



**HAL**  
open science

# Dataset Augmentation and Dimensionality Reduction of Pinna-Related Transfer Functions

Corentin Guezenoc, Renaud Segquier

► **To cite this version:**

Corentin Guezenoc, Renaud Segquier. Dataset Augmentation and Dimensionality Reduction of Pinna-Related Transfer Functions. Audio Engineering Society Convention, May 2020, Vienna, Austria. 10.17743/aesconv.2020.978-1-942220-32-9 . hal-02484088

**HAL Id: hal-02484088**



**<https://hal.science/hal-02484088v1>**

Submitted on 7 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Dataset Augmentation and Dimensionality Reduction of Pinna-Related Transfer Functions

 Corentin Guezenoc\* and  Renaud Séguier†

*FAST Research Team  
IETR (CNRS UMR 6164)  
CentraleSupélec  
Rennes, France*

Efficient modeling of the inter-individual variations of head-related transfer functions (HRTF) is a key matter to the individualization of binaural synthesis. In previous work, we augmented a dataset of 119 pairs of ear shapes and pinna-related transfer functions (PRTFs), thus creating a wide dataset of 1005 ear shapes and PRTFs generated by random ear drawings (WiDESPREaD) and acoustical simulations. In this article, we investigate the dimensionality reduction capacity of two principal component analysis (PCA) models of magnitude PRTFs, trained on WiDESPREaD and on the original dataset, respectively. We find that the model trained on the WiDESPREaD dataset performs best, regardless of the number of retained principal components.

## I. INTRODUCTION

Head-related transfer functions (HRTFs) individualization is a key matter in binaural synthesis [1]. A lot of work has been done towards proposing user-friendly personalization methods, either based on anthropometric measurements [2–4] or on perceptual feed-back [5–9], most of which rely on HRTF databases. Although some of these approaches propose to select a best-mach HRTF set among a database [6], others rely on a statistical model of the inter-individual variations of HRTF sets [4, 7–9]. In particular, an interesting direction of work consists in fitting an HRTF model to the listener based on his perceptual feed-back [7–9], either by tuning himself the model parameters, or by using an optimization process that prompts the listener for perceptual feed-back. In this context, tuning time must be kept as low as possible. Thus, it is essential for the HRTF model to be as compact as possible, i.e. having as few parameters as possible.

However, the databases that are currently available have few subjects compared to the dimensionality of the data. Indeed, the largest one, the Acoustics Research Institute (ARI) dataset [10], includes HRTF sets for 201 subjects, while the dimensionality of the data is of roughly  $1.2 \cdot 10^6$  (256 time-domain samples  $\times$  2300 directions  $\times$  2 ears) for a typical high-resolution HRTF set [11]. To address this matter, in previous work [12] we proposed a method to augment a dual dataset of 119 ear point clouds [13] and corresponding pinna-related transfer functions (PRTFs). The method consists in drawing new ear shapes according to the observed statistical distribution then computing the corresponding PRTF sets by means of fast-multipole boundary element method (FM-BEM). The resulting augmented dataset, named WiDESPREaD (wide dataset of ear shapes and pinna-related transfer functions generated

by random ear drawings) includes 1005 artificial subjects and is freely available on the sofacooustics website<sup>1</sup>.

In the present article, we look into how using this new dataset instead of the original one for training improves the capacity of principal component analysis (PCA) to reduce the dimensionality of log-magnitude<sup>2</sup> PRTF sets. To this end, a 20-fold cross-validation of PCA is performed on each dataset. The training and validation reconstruction errors are then compared for various numbers of retained principal components (PCs).

The paper is organized as follows. First, we present the original dataset of 119 pairs of ear shape and corresponding simulated PRTF sets. Second, we summarize the process of data augmentation that led to the creation of the WiDESPREaD dataset. Third, we present the construction of two PCA models of magnitude PRTF set, trained on the original and augmented dataset, respectively. Fourth, we discuss the dimensionality reduction capacity of both PCA models by studying the cumulative proportion of variance, then by performing a 20-fold cross-validation in order to evaluate their performance on data left out of the training set. Finally, we discuss the results.

## II. A DATASET OF 3-D EAR SCANS AND MATCHING SIMULATED PRTFS

The original dataset is composed of pairs of ear mesh and matching simulated PRTF set for 119 human subjects.

