is the diagonal matrix that contains the eigenvalues of $\mathbf{\Gamma}_E$, $\sigma_{E_1}{}^2$, $\sigma_{E_2}{}^2$, $\ldots \sigma_{E_{n-1}}{}^2$, ordered so that $\sigma_{E_1}{}^2 \geq \sigma_{E_2}{}^2 \geq \cdots \geq \sigma_{E_{n-1}}{}^2$

$$\mathbf{\Sigma}_E{}^2 = \begin{bmatrix} \sigma_{E_1}{}^2 & & \\ & \ddots & \\ & & \sigma_{E_{n-1}}{}^2 \end{bmatrix}, \quad (5)$$

and $\mathbf{U}_E \in \mathbb{R}^{(n-1)\times 3n_v}$ is an orthogonal matrix that contains the corresponding eigenvectors $\mathbf{u}_{E_1}$, $\mathbf{u}_{E_2}$, $\ldots$ $\mathbf{u}_{E_{n-1}} \in \mathbb{R}^{3n_v}$

$$\mathbf{U}_E = \begin{bmatrix} \mathbf{u}_{E_1} \\ \vdots \\ \mathbf{u}_{E_{n-1}} \end{bmatrix}. \quad (6)$$

The eigenvalues denote how much variance in the input data is explained by the corresponding eigenvectors.

In the equations above, we implicitly set the number of principal components (PCs) to $n-1$, because all PCs after the $(n-1)^{th}$ are trivial, i.e. of null associated eigenvalue. Indeed, the number of examples n is lower than the data dimension $3n_v$ and the data is centered, thus

$$r = \operatorname{rank}\left(\mathbf{X}_E - \bar{\mathbf{X}}_E\right) \leq n-1. \quad (7)$$

Hence, the rank of the covariance matrix does not exceed $n-1$ either:

$$\operatorname{rank}\left(\mathbf{\Gamma}_E\right) \leq \min\left(r, r\right) = r \leq n-1. \quad (8)$$

The behavior of the first 3 principal components is illustrated as follows. For each PC of index $j \in \{1,\, 2,\, 3\}$, we set the j^{th} PC weight to $\lambda\sigma_{E_j}$ and all other PC weights to zero, with $\lambda \in \{-5,\, -3,\, -1,\, +1,\, +3,\, +5\}$ and reconstructed the corresponding ear shape $\mathbf{e}_{v_j}(\lambda)$ by inverting Equation (3)

$$\mathbf{e}_{v_j}(\lambda) = \begin{pmatrix} 0 & \ldots & 0 & \lambda\sigma_{E_j} & 0 & \ldots & 0 \end{pmatrix} \mathbf{U}_E + \bar{\mathbf{e}}. \quad (9)$$

Meshes derived from said ear shapes are displayed in Figure 3, colored with the vertex-to-vertex euclidean distance to the average shape.

The first one seems to control vertical pinna elongation including concha height and lobe length up to disappearance, as well as some pinna vertical axis rotation. The second one seems to encode the intensity of some topography features such as triangular fossa depth or helix prominence. It also has an impact on concha shape and vertical axis rotation. The third PC seems to have a strong influence on concha depth, triangular fossa depth as well as upper helix shape.

V. PCA OF LOG-MAGNITUDE PRTFS

In the following, we focus on the log-magnitude spectrum of the PRTFs. One reason is that HRTFs can be well modeled by a combination of minimum phase spectrum and pure delay [24]. Another one is the fact that, the TOA due to the pinnae, i.e. the one contained in PRTFs, is negligible compared to the effect of head and torso shadowing in HRTFs. The logarithmic scale was chosen for its coherence with human perception.

For all ear shape $\mathbf{e}_i \in E$, let $\mathbf{p}_i = \varphi\left(\mathbf{e}_i\right) \in \mathbb{C}^{n_f \times n_d}$ be the corresponding PRTF set, computed according to the process described in Section III, and let $\mathbf{q}_i = 20 \cdot \log_{10}\left(|\mathbf{p}_i|\right) \in \mathbb{R}^{n_f \times n_d}$ be the corresponding log-magnitude PRTF set, where the $|\cdot|$ and $\log_{10}$ operators are considered element-wise. Accordingly, let $\phi : \mathbb{R}^{3n_v} \longrightarrow \mathbb{R}^{n_f \times n_d}$, defined by $\mathbf{e} \longmapsto \mathbf{q} = 20 \cdot \log_{10}\left(|\varphi(\mathbf{e})|\right)$, be the process of deriving a log-magnitude PRTF set from an ear point cloud, and let $Q = \{\mathbf{q}_1,\, \ldots\, \mathbf{q}_n\} = \{\phi(\mathbf{e}_1)\, \ldots\, \phi(\mathbf{e}_n)\}$ be the set of log-magnitude PRTF sets derived from E.

