A Wide Dataset of Ear Shapes and Pinna-related transfer functions generated by Random Ear Drawings: WiDESPREaD

Corentin Guezenoc¹ and Renaud Séguiern
IETR / CNRS UMR 6164, CentraleSupélec, avenue de la Boulaie, 35510 Cesson-Sévigné, France¹

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) simulations. 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, due to a non-linearity in the process of deriving PRTF sets from pinna morphology. 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 simulations.

©2019 Acoustical Society of America. [http://dx.doi.org(DOI number)]

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 (Møller, 1992). They are the cornerstone of a technique called binaural synthesis that allows to create 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 (Wenzel et al., 1993). Thus, a lot of work has been done for the past decades towards user-friendly HRTF individualization, among which four categories can be identified (Guezenoc and Seguier, 2018). Acoustical measurement (Wightman and Kistler, 1989) is the state-of-the-art method and relies on a heavy measurement apparatus and is time-intensive. Numerical simulation allows to simulate HRTFs from a 3D scan of a listener’s morphology (Kahana et al., 1998). 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, par-

¹corentin.guezenoc@centralesupelec.fr;
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 (Hu et al., 2008; Middlebrooks, 1999; Zotkin et al., 2002) or on perceptual feed-back (Hwang et al., 2008; Middlebrooks et al., 2000; Seeber and Fastl, 2003; Yamamoto and Igarashi, 2017) and often rely heavily on HRTF databases.

However, currently available measured HRTF databases (Algazi et al., 2001; Bomhardt et al., 2016; Brinkmann et al., 2019; Carpentier et al., 2014; Majdak et al., 2010; Watanabe et al., 2014) are small compared to the dimensionality of the data. Indeed, the largest that we know of, the ARI database (Majdak et al., 2010), features 120 subjects, while the dimension of a typical high-resolution HRIR set (Bomhardt et al., 2016) is about $1.2 \cdot 10^6$ (256 time-domain samples × 2300 directions × 2 ears). While work has been done towards combining existing databases (Andreopoulou and Roginska, 2011; Tsui and Kearney, 2018), 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 (Brinkmann et al., 2019; Jin et al., 2014; Kaneko et al., 2016; Rui et al., 2013). However, to the best of our knowledge, only the HUTUBS database (Brinkmann et al., 2019) is fully public. Moreover, such datasets are not larger than acoustically measured ones. Indeed, the largest that we know of, HUTUBS (Brinkmann et al., 2019), 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 data by performing Principal Component Analysis (PCA) on each dataset. Although it is a coarse machine learning technique whose limitations include linearity, PCA is a good starting point thanks to its algorithmic simplicity and high interpretability. We find that the transformation from ear shape to PRTFs is non-linear and that PCA modeling works best on ear shapes. 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 (Asano et al., 1990). 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 $n = 119$ human left ear scans. It was constituted in previous work by Ghorbal, Séguiere and Bonjour (Ghorbal et al., 2019), where pinna shapes were acquired using a commercial structured-light-based scanner, before being aligned, registered and normalized in size. The resulting meshes feature $n_v = 18176$ vertices and 35750 triangular faces.

In the following, we note $E = \{e_1, e_2, \ldots, e_n\}$ this set of $n$ ear point clouds represented as column vectors $e_i$ of $\mathbb{R}^{3n_v}$, where $3n_v = 54528$.

III. NUMERICAL SIMULATIONS OF PRTFS

For all ear shape $e_i$ in $E$, we simulate numerically the corresponding PRTF set $p_i \in \mathbb{C}^{n_f \times n_d}$, where $n_f$ and $n_d$ denote respectively the number of frequency bins and the number of directions of measurements on a spherical grid. Simulations are carried out using the Fast-Multipole Boundary Element Method (FM-BEM) (Gumerov and Duraiswami, 2005), thanks to the mesh2hrtf software developed by the Acoustics Research Institute (ARI) team (Ziegelwanger et al., 2015a,b).
Fig. 1. Ear shape $e_1$ after closure of the ear canal hole and completion with a disk-like basis mesh.