The database of 119 ear meshes was constituted in previous work by Ghorbal *et al.* [13]. As part of that work, 3D scans were acquired using a structure-light based scanner, before being normalized in size and rigidly aligned.

---

\*Electronic address: corentin.guezenoc@centralesupelec.fr

†Electronic address: renaud.seguier@centralesupelec.fr

<sup>1</sup> <https://www.sofacooustics.org/data/database/widespread>

<sup>2</sup> We focus here on the magnitude spectra, as the perceptual defects due to a lack of individualization mostly derive from the distortion of the spectral cues [14].

Finally, the point clouds were registered: every point cloud has the same number  $n_v = 18176$  of vertices, and the vertex indexing is semantically coherent from one subject to the other. Meshes can be derived from the point clouds thanks to a set of 35750 triangular faces defined identically for all point clouds by the indices of the  $n_v$  vertices.

In order to compute the PRTF sets, we closed the ear meshes using a cylinder-like support mesh. We then performed the acoustical simulations by FM-BEM using the mesh2hrtf [15] software by Ziegelwanger *et al.* This step is described in more details in [12].

### III. DATASET AUGMENTATION

In a previous paper [12], exploiting the initial dataset of 119 pairs of ear shapes and PRTFs, we proposed a method to generate new examples based on a generative model of 3-D ear shapes and boundary element method simulations. As a result, the WiDESPREaD dataset was created and made available to other researchers, featuring 1005 artificial subjects. In this section, we summarize the data augmentation process (the curious reader can refer to [12] for more details).

#### A. Statistical Ear Shape Model

Thanks to the fact that the ear point clouds are in correspondence, we were able to perform a PCA.

By concatenating the  $x$ ,  $y$  and  $z$  coordinates, every point cloud was represented as a row vector  $\mathbf{e}_i$  of  $\mathbb{R}^{3n_v}$ , where  $i = 1, \dots, N_O$ , with  $N_O$  the number of subjects in the original dataset. Let us denote by  $\mathbf{E} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{N_O}\}$  the dataset of 119 ear point clouds. The corresponding data matrix  $\mathbf{X}_E \in \mathbb{R}^{N_O \times 3n_v}$  was constructed by stacking the ear point cloud row vectors vertically  $\mathbf{X}_E = (\mathbf{e}_1^t, \dots, \mathbf{e}_{N_O}^t)^t$ . Additionally, let  $\bar{\mathbf{e}} = \frac{1}{N_O} \sum_{i=1}^{N_O} \mathbf{e}_i$  be the average ear shape and  $\bar{\mathbf{X}}_E = (\bar{\mathbf{e}} \dots \bar{\mathbf{e}})^t \in \mathbb{R}^{N_O \times 3n_v}$  the matrix constituted of the average shape stacked  $N_O$  times, and let  $\mathbf{\Gamma}_E \in \mathbb{R}^{3n_v \times 3n_v}$  be the covariance matrix of  $\mathbf{X}_E$ :

$$\mathbf{\Gamma}_E = \frac{1}{N_O - 1} (\mathbf{X}_E - \bar{\mathbf{X}}_E)^t (\mathbf{X}_E - \bar{\mathbf{X}}_E). \quad (1)$$

PCA can thus be written as

$$\mathbf{Y}_E = (\mathbf{X}_E - \bar{\mathbf{X}}_E) \mathbf{U}_E^t, \quad (2)$$

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

$$\mathbf{\Gamma}_E = \mathbf{U}_E^t \mathbf{\Sigma}_E^2 \mathbf{U}_E, \quad (3)$$

with  $\mathbf{\Sigma}_E^2 \in \mathbb{R}^{(N_O-1) \times (N_O-1)}$  a diagonal matrix that contains its eigenvalues  $\sigma_{E_1}^2, \sigma_{E_2}^2, \dots, \sigma_{E_{N_O-1}}^2$

$$\mathbf{\Sigma}_E^2 = \begin{bmatrix} \sigma_{E_1}^2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \sigma_{E_{N_O-1}}^2 \end{bmatrix} \quad (4)$$

ordered so that  $\sigma_{E_1}^2 \geq \sigma_{E_2}^2 \geq \dots \geq \sigma_{E_{N_O-1}}^2$ , and with  $\mathbf{U}_E \in \mathbb{R}^{(N_O-1) \times 3n_v}$  an orthogonal matrix that contains the corresponding eigenvectors  $\mathbf{u}_{E_1}, \mathbf{u}_{E_2}, \dots, \mathbf{u}_{E_{N_O-1}} \in \mathbb{R}^{3n_v}$

$$\mathbf{U}_E = \begin{bmatrix} \mathbf{u}_{E_1} \\ \vdots \\ \mathbf{u}_{E_{N_O-1}} \end{bmatrix}. \quad (5)$$

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

#### B. Random Drawing of Ear Shapes

The PCA model was then used as a generative model to randomly draw an arbitrary large number  $N_W$  of new ear shapes.