Most work in the literature either stacks the HRTFs of various directions and subjects into a data matrix of size $(n \cdot n_d) \times n_f$ prior to PCA [21, 28], or performs PCA one direction at a time on n_d different $n \times n_f$-sized matrices [31, 39]. In contrast, we chose to concatenate PRTFs from the n_d directions into a row vector $\mathbf{q}_i \in \mathbb{R}^{n_f n_d}$ for each subject $i = 1,\, \ldots\, n$. The n row vectors were then stacked into the data matrix $\mathbf{X}_Q = \left(\mathbf{q}_1,\, \ldots\, \mathbf{q}_n\right)^{\mathrm{t}} \in \mathbb{R}^{n \times (n_f n_d)}$. This approach has the advantage of parsing only the across-subject variability, instead of mixing the contributions of directionality and inter-individuality into the statistical analysis.

As in the case of ear shapes, we performed PCA on the data matrix $\mathbf{X}_Q$ according to Equations (3), (4) and (6). The number of non-trivial PCs is $(n-1)$ in this case as well, due to the fact that $n < n_f n_d$.

Various PRTF sets that illustrate the behavior of the three first PCs were reconstructed as explained in Section IV and Equation (9). They can be observed in Figure 4 for directions that belong to the median sagittal plane. As it was expected, no variations are visible below 5 kHz: at these wavelengths the pinna have little impact on sound propagation. Each PC appears to represent a different pattern of change in anterior and posterior directions, although only the first one seems to have a strong influence on directions above the head. However, it does not seem possible to distinguish patterns that are limited to a certain range of directions and/or frequencies. Furthermore, we are not able to identify a PC or a combination of PCs that represents a frequency shift in the PRTFs. Although the pinnae used to construct the model are normalized in size, one could have expected to observe frequency shifts due to variations in concha volume, for example.

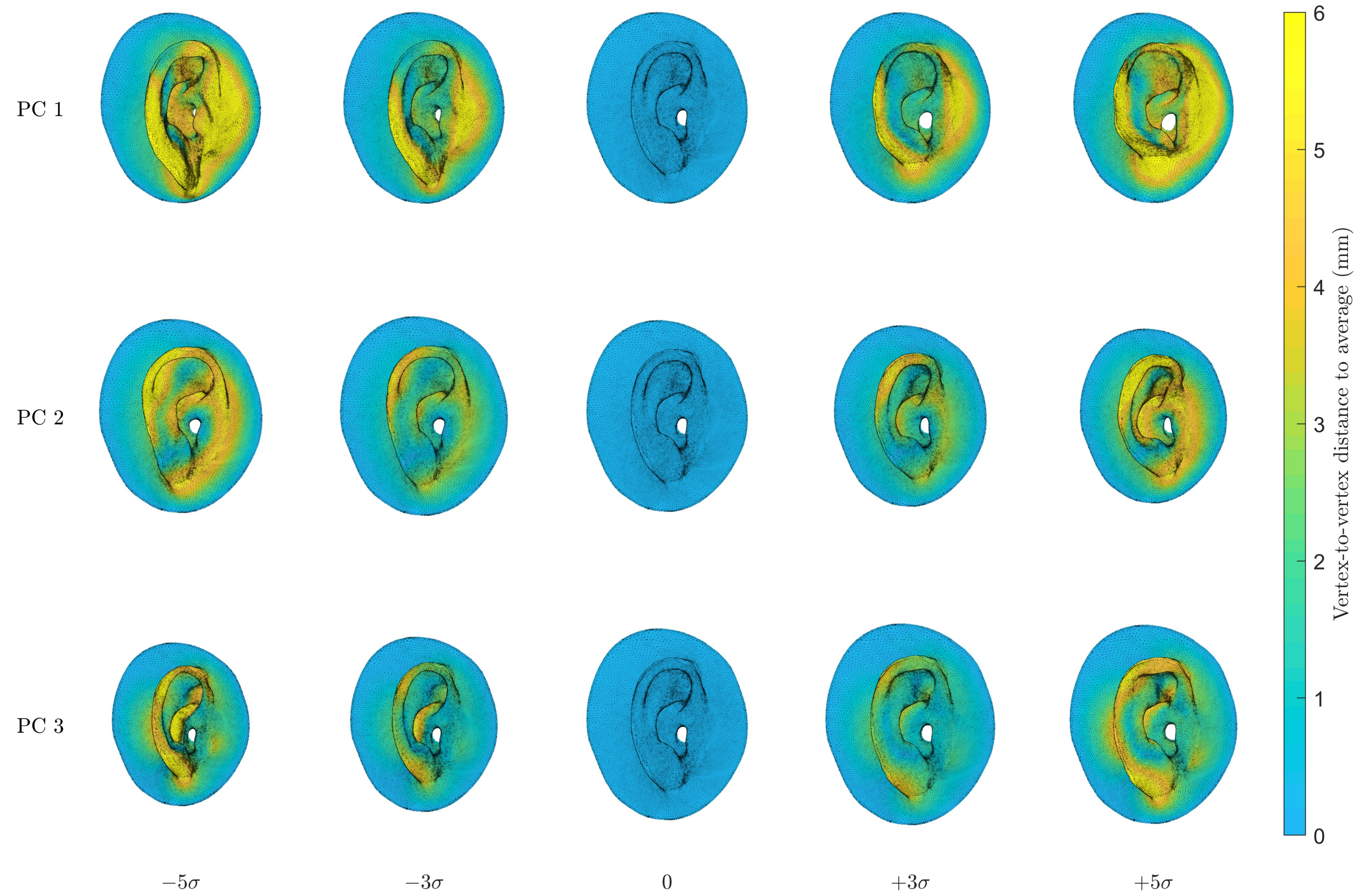


FIG. 3: (color online) First three principal components (PCs) of the PCA ear shape model. Rows: PC of index $j \in \{1, 2, 3\}$. Columns: Weight assigned to given PC, indicated in proportion of its standard deviation σ_{E_j}.