We denote
\[ \varphi : \mathbb{R}^{3n_v} \rightarrow \mathbb{C}^{n_f \times n_d} \]
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 $(n_f - 1) = 160$ frequency bins from 0.1 to 16 kHz by steps of 100 Hz. Frequency zero is not simulated and padded during the post-processing step (see Subsection III C).

A. Mesh closing and grading

Prior to simulation, we close 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. These steps are scripted in Blender\(^1\) Python and performed automatically using various Blender built-in mesh treatments.

Then, we re-mesh the closed surface using the progressive mesh grading algorithm proposed in (Ziegelwanger et al., 2016) and made available on-line as an OpenFlipper\(^2\) plug-in by Ziegelwanger, Kreuzer and Majdak. In this case, we use the cosine-based approach with the grading factor set to 10. In order to reduce the computational cost, target mesh resolution is adapted to each of four frequency bands. An example of simulation-ready mesh is displayed in Figure 1.

B. Simulation settings

According to the reciprocity principle (Zotkin et al., 2006), a sound source is placed at the entry of the filled ear canal, while virtual microphones are placed on 5 spherical arrays centered on the pinna: an icosahedral geodesic polyhedron of frequency 256 ($n_d = 2562$ vertices) for radii of 2 m and 1 m, and an equiangular vertical polar grid with an angular resolution of 5° ($n_d = 2522$ vertices) for radii of 2 m, 1 m and 0.5 m. The two types of spherical grids are displayed in Figure 2.

The sound source is 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 is set to infinitely reflective.

C. Post-processing

Once the simulation of a PRTF set is complete, it is post-processed as follows. Let $p_{raw} \in \mathbb{C}^{(n_f - 1) \times n_d}$ be a PRTF set simulated for $(n_f - 1) = 160$ frequency bins that exclude the constant component and for $n_d$ vertices of a spherical grid.

First, PRTFs are padded in frequency zero using the values of PRTFs at the first simulated frequency i.e. 100 Hz.

Then, diffuse field equalization (Middlebrooks, 1999) is 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 $d = 1, \ldots n_d$,

\[ p(f, d) = \frac{p_{raw}(f, d)}{c(f)}, \]
where the CTF $c \in \mathbb{C}^{n\times 3}$ is obtained by calculating a Voronoi-diagram-based (Augenbaum and Peskin, 1985) weighted average of the log-magnitude spectra of $p$ over all directions $d = 1, \ldots, n_d$, then deriving the corresponding minimal phase spectrum.

Based on the classical modeling of HRTFs as a combination of minimum phase spectra and pure delays (Kulkarni et al., 1995), we then extract the time-of-arrivals (TOA) from the PRIRs before re-introducing them as pure delays into the minimal phase spectra derived from the magnitude PRTFs.

IV. PCA OF EAR SHAPES

From the set of ear shapes $E$ described in Section II, we classically construct a statistical shape model of the ear using PCA (Cootes et al., 1995). Let there be $X_E = (e_1, \ldots, e_n)^t \in \mathbb{R}^{n \times 3n_v}$ the data matrix, $\bar{e} = \frac{1}{n} \sum_{i=1}^{n} e_i$ the average ear shape and $\tilde{X}_E = (e \ldots e)^t \in \mathbb{R}^{n \times 3n_v}$ the matrix constituted of the average shape stacked $n$ times. Finally, let $\Gamma_E \in \mathbb{R}^{3n_v \times 3n_v}$ be the covariance matrix of $X_E$:

$$\Gamma_E = \frac{1}{n-1} (X_E - \bar{X}_E)^t (X_E - \bar{X}_E).$$  \hfill (3)