For all subject of index  $j = 1, \dots, N_W$ , a principal component (PC) weights vector  $\mathbf{y}_{E_{Wj}} = (y_{E_{Wj,1}}, \dots, y_{E_{Wj,N_O-1}}) \in \mathbb{R}^{N_O-1}$  was obtained by drawing the  $(N_O - 1)$  PC weights  $y_{E_{Wj,1}}, \dots, y_{E_{Wj,N_O-1}}$  independently according to their respective observed probability laws  $\mathcal{N}(0, \sigma_{E_1}^2), \dots, \mathcal{N}(0, \sigma_{E_{N_O-1}}^2)$ .

Then, the corresponding ear shapes were reconstructed by inverting Equation (2)

$$\begin{bmatrix} \mathbf{e}_{W1} \\ \vdots \\ \mathbf{e}_{W_{N_W}} \end{bmatrix} = \mathbf{X}_{E_W} = \mathbf{U}_E \mathbf{Y}_{E_W} + \bar{\mathbf{X}}_E, \quad (6)$$

where  $\mathbf{Y}_{E_W} \in \mathbb{R}^{N_W \times (N_O-1)}$  is the matrix whose rows are the  $N_W$  PC weights vectors

$$\mathbf{Y}_{E_W} = \begin{bmatrix} \mathbf{y}_{E_{W1}} \\ \vdots \\ \mathbf{y}_{E_{W_{N_W}}} \end{bmatrix} = \begin{bmatrix} y_{E_{W1,1}} & \dots & y_{E_{W1,N_O-1}} \\ \vdots & \ddots & \vdots \\ y_{E_{W_{N_W},1}} & \dots & y_{E_{W_{N_W},N_O-1}} \end{bmatrix}. \quad (7)$$

#### C. Simulation of PRTF Sets

Finally,  $N_W$  PRTF sets were calculated from the  $N_W$  ear point clouds as it was done in the case of the original dataset (see Sub-Section II).

For illustration purposes, the ear shapes and PRTF sets of the first 10 artificial subjects of WiDESPREaD are displayed in Figure 1.