VI. COMPARISON OF BOTH PCA MODELS

A. Dimensionality reduction capacity

Let $S \in \{E, Q\}$ be either dataset. PCA can be used as a dimensionality reduction technique by retaining only the first p PCs and setting the weights of the discarded PCs to zero [17], where $p \in \{1, \ldots n-1\}$:

$$\tilde{\mathbf{Y}}_S = \begin{bmatrix} y_{S_{1,1}} & \ldots & y_{S_{1,p}} & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{S_{n,1}} & \ldots & y_{S_{n,p}} & 0 & \ldots & 0 \end{bmatrix}, \tag{10}$$

where $y_{S_{i,j}}$ is the value of matrix $\mathbf{Y}_S$ at the i^{th} row and j^{th} column for all $i = 1, \ldots n$ and $j = 1, \ldots n-1$.

Indeed, the change of basis defined by ${\mathbf{U}_S}^{\text{t}}$ allows us to transform the dataset $\mathbf{X}_S$ into a domain where the associated covariance matrix ${\mathbf{\Sigma}_S}^2$ is diagonal with its diagonal values in decreasing order. In other words, PCs are independent up to the second-order statistical moment and are ordered so that the first PCs describe more variability in the data than the last ones.

Approximated data can then be reconstructed by inverting Equation (3):

$$\tilde{\mathbf{X}}_S = \tilde{\mathbf{Y}}_S \mathbf{U}_{\text{S}} + \bar{\mathbf{X}}_S. \tag{11}$$

A simple but useful metric to evaluate the capacity of a PCA model to reduce dimensionality is the cumulative percentages of total variance (CPV) [17, section 6.1]

$$\tau_{S_p} = 100 \cdot \left(\sum_{j=1}^{p} {\sigma_{S_j}}^2 \right) \Big/ \left(\sum_{j=1}^{n-1} {\sigma_{S_j}}^2 \right), \tag{12}$$

where $S \in \{E, Q\}$ represents either the set of ear shapes E or the set of log-magnitude PRTFs Q and $p \in \{1, \ldots n-1\}$ is the number of retained PCs. CPVs for both models are plotted in Figure 5.

A first notable result is that, for the ear shape model, the 99%-of-total-variance threshold is reached for $p = 80$ retained PCs, i.e. only $\frac{p}{n-1} = \frac{80}{118} = 67.8\%$ of the maximum number of PCs. In other words, the 118-dimensional linear subspace of $\mathbb{R}^{3n_v} = \mathbb{R}^{56661}$ defined