PCA can thus be written as

$$Y_E = (X_E - \bar{X}_E) U_E^t,$$  \hfill (4)

where $U_E$ is obtained by diagonalizing $\Gamma_E$

$$\Gamma_E = U_E^t \Sigma_E^2 U_E,$$  \hfill (5)

with $\Sigma_E^2 \in \mathbb{R}^{(n-1) \times (n-1)}$ a diagonal matrix that contains its eigenvalues $\sigma_{E_1}^2, \sigma_{E_2}^2, \ldots, \sigma_{E_{n-1}}^2$

$$\Sigma_E^2 = \begin{bmatrix}
\sigma_{E_1}^2 \\
\vdots \\
\sigma_{E_{n-1}}^2
\end{bmatrix}$$  \hfill (6)

ordered so that $\sigma_{E_1}^2 \geq \sigma_{E_2}^2 \geq \cdots \geq \sigma_{E_{n-1}}^2$, and with $U_E \in \mathbb{R}^{(n-1) \times 3n_v}$ an orthogonal matrix that contains the corresponding eigenvectors $u_{E_1}, u_{E_2}, \ldots, u_{E_{n-1}} \in \mathbb{R}^{3n_v}$

$$U_E = \begin{bmatrix}
u_{E_1} \\
u_{E_2} \\
u_{E_{n-1}}
\end{bmatrix}.$$

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 samples $n$ is lower than the data dimension $3n_v$ and the data is centered, thus

$$r = \text{rank} (X_E - \bar{X}_E) \leq n - 1.$$  \hfill (8)

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

$$\text{rank} (\Gamma_E) \leq \min (r, r) = r \leq n - 1.$$  \hfill (9)

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 reconstruct the corresponding ear shape $e_{v_j}(\lambda)$ by inverting Equation (4)

$$e_{v_j}(\lambda) = (0 \ldots 0 \lambda \sigma_{E_j} 0 \ldots 0) U_E + \bar{e}.$$  \hfill (10)

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 of concha shape and $y$-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, as mentioned above, HRTFs can be well modeled by a combination of minimum phase spectrum and pure delay (Kulkarni et al., 1995). Another one is the fact that, the time-of-arrival (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 is chosen for its coherence with human perception.
Further on, for all PRTF set $p \in \mathbb{C}^{n_f \times n_d}$ we denote

$$q = 20 \cdot \log_{10}(|p|) \in \mathbb{R}^{n_f \times n_d} \quad (11)$$

the corresponding set of log-magnitude spectra, where the $|\cdot|$ and $\log_{10}$ operators are considered element-wise. Accordingly, we denote

$$\phi : \mathbb{R}^{3n_v} \rightarrow \mathbb{R}^{n_f \times n_d} \quad (12)$$

$$\phi : e \mapsto 20 \cdot \log_{10}(|\phi(e)|) \quad (12)$$

PCA is performed on the set of $n$ log-magnitude PRTF sets computed from the $n$ ear shapes, noted $Q = \{q_1, \ldots, q_n\} = \{\phi(e_1), \ldots, \phi(e_n)\}$. While most work in the literature (Kistler and Wightman, 1992; Middlebrooks and Green, 1992) stack the HRTFs of various directions and subjects prior to PCA, we choose to flatten the PRTFs from all directions for each subject separately. Thus, each vector $q_i \in \mathbb{R}^{n_f \times n_d}$ corresponds to a subject and is stacked into the data matrix $X_Q = (q_1, \ldots, q_n)^t \in \mathbb{R}^{n \times (n_f n_d)}$. This has the advantage of parsing only the across-subject variability, instead of mixing the contributions of directionality and inter-individuality into the statistical analysis. Using the same notations as in Section IV, PCA can be written as follows:

$$Y_Q = (X_Q - \bar{X}_Q) U_Q^t \quad (13)$$

where $U_Q$ is obtained by diagonalizing the covariance matrix of $X_Q$, $\Gamma_Q \in \mathbb{R}^{(n_f n_d) \times (n_f n_d)}$:

$$\Gamma_Q = U_Q^t \Sigma_Q^2 U_Q \quad (14)$$

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 $\sigma_{E_j}$.
with $\Sigma_Q^2 \in \mathbb{R}^{(n-1) \times (n-1)}$ a diagonal matrix that contains its eigenvalues $\sigma_{G_1}^2, \sigma_{G_2}^2, \ldots \sigma_{G_{n-1}}^2$ ordered so that $\sigma_{G_1}^2 \geq \sigma_{G_2}^2 \geq \cdots \geq \sigma_{G_{n-1}}^2$, and with $U_Q \in \mathbb{R}^{(n-1) \times (n-1)}$ an orthogonal matrix that contains the corresponding eigenvectors $U_Q = \begin{bmatrix} u_{Q_1} \\ \vdots \\ u_{Q_{n-1}} \end{bmatrix}$.