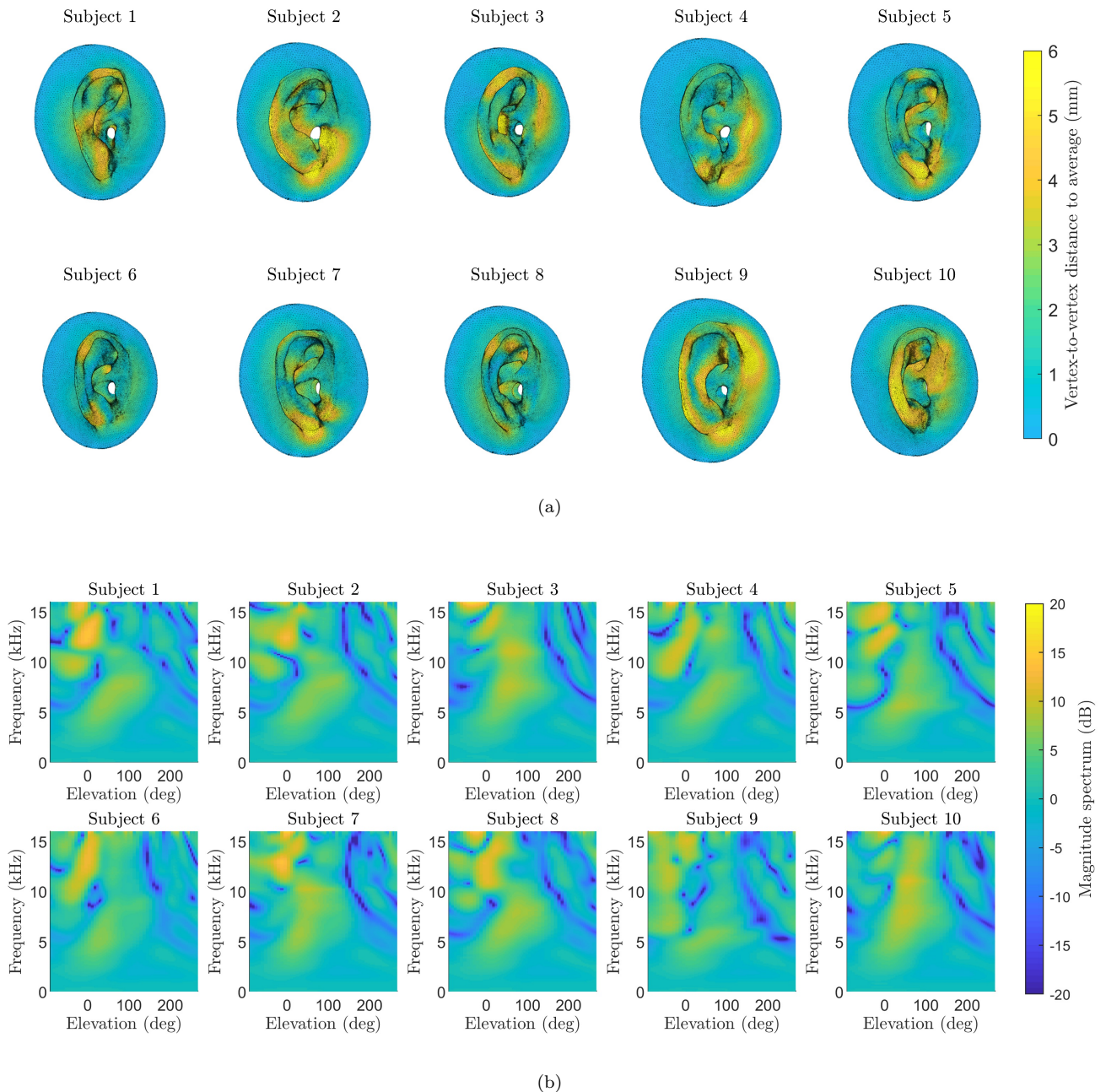


FIG. 1: Visualization of the first 10 subjects of WiDESPREaD. (a) Meshes derived from the synthetic ear shapes  $\mathbf{e}_{W_1}, \dots, \mathbf{e}_{W_{10}}$ . Color represents the vertex-to-vertex euclidean distance to the generative model's average  $\bar{\mathbf{e}}$ . (b) Log-magnitude PRTF sets  $20 \cdot \log_{10}(\mathbf{p}_{W_1}), \dots, 20 \cdot \log_{10}(\mathbf{p}_{W_{10}})$  displayed in the median sagittal plane.

#### IV. PRINCIPAL COMPONENT ANALYSIS OF MAGNITUDE PRTFS

Let us denote by  $\mathbf{p}_{O_1}, \dots, \mathbf{p}_{O_{N_O}} \in \mathbb{R}^{n_f \times n_d}$  and  $\mathbf{p}_{W_1}, \dots, \mathbf{p}_{W_{N_W}} \in \mathbb{R}^{n_f \times n_d}$  the log-magnitude PRTF sets of the original and WiDESPREaD datasets, respectively, where log-magnitude PRTFs are meant here as  $20 \cdot \log_{10}(|\cdot|)$  of the complex PRTFs.

We concatenated the filters from all  $n_d$  directions of

each log-magnitude PRTF set  $\mathbf{p}_{S_i} \in \mathbb{R}^{n_f \times n_d}$  into a row vector of  $\mathbb{R}^{n_f n_d}$ , where  $i = 1, \dots, N_S$  and  $S \in \{O, W\}$ . Further on,  $\mathbf{p}_{S_i}$  designates the row vector representation of the log-magnitude PRTF set. The logarithmic scale was chosen for its coherence with human perception.