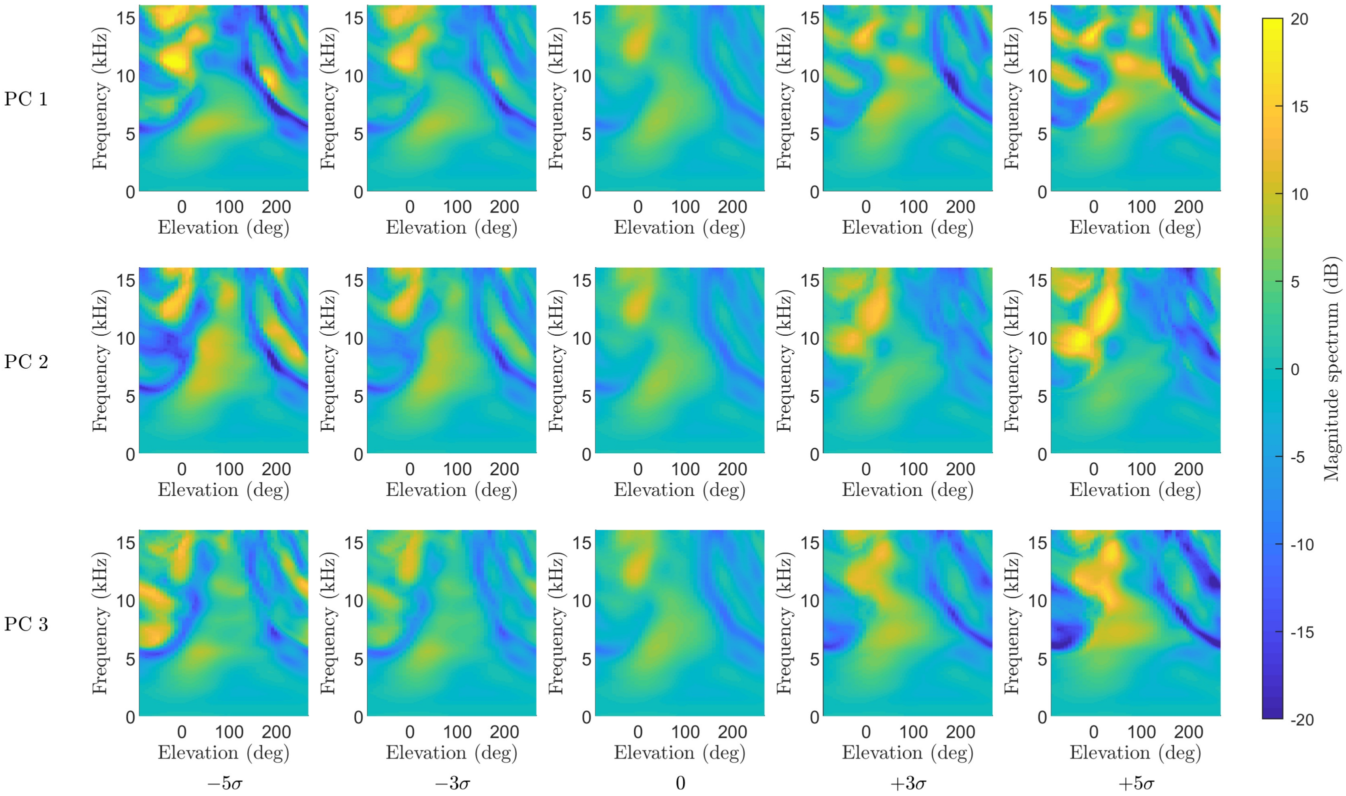


FIG. 4: (color online) First Principal Components (PCs) of the PCA model of log-magnitude PRTFs. Reconstructed PRTF sets are plotted in the median sagittal plane. Rows: PC. Columns: Weight assigned to given PC, indicated in proportion of its standard deviation σ.

by the $n = 119$ pinnae of our database can be described using only 80 parameters with reasonable reconstruction accuracy, in the sense of a vertex-to-vertex mean-square error.

More importantly, PCA appears to be significantly more successful at reducing the dimension of ear shapes $\mathbf{e}_i$ than that of PRTF sets generated from the same ear shapes $\mathbf{q}_i = \phi(\mathbf{e}_i)$. Indeed, the PRTF CPV is significantly lower than the ear shape CPV for any number of retained PCs. For instance, the 99%-of-total-variance threshold is reached for 112 PCs out of 118 for the PRTF model against 80 out of 118 for the ear shape one.

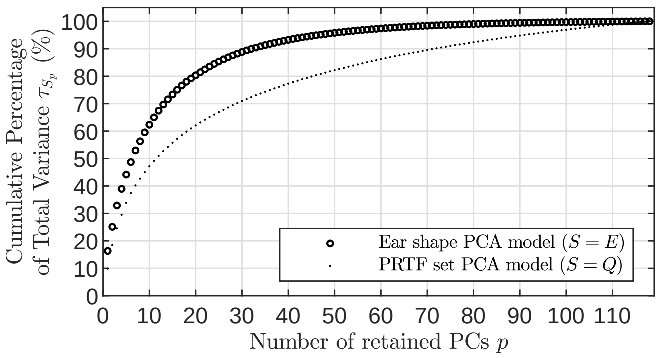


FIG. 5: CPV τ_{S_p} as a function of the number of retained PCs $p \in \{1, \dots\ n-1\}$ for either PCA model. Circles: ear shape model ($S = E$). Dots: PRTF set model ($S = Q$).

B. Statistical distribution

Furthermore, in order to get a better idea of the repartition of the data in both 118-dimensional linear subspaces, we tested the PCs of each model for multivariate normal distribution using Royston's test [32]. The test was performed on the columns of the PC weights matrix $\mathbf{Y}_S$, where $S \in \{E, Q\}$ denotes the dataset.

The outcome of the test is an associated p-value of 3.7% in the case of ear PCs, and 0.0% in the case of PRTF PCs, where the p-value refers to the null hypothesis that the distribution is not multivariate normal. In other words, the ear model's PC weights can be considered to be multivariate-normally distributed with a significance level of 3.7%, while its PRTF counterpart's fail the test for any significance level.

C. Summary

Overall, it appears that PCA performs better at modeling and reducing the dimensionality of ear shapes than of the corresponding log-magnitude PRTF sets.

Hence, linear techniques such as PCA seem ill-suited to reduce the dimensionality of PRTF sets. As, in addition, the ear shape model's PC weights follow a multivariate normal distribution, it appears to be more suitable than its PRTF counterpart for generalization and the generation of new data.

VII. DATABASE GENERATION

Non-linear machine learning methods may thus be more suited to model PRTFs sets than linear ones. However, such more complex techniques usually require larger amounts of data. Nevertheless, as mentioned in the introduction, currently available databases of HRTFs feature about 10^2 subjects in the best case, for a data dimension of about 10^6, that is a proportion of 10^{-4} of the data's dimension. Hence, we propose a scalable method to construct a large database of synthetic PRTF sets by using the ear shapes space as a back door where to generate relevant artificial data.

A. Random drawing of ear shapes

The statistical ear shape model learned from dataset E presented in Section IV can be used as a generative model. Indeed, based on the results from Section VI, we assume hereafter that the model's PCs (i.e. the columns of $\mathbf{Y}_E$) are mutually statistically independent and follow normal probability laws of zero mean and σ_{E_j} standard deviation $\mathcal{N}(0, \sigma_{E_j})$, where $j \in \{1, \dots n-1\}$ represents the PC index.