As in Section IV, the number of non-trivial PCs is $n-1$, due to the fact that $n < n_{fnd}$.

Various PRTF sets that illustrate the behavior of the three first PCs are reconstructed as in the ear model case (see Section IV and Equation (10)). They can be seen in Figure 4 for directions that belong to the median sagittal plane of the equiangular vertical polar grid of radius 2 meters.

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 (Jolliffe, 2002), where $p \in \{1, \ldots n-1\}$:

$$\tilde{Y}_S = \begin{bmatrix} y_{S_1,1} & \ldots & y_{S_1,p} & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{S_n,1} & \ldots & y_{S_n,p} & 0 & \ldots & 0 \end{bmatrix},$$

(17)

where $y_{S_{i,j}}$ is the value of matrix $Y_S$ 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 $U_S^t$ allows us to transform the dataset $\tilde{X}_S$ into a frame where the associated covariance matrix $\Sigma_S^2$ 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 is then reconstructed by inverting Equation (4):

$$\tilde{X}_S = \tilde{Y}_S U_S + \tilde{X}_S.$$

(18)

A simple but useful metric to evaluate the capacity of a PCA model to reduce dimensionality is the cumulative percentages of total variance (CPV) (Jolliffe, 2002, section 6.1)

$$\tau_{Sp} = 100 \cdot \left( \frac{\sum_{j=1}^{p} \sigma_{S_{j}^2}}{\sum_{j=1}^{n-1} \sigma_{S_{j}^2}} \right),$$

(19)

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 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 $e_i$ than that of PRTF sets generated from the same ear shapes $q_i = \phi(e_i)$. 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.

B. Statistical distribution

Furthermore, in order to get a better idea of the repartition of the data in both 118-dimensional linear subspaces, we test whether the PCs of each model, that is the columns of $Y_S$ where $S \in \{E, Q\}$ denotes the dataset, are normally distributed using the Shapiro-Wilk test on each PC. This is motivated by the fact that, as the PCs are by construction decorrelated (for all $S \in \{E, Q\}$ the covariance matrix of $Y_S$ is $\Sigma_S^2$ which is diagonal), if all of them are normally distributed it follows that they are statistically...
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 $\sigma$.

FIG. 5. CPV $\tau_S$ as a function of the number of retained PCs $p \in \{1, \ldots, n-1\}$ for either PCA model. Continuous line: ear shape model ($S = E$). Discontinuous line: PRTF set model ($S = Q$).

As can be seen in Figure 6, all of the ear shape model’s PCs, i.e. the columns of $Y_E$, pass the test with a significance level of 1% except for 2 PCs (the 12th and 83rd) which account for only 2.5% of the total variance.

Concerning the PCs of the PRTF set model, i.e. the columns of $Y_Q$, 8 PCs (1st, 110th, 116th, 117th and 118th) are rejected with a significance level of 1% (see Figure 6), accounting for a total variance of 12%. Interestingly, the first PC i.e. the one accounting for the most variance (9.8% of total) is rejected and presents a significantly asymmetric distribution, its skewness being of $-0.6$.

In brief, the ear model’s PC weights can reasonably be considered as multivariate-normally distributed while its PRTF counterpart’s cannot.

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. In particular, an interest-
Nevertheless, as mentioned in the introduction, linear techniques usually require larger amounts of data. 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 PRTFs 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 \( \sigma_{E,j} \) standard deviation \( \mathcal{N}(0, \sigma_{E,j}) \), where \( j \in \{1, \ldots, n - 1\} \) represents the PC index.