The data matrices  $\mathbf{X}_O \in \mathbb{R}^{N_O \times n_f n_d}$  and  $\mathbf{X}_W \in \mathbb{R}^{N_W \times n_f n_d}$  were then constituted by stacking vertically the log-magnitude PRTF sets:  $\mathbf{X}_O = (\mathbf{p}_{O_1}^t \dots \mathbf{p}_{O_{N_O}}^t)^t$

and  $\mathbf{X}_W = (\mathbf{p}_{W_1}^t \dots \mathbf{p}_{W_{N_W}}^t)^t$ . PCA was then performed on each dataset as described by Equations (2) and (4) in the case of ear point clouds.

## V. DIMENSIONALITY REDUCTION CAPACITY

PCA can be used as a dimensionality reduction technique by only retaining the first  $m$  PCs [16], where  $m \in \{1, \dots, N_S - 1\}$  and  $S \in \{O, W\}$  designates the dataset:

$$\tilde{\mathbf{Y}}_S^{(m)} = \begin{bmatrix} y_{S1,1} & \dots & y_{S1,m} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{SN_S,1} & \dots & y_{SN_S,m} & 0 & \dots & 0 \end{bmatrix}, \quad (8)$$

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

### A. Cumulative Percentage of Total Variance

Approximated data can be then reconstructed by inverting Equation (2):

$$\tilde{\mathbf{X}}_S^{(m)} = \tilde{\mathbf{Y}}_S^{(m)} \mathbf{U}_S + \bar{\mathbf{X}}_S. \quad (9)$$

A simple but useful metric to evaluate the dimensionality reduction capacity of a PCA model is the cumulative percentage of total variance (CPV) [16, section 6.1]:

$$\text{CPV}_S(m) = 100 \cdot \left( \frac{\sum_{j=1}^m \sigma_{S_j}^2}{\sum_{j=1}^{N_S-1} \sigma_{S_j}^2} \right), \quad (10)$$

where  $S \in \{O, W\}$  represents either the original or the WiDESPREaD set of log-magnitude PRTFs and where  $m \in \{1, \dots, N_S - 1\}$  is the number of retained PCs.

Let us note that CPV is closely related to the dimensionality reduction-related mean-square reconstruction error (MSE) of the training set [16, section 6.1]. This relation can be expressed as follows:

$$\text{CPV}_S(m) = 100 \cdot \frac{1 - \text{MSE}(\tilde{\mathbf{X}}_S^{(m)}, \mathbf{X}_S)}{\text{MSE}(\bar{\mathbf{X}}_S, \mathbf{X}_S)}, \quad (11)$$

where

$$\text{MSE}(\mathbf{A}, \mathbf{B}) = \frac{1}{q} \frac{1}{r} (\mathbf{A} - \mathbf{B}) (\mathbf{A} - \mathbf{B})^t, \quad (12)$$

for all  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{q \times r}$  and  $q, r \in \mathbb{N}$ .

CPVs of both original and WiDESPREaD log-magnitude PRTF PCA models are plotted in Figure 2. In the case of the original dataset, we observe, as in [12], that most PCs i.e. 95% (112 out of 118) of them are required to reach the threshold of 99% of the total variance.

It would seem that we can not find a linear sub-space of the space generated by our  $N_O = 119$  examples in which log-magnitude PRTF sets are well represented. As far as we can see, a first possible explanation is that log-magnitude PRTF sets populate a linear sub-space of  $\mathbb{R}^{n_f n_d}$  whose dimension is greater than  $N_O - 1 = 118$ . A second explanation may be the existence of a non-linear manifold which would cause PCA to require all principal components to describe accurately the training set, independently of the number of training examples. Either way, it is hard to conclude without more subjects, which initially motivated the creation of the WiDESPREaD dataset.

It can be observed that, for equal numbers of retained PCs, the CPV is lower in the original case than in the WiDESPREaD one. For instance, a CPV of 99% is reached using 112 PCs for the first while many more, i.e. 866, are needed for the latter. Although this may appear as a regression, let us bear in mind that the CPV represents a proportion of the total variance observed in the dataset. However, the total variance is not the same in both datasets. Indeed, the examples of magnitude PRTFs in WiDESPREaD are more numerous and were synthesized so that their statistical distribution was realistic. Hence, as expected, the examples of this dataset are more diverse than that of the original dataset, resulting in a higher total variance. In other words, the new PCA model captures variations of magnitude PRTFs that were not present in the original dataset, and naturally uses more principal components to do so.