An arbitrarily large number N of ear shapes $\mathbf{e}'_1, \dots \mathbf{e}'_N \in \mathbb{R}^{3n_v}$ could thus be generated as follows. First, for all $i = 1, \dots N$, a PC weights vector $\mathbf{y}_{E_i} = (y_{E_{i,1}}, \dots y_{E_{i,n-1}}) \in \mathbb{R}^{n-1}$ was obtained by drawing the $(n-1)$ PC weights $y_{E_{i,1}}, \dots y_{E_{i,n-1}}$ independently according to their respective probability laws $\mathcal{N}(0, \sigma_{E_1}), \dots \mathcal{N}(0, \sigma_{E_{n-1}})$. Second, the corresponding ear shapes were reconstructed by inverting Equation (3)

$$\mathbf{X}'_E = \mathbf{U}_E \mathbf{Y}'_E + \bar{\mathbf{X}}_E, \tag{13}$$

where $\mathbf{Y}'_E \in \mathbb{R}^{N \times (n-1)}$ is the matrix whose rows are the N PC weights vectors

$$\mathbf{Y}'_E = \begin{bmatrix} \mathbf{y}'_{E_1} \\ \vdots \\ \mathbf{y}'_{E_N} \end{bmatrix} = \begin{bmatrix} y'_{E_{1,1}} & \dots & y'_{E_{1,n-1}} \\ \vdots & \ddots & \vdots \\ y'_{E_{N,1}} & \dots & y'_{E_{N,n-1}} \end{bmatrix}, \tag{14}$$

and $\mathbf{X}'_E \in \mathbb{R}^{N \times 3n_v}$ is the data matrix whose rows are the N ear shapes $\mathbf{e}'_1, \dots \mathbf{e}'_N \in \mathbb{R}^{3n_v}$

$$\mathbf{X}'_E = \begin{bmatrix} \mathbf{e}'_1 \\ \vdots \\ \mathbf{e}'_N \end{bmatrix}. \tag{15}$$

B. Ear shapes quality check

At the end of the ear shape generation process, meshes were derived from the point clouds as in the case of the original dataset (see Section II). We then verified that the meshes were not aberrant and that they were fit for numerical simulation: any mesh that presented at least one self-intersecting face was left out.

In total, 24% (320 out of 1325) of the meshes were discarded. Performing the Royston's multivariate normality test on the 1325 randomly drawn ear PC weights then on the 1005 remaining ones, we observed a decrease in the significance level of the test from 4.8% to 0.8%: the it appears that the statistical distribution of the ear PC weights was degraded by the selection process. However, when looking into the distribution of each PC of the selected ear shapes separately (using the Shapiro-Wilk univariate normality test with a significance level of 5%), we observe that the 9 rejected PCs account only for 3.7% of the total variance.

For simplicity, we consider further on that N is the number of retained meshes i.e. $N = 1005$.

C. Numerical simulation

Finally, PRTF sets were numerically simulated from the ear shapes of the new set E' according to the process described in Section III

$$\mathbf{p}'_j = \varphi\left(\mathbf{e}'_j\right), \quad \forall j = 1, \dots N. \tag{16}$$

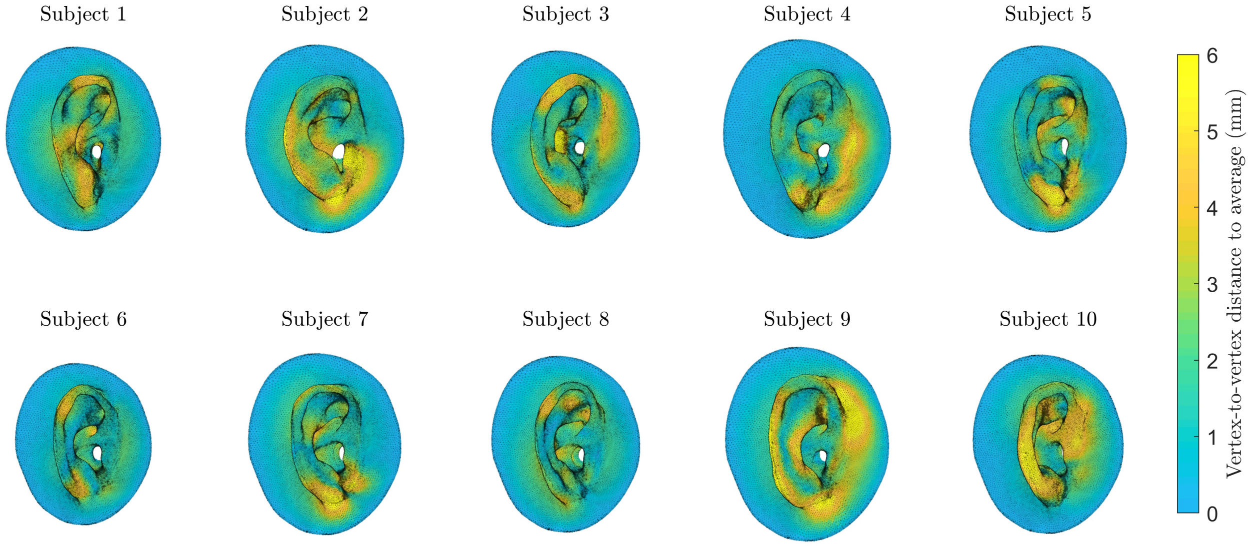
Computing time for the simulation of the 1005 PRTF sets was of 40 days on a workstation that features 12 CPU and 32 GB of RAM.