An arbitrarily large number \( N \) of ear shapes \( e'_1, \ldots, e'_N \in \mathbb{R}^{3n_e} \) can thus be generated as follows. First, for all \( i = 1, \ldots, N \), a PC weights vector \( \mathbf{y}_{E_i} = (y_{E_{i,1}}, \ldots, y_{E_{i,n-1}}) \in \mathbb{R}^{n-1} \) is obtained by drawing the \( (n - 1) \) PC weights \( y_{E_{i,1}}, \ldots, y_{E_{i,n-1}} \) independently according to their respective probability laws \( \mathcal{N}(0, \sigma_{E_i,j}) \), \( j = 1, \ldots, n-1 \). Second, the corresponding ear shapes are reconstructed by inverting Equation (4)

\[
\mathbf{X}'_E = \mathbf{U}_E \mathbf{Y}'_E + \mathbf{X}_E,
\]

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

\[
\mathbf{Y}'_E = \begin{bmatrix}
\mathbf{y}'_{E_{1,1}} & \cdots & y'_{E_{1,n-1}} \\
\vdots & \ddots & \vdots \\
\mathbf{y}'_{E_{N,1}} & \cdots & y'_{E_{N,n-1}}
\end{bmatrix},
\]

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

\[
\mathbf{X}'_E = \begin{bmatrix}
e'_1 \\
\vdots \\
e'_N
\end{bmatrix}.
\]

B. Ear shapes quality check

At the end of the ear shape generation process, we verify that the meshes are not aberrant and that they are fit for numerical simulation. Any mesh that presents at least one self-intersecting face is discarded.

In total, 24% (320 out of 1325) of the meshes are discarded. Thus, before going further, we check whether the selection process introduces any bias in the distribution of the PC weights.
of the ear shapes dataset. Using the Shapiro-Wilk test with a significance level of 1%, we test the normality of the distribution of the retained 1005 PC weights: only 4 PCs (12, 54, 55 and 62) accounting for 2.9% of total variance are rejected. As a reference, 3 PCs (21, 62 and 117) accounting for 1.2% of total variance are rejected when performing the test on the PC weights of all drawn 1325 ear shapes.

We conclude that the selection process does not introduce any major bias in the statistical distribution of the generated ear shapes and thus pursue with the 1005 retained ones. For simplicity, we consider further on that \( N \) is the number of retained meshes i.e. \( N = 1005 \).

C. Numerical simulation

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

\[
p_j' = \phi \left( e_j' \right), \quad \forall j = 1, \ldots, N. \tag{23}
\]

Computing time for the simulation of the 1005 PRTF sets is 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. In guise of example, the first 10 ear shapes and matching PRTF sets of the WiDESPREaD dataset are displayed in Figure 7. These first 10 randomly drawn subjects illustrate quite 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 ear shapes, we constituted a corresponding dataset of 119 PRTF sets by FM-BEM numerical simulations. 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, due to a non-linearity in the process of deriving PRTF sets from pinna morphology. 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 simulation. The resulting dataset of 1005 ear meshes and corresponding PRTF sets, named Wide Dataset of Ear Shapes and Pinna-Related transfer functions obtained by Random Ear Drawings (WiDESPREaD), is made freely available to fellow researchers.

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 non-linear character 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 features about 10 times more subjects than any available HRTF dataset, we think it can be an interesting dataset on which to try 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 (Algazi et al., 2001) 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 (Lüthi et al., 2012, 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 bust and right pinna could extend the method to the synthesis of full-bust HRTFs.
FIG. 7. (color online) Visualization of the first 10 subjects of WiDESPREaD. (a) Synthetic ear shapes $e'_1, \ldots, e'_{10}$. Color represents the vertex-to-vertex euclidean distance to the generative model’s average $\bar{e}$. (b) Log-magnitude PRTF sets $20 \cdot \log_{10}(p'_1), \ldots, 20 \cdot \log_{10}(p'_{10})$ displayed in the median sagittal plane.

ACKNOWLEDGMENTS

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

1https://www.blender.org/
2http://www.openflipper.org/