In the case of WiDESPREaD, the 99% CPV threshold is reached by retaining 86% (866 out of 1004) of the PCs. In other words, the last 138 principal components are not needed to approximate log-magnitude PRTF sets of the training set with reasonable accuracy. The ratio of PCs to be retained is noticeably smaller than in the case of the original dataset, which is an indication that the model trained on WiDESPREaD PRTFs may perform better at representing log-magnitude PRTF sets in general.

### B. 20-Fold Cross-Validation

In order to assess and compare the capacity of both PCA models to generalize to new examples, we performed a 20-fold cross-validation for each one of them.

Each dataset  $S \in \{O, W\}$  was equally divided in  $K = 20$  sub-groups, each containing about 5% of the subjects. Each sub-group of index  $k = 1, \dots, K$  was then used in turn as a validation set for a PCA model trained the subjects of the remaining  $K - 1$  folds. For all  $k = 1, \dots, K$  and for all dataset  $S \in \{O, W\}$ , let there be  $I_{S_{\text{train},k}} \subset \{1, \dots, N_S\}$  and  $I_{S_{\text{val},k}} \subset \{1, \dots, N_S\}$  the sets of subject indices that constitute the  $k^{\text{th}}$  fold's training and validation sets, respectively.

Let there be a fold  $k = 1, \dots, K$  and a dataset  $S \in \{O, W\}$ . PCA was performed on the data matrix  $\mathbf{X}_{S_{\text{train},k}} = (\mathbf{p}_{S_i})_{i \in I_{S_{\text{train},k}}}$ . Re-writing Equation (2) using

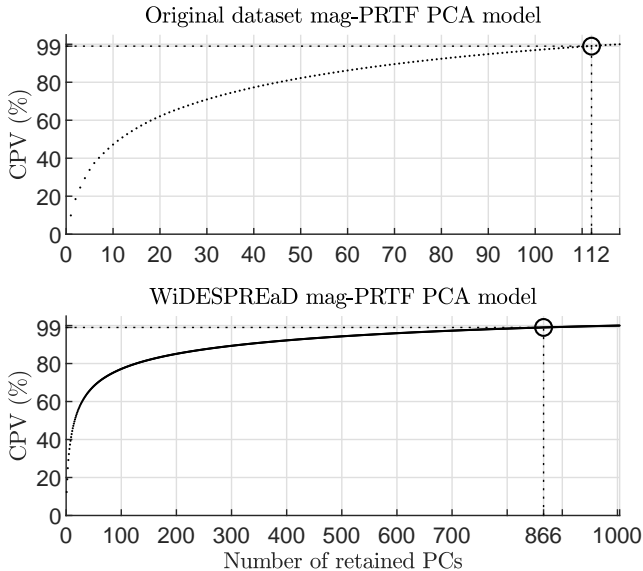


FIG. 2: Cumulative percentage of total variance (CPV) of log-magnitude PRTF PCA models as a function of number of retained principal components. Please note that the x-axes are on different scales, due to the large difference in number of subjects (and thus number of PCs) between models. Up: Original dataset. Down: WiDESPREaD.

this notation, the PCA transform can be written:

$$\mathbf{Y}_{S_{\text{train},k}} = (\mathbf{X}_{S_{\text{train},k}} - \bar{\mathbf{X}}_{S_{\text{train},k}}) \mathbf{U}_{S_{\text{train},k}}^t. \quad (13)$$

Examples from the validation set  $\mathbf{X}_{S_{\text{val},k}} = (\mathbf{p}_{S_i})_{i \in I_{S_{\text{val},k}}}$  were then projected in the training space as follows:

$$\mathbf{Y}_{S_{\text{val},k}} = (\mathbf{X}_{S_{\text{val},k}} - \bar{\mathbf{X}}_{S_{\text{train},k}}) \mathbf{U}_{S_{\text{train},k}}^t. \quad (14)$$