D. Data visualization

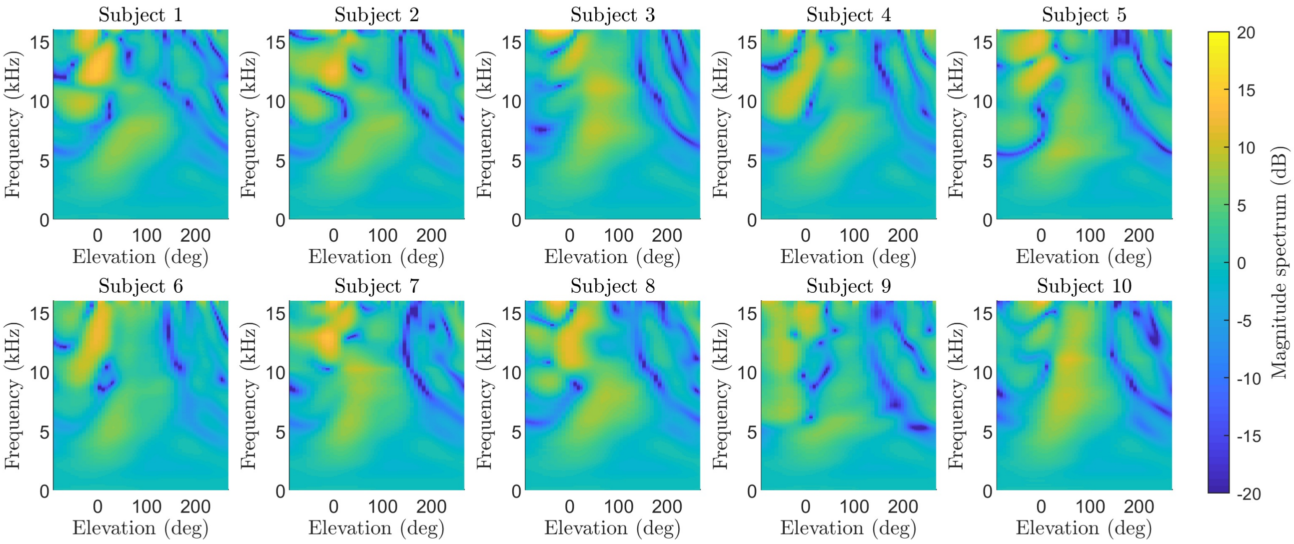
By checking visually, we find that the synthesized ear shapes and PRTF sets look as realistic as hoped. For purposes of illustration, the first 10 ear shapes and matching PRTF sets of the WiDESPREaD dataset are displayed in Figure 6. These first 10 randomly drawn subjects illustrate well how ear shapes and matching PRTFs can be diverse and highlight the interest of this dataset.

VIII. CONCLUSIONS

In this paper, based on a proprietary dataset of 119 left ear meshes, we constituted a corresponding dataset of 119



(a)



(b)

FIG. 6: (color online) Visualization of the first 10 subjects of WiDESPREaD. (a) Meshes derived from the synthetic ear shapes $\mathbf{e}'_1$, ... $\mathbf{e}'_{10}$. Color represents the vertex-to-vertex euclidean distance to the generative model's average $\bar{\mathbf{e}}$. (b) Log-magnitude PRTF sets $20 \cdot \log_{10}(\mathbf{p}'_1)$, ... $20 \cdot \log_{10}(\mathbf{p}'_{10})$ displayed in the median sagittal plane.

PRTF sets by FM-BEM calculations. We then applied a simple linear machine learning technique, PCA, independently to each dataset and found that it performed better at modeling and reducing the dimensionality of data on ear shapes than on PRTF sets. Based on this result, we proposed a method to generate an arbitrarily large synthetic PRTF database by means of random drawing of ear shapes and FM-BEM calculation. The resulting dataset of 1005 ear meshes and corresponding PRTF sets, named WiDESPREaD, is freely available on the Sofacoustics website: `https://www.sofacoustics.org/data/database/widespread`.

Increasing the number of PRTF sets by generating new artificial subjects in the ear shape space, where linear modeling seems adequate, may allow us to better understand the complexity of PRTF and HRTF generation from listener morphology and help to model them better. In particular, non-linear machine learning techniques such as neural networks can benefit from the scalability of this synthetic dataset generation process, as they usually require a large amount of data. As it is, WiDESPREaD is the first database, to our knowledge, with over a thousand PRTF sets and matching registered ear meshes. Although PRTFs are not complete HRTFs, they include an important part of the information relevant to HRTF individualization and, as the dataset includes about 8 times more subjects than any available HRTF dataset, it has great potential to help develop and improve methods for HRTF modeling, dimensionality reduction and manifold learning, as well as spatial interpolation of sparsely measured HRTFs.