Finally, the training and validation data matrices were reconstructed from the PC weights. The number of PCs retained for reconstruction,  $m$ , varied in  $\{0, \dots, \text{card}(I_{S_{\text{val},k}}) - 1\}$ , where  $\text{card}(I_{S_{\text{train},k}}) = (K - 1) \lfloor \frac{N_S}{K} \rfloor$  is the number of training subjects. Thus, using the same notation as in Equation (8) and according to (9), training and validation sets were reconstructed according to the following equations:

$$\tilde{\mathbf{X}}_{S_{\text{train},k}}^{(m)} = \tilde{\mathbf{Y}}_{S_{\text{train},k}}^{(m)} \mathbf{U}_{S_{\text{train},k}} + \bar{\mathbf{X}}_{S_{\text{train},k}}, \quad (15)$$

and

$$\tilde{\mathbf{X}}_{S_{\text{val},k}}^{(m)} = \tilde{\mathbf{Y}}_{S_{\text{val},k}}^{(m)} \mathbf{U}_{S_{\text{train},k}} + \bar{\mathbf{X}}_{S_{\text{train},k}}. \quad (16)$$

The MSE reconstruction error was then averaged across all folds for both training sets

$$\epsilon_{\text{MSE,train}}(S) = \frac{1}{K} \sum_{k=1}^K \text{MSE} \left( \tilde{\mathbf{X}}_{S_{\text{train},k}}^{(m)}, \mathbf{X}_{S_{\text{train},k}} \right), \quad (17)$$

and validation sets

$$\epsilon_{\text{MSE,val}}(S) = \frac{1}{K} \sum_{k=1}^K \text{MSE} \left( \tilde{\mathbf{X}}_{S_{\text{val},k}}^{(m)}, \mathbf{X}_{S_{\text{train},k}} \right). \quad (18)$$

The training and validation reconstruction errors for both original and WiDESPREaD dataset are displayed in Figure 3. A first observation is that, when retaining all principal components in both models, the WiDESPREaD PCA model reconstructs validation data with a notably lower error (2.3 dB<sup>2</sup>) than the original model (6.0 dB<sup>2</sup>), that is a factor of 2.6.

More interestingly, for identical numbers of retained components, the WiDESPREaD cross-validation error is always lower than that of the original model. In particular, the lowest cross-validation error ever attained by the original model, 6.0 dB<sup>2</sup> (for  $m = 94$  retained PCs) is reached by the WiDESPREaD one using as few as 35 retained PCs. In other words, 35 PCs are sufficient to obtain a generalization error lower than when using the original model with all of its 94 PCs.

## VI. DISCUSSION

One of the observations that led to the creation of WiDESPREaD was the fact that, in the original log-magnitude PRTF PCA model, almost all PCs, i.e. 95% (112 / 118) of the PCs were necessary to retain at least 99% of the total variance observed in the dataset [12]. Reassuringly, by training in the same manner a log-magnitude PRTF PCA model on the augmented PRTF dataset, we observe that a smaller ratio of PCs of 86% (866 / 1004) is required to reach a CPV of 99% (see Sub-Section V A). Thus, it would seem that increasing the number of subjects allowed us indeed to identify a linear sub-space of  $\mathbb{R}^{N_S-1}$  that is able to contain most of the inter-individual variability of log-magnitude PRTF sets.

Nevertheless, this is to be taken with caution. Indeed, while we increased the number of subjects by a factor 8.5, the decrease in the ratio of PCs required to reach a CPV of 99% is rather modest (95% to 86%). Hence, the possibility remains that adding more subjects would keep increasing significantly the number of PCs required to represent 99% the information, which would be the case if there was a non-linear manifold, for instance.

However, cross-validation yielded promising results as to providing a compact representation of log-magnitude PRTF sets in general (see Sub-Section V B). Indeed, the PCA model based on WiDESPREaD seemed able to generalize to new examples better (lower reconstruction error) than the model trained on the original dataset.

When comparing both models with all their PCs retained, it could be expected. Indeed, approximating new data thanks to a PCA model while retaining all PCs is equivalent to a projection into the  $(N_S - 1)$ -dimensional space generated by linear combinations of the  $N_S$  training examples, and the WiDESPREaD dataset has about 9 times more examples than the original one.



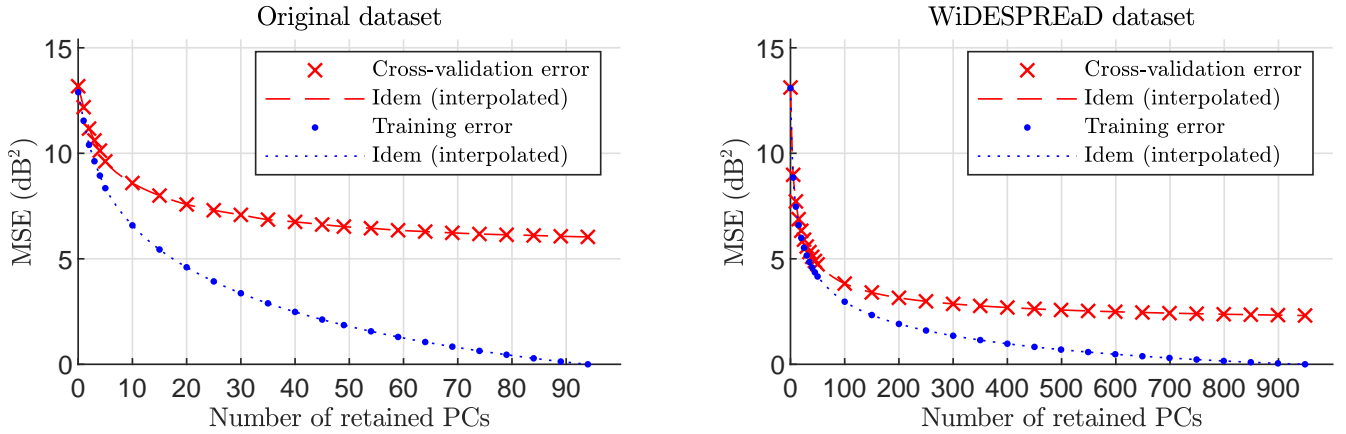


FIG. 3: Mean-square reconstruction errors (MSE) of log-magnitude PRTF sets, averaged over 20 cross-validation folds, for various numbers of retained PCs, for both original (left) and WiDESPREaD (right) datasets. Training errors are displayed as blue points while validation errors are represented as red crosses. The blue dashed and red dotted lines are a cubic interpolation of the training and validation errors, respectively. Please note that the x-axes are on different scales, due to the large difference in number of subjects between models.

However, and more interestingly, the generalization error curve for the WiDESPREaD case is lower than that of the original dataset, regardless of the number of retained PCs.

Furthermore, only 35 components (out of 949) are needed for the WiDESPREaD cross-validation reconstruction error to subceed the lowest cross-validation error ever attained in the case of the original dataset i.e. with 94 retained PCs out of 94.

## VII. CONCLUSION

In this article, we presented a dataset of 119 pairs of ear mesh and matching calculated PRTF set. Thereupon, we presented briefly how this dataset was augmented in previous work [12], resulting in a wide dataset of 1005 pairs of ear mesh and matching calculated PRTF set, named WiDESPREaD. Building upon this original dataset, we then summarized how an augmented dataset of 1005 pairs of ear mesh and matching calculated PRTF set was generated, according to a method introduced in a previous article. We then trained PCA models of log-magnitude PRTF sets on each original and augmented dataset, before

comparing their dimensionality reduction performance on both training and validation data using tools such as the cumulative proportion of total variance and k-fold cross-validation.

Overall, we found that using the WiDESPREaD dataset improved the performances of PCA at modeling and reducing the dimensionality of log-magnitude PRTF sets, in comparison with the original dataset. These results are encouraging and tend to corroborate the promising character of the process of synthetic data generation that we proposed in [12], which could benefit to applications such as HRTF adaptation based on listener feed-back or anthropometry.

As the possibility for a non-linear manifold in the space of log-magnitude PRTFs is not ruled out (see Section VI), future work includes using a non-linear dimensionality reduction technique on the database of WiDESPREaD PRTFs and compare its performance with PCA.

Furthermore, it may be interesting to use other metrics in addition to MSE to evaluate dimensionality reduction error, such as inter-subject spectral difference (ISSD) [2] and simulated sagittal plane sound localization errors using the psycho-acoustic model by Baumgartner *et al.* [17].

- 
- [1] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, *The Journal of the Acoustical Society of America* **94**, 111 (1993).
  - [2] J. C. Middlebrooks, *The Journal of the Acoustical Society of America* **106**, 1480 (1999).
  - [3] D. N. Zotkin, R. Duraiswami, and L. S. Davis, in *International Conference on Auditory Display* (Kyoto, Japan, 2002).
  - [4] H. Hu, L. Zhou, J. Zhang, H. Ma, and Z. Wu, in *Inter-*

- national Conference on Computational Intelligence and Security