Future work includes the analysis of the augmented PRTF dataset and the search for a non-linear manifold. If needed, new data can be generated to increase the size of the dataset, providing computing power and time. Furthermore, anthropometric measurements of the pinnae such as introduced with the CIPIC dataset [1] can be directly derived from the registered meshes, which may prove useful for the active field of HRTF individualization based on anthropometric features. Finally, the method for data generation itself could be further improved on several aspects. Indeed, our rudimentary generative ear shape model could be ameliorated by using either simple upgrades like probabilistic PCA [25, p. 5] or other modeling techniques altogether, although our results suggest that linear modeling techniques may be sufficient. Going one step further, including statistical models of the human head, shoulders and right pinna could extend the method to the synthesis of HRTFs.

IX. ACKNOWLEDGMENTS

Special thanks go to our colleague Simon Leglaive for his thorough and insightful proofreading work.

[1] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, New Platz, NY, USA, pp. 99–102.

[2] Andreopoulou, A., and Roginska, A. (**2011**). "Towards the creation of a standardized HRTF repository," in *Proceedings of the 131th Audio Engineering Society Convention*, Audio Engineering Society, New York, NY, USA.

[3] Asano, F., Suzuki, Y., and Sone, T. (**1990**). "Role of spectral cues in median plane localization," The Journal of the Acoustical Society of America **88**(1), 159–168.

[4] Augenbaum, J. M., and Peskin, C. S. (**1985**). "On the construction of the Voronoi mesh on a sphere," Journal of Computational Physics **59**(2), 177–192.

[5] Bomhardt, R., de la Fuente Klein, M., and Fels, J. (**2016**). "A high-resolution head-related transfer function and three-dimensional ear model database," in *Proceedings of the 172nd Meeting of the Acoustical Society of America*, Honolulu, HI, USA, Vol. 29, p. 050002.

[6] Brinkmann, F., Dinakaran, M., Pelzer, R., Grosche, P., Voss, D., and Weinzierl, S. (**2019**). "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 **67**(9), 705–718.

[7] Carpentier, T., Bahu, H., Noisternig, M., and Warusfel, O. (**2014**). "Measurement of a head-related transfer function database with high spatial resolution," in *Proceedings of the 7th Forum Acusticum*, European Acoustics Association, Kraków, Poland.

[8] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (**1995**). "Active Shape Models - Their Training and Application," Computer Vision and Image Understanding **61**(1), 38–59.

[9] Ghorbal, S., Seguier, R., and Bonjour, X. (**2019**). "Method for establishing a deformable 3d model of an element, and associated system," US Patent 16/300, 044.

[10] Guezenoc, C., and Seguier, R. (**2018**). "HRTF Individualization: A Survey," in *Proceedings of the 145th Audio Engineering Society Convention*, Audio Engineering Society, New York, NY, USA.

[11] Gumerov, N. A., and Duraiswami, R. (**2005**). *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions* (Elsevier Science).

[12] Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., and Zotkin, D. N. (**2010**). "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), 370–386.

[13] Hu, H., Zhou, L., Ma, H., and Wu, Z. (**2008**). "HRTF personalization based on artificial neural network in individual virtual auditory space," Applied Acoustics **69**(2), 163–172.

[14] Hu, H., Zhou, L., Zhang, J., Ma, H., and Wu, Z. (**2006**). "Head Related Transfer Function Personalization Based on Multiple Regression Analysis," in *Proceedings of the 2006 International Conference on Computational Intelligence and Security*, Guangzhou, China, Vol. 2, pp. 1829–1832.

[15] Hwang, S., Park, Y., and Park, Y.-s. (**2008**). "Modeling and Customization of Head-Related Impulse Responses Based on General Basis Functions in Time Domain," Acta Acustica united with Acustica **94**(6), 965–980.

[16] Jin, C. T., Guillon, P., Epain, N., Zolfaghari, R., Schaik, A. v., Tew, A. I., Hetherington, C., and Thorpe, J. (**2014**). "Creating the Sydney York Morphological and Acoustic Recordings of Ears Database," IEEE Transactions on Multimedia **16**(1), 37–46.

[17] Jolliffe, I. T. (**2002**). "Principal components in regression analysis," Principal component analysis 167–198.

[18] Kahana, Y., Nelson, P. A., Petyt, M., and Choi, S. (**1998**). "Boundary Element Simulation of HRTFs and Sound Fields Produced by Virtual Acoustic Imaging Systems," in *Proceedings of the 105th Audio Engineering Society Convention*, Audio Engineering Society, San Francisco, CA, USA.

[19] Kaneko, S., Suenaga, T., Fujiwara, M., Kumehara, K., Shirakihara, F., and Sekine, S. (**2016**). “Ear Shape Modeling for 3d Audio and Acoustic Virtual Reality: The Shape-Based Average HRTF,” in *Proceedings of the Audio Engineering Society 61st International Conference: Audio for Games*, Audio Engineering Society, London, UK.

[20] Kimura, M., Kunio, J., Schuhmacher, A., and Ryu, Y. (**2014**). “A new high-frequency impedance tube for measuring sound absorption coefficient and sound transmission loss,” in *Proceedings of Inter-Noise*, Institute of Noise Control Engineering, Melbourne, Australia.

[21] Kistler, D. J., and Wightman, F. L. (**1992**). “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,” The Journal of the Acoustical Society of America **91**(3), 1637–1647.

[22] Breebaart, J., and Kohlrausch, A. (**2001**). “Perceptual (ir)relevance of HRTF magnitude and phase spectra,” in *Proceedings of the 110th Audio Engineering Society Convention*, Audio Engineering Society, Amsterdam, The Netherlands.

[23] Kreuzer, W., Majdak, P., and Chen, Z. (**2009**). “Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range,” The Journal of the Acoustical Society of America **126**(3), 1280–1290.

[24] Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1995**). “On the minimum-phase approximation of head-related transfer functions,” in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE, New Paltz, NY, USA, pp. 84–87.

[25] Lüthi, M., Blanc, R., Albrecht, T., Gass, T., Goksel, O., Büchler, P., Kistler, M., Bousleiman, H., Reyes, M., and Cattin, P. (**2012**). “Statismo-A framework for PCA based statistical models,” The Insight Journal **2012**, 1–18.

[26] Majdak, P., Goupell, M. J., and Laback, B. (**2010**). “3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training,” Attention, Perception & Psychophysics **72**(2), 454–469.

[27] Middlebrooks, J. C. (**1999**). “Individual differences in external-ear transfer functions reduced by scaling in frequency,” The Journal of the Acoustical Society of America **106**(3), 1480–1492.

[28] Middlebrooks, J. C., and Green, D. M. (**1992**). “Observations on a principal components analysis of head-related transfer functions,” The Journal of the Acoustical Society of America **92**(1), 597–599.

[29] Middlebrooks, J. C., Macpherson, E. A., and Onsan, Z. A. (**2000**). “Psychophysical customization of directional transfer functions for virtual sound localization,” The Journal of the Acoustical Society of America **108**(6), 3088–3091.

[30] Møller, H. (**1992**). “Fundamentals of binaural technology,” Applied Acoustics **36**(3), 171–218.

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

[32] Royston, J. P. (**1983**). “Some techniques for assessing multivarate normality based on the Shapiro-Wilk W,” Journal of the Royal Statistical Society: Series C (Applied Statistics) **32**(2), 121–133.

[33] Rui, Y., Yu, G., Xie, B., and Liu, Y. (**2013**). “Calculation of Individualized Near-Field Head-Related Transfer Function Database Using Boundary Element Method,” in *Proceedings of the 134th Audio Engineering Society Convention*, Audio Engineering Society, Rome, Italy.

[34] Seeber, B. U., and Fastl, H. (**2003**). “Subjective selection of non-individual head-related transfer functions,” in *Proceedings of the 2003 International Conference on Auditory Display*, Boston, MA, USA.

[35] Tsui, B., and Kearney, G. (**2018**). “A Head-Related Transfer Function Database Consolidation Tool For High Variance Machine Learning Algorithms,” in *Proceedings of the 145th Audio Engineering Society Convention*, New York, NY, USA, p. 5.

[36] Watanabe, K., Iwaya, Y., Suzuki, Y., Takane, S., and Sato, S. (**2014**). “Dataset of head-related transfer functions measured with a circular loudspeaker array,” Acoustical Science and Technology **35**(3), 159–165.

[37] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). “Localization using nonindividualized head-related transfer functions,” The Journal of the Acoustical Society of America **94**(1), 111–123.

[38] Wightman, F. L., and Kistler, D. J. (**1989**). “Headphone simulation of free-field listening. I: stimulus synthesis,” The Journal of the Acoustical Society of America **85**(2), 858–867.

[39] Xu, S., Li, Z., and Salvendy, G. (**2008**). “Improved method to individualize head-related transfer function using anthropometric measurements,” Acoustical Science and Technology **29**(6), 388–390.

[40] Yamamoto, K., and Igarashi, T. (**2017**). “Fully perceptual-based 3D spatial sound individualization with an adaptive variational autoencoder,” Association for Computing Machinery (ACM) Transactions on Graphics **36**(6), 1–13.

[41] Ziegelwanger, H., Kreuzer, W., and Majdak, P. (**2015**). “Mesh2hrtf: Open-source software package for the numerical calculation of head-related transfer functions,” in *Proceedings of the 22nd International Congress on Sound and Vibration*, Florence, Italy.

[42] Ziegelwanger, H., Kreuzer, W., and Majdak, P. (**2016**). “A priori mesh grading for the numerical calculation of the head-related transfer functions,” Applied Acoustics **114**, 99–110.

[43] Ziegelwanger, H., Majdak, P., and Kreuzer, W. (**2015**). “Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization,” The Journal of the Acoustical Society of America **138**(1), 208–222.

[44] Zotkin, D. N., Duraiswami, R., and Davis, L. S. (**2002**). “Customizable auditory displays,” in *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan.

[45] Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A. (**2006**). “Fast head-related transfer function measurement via reciprocity,” The Journal of the Acoustical Society of America **120**(4), 2202–2215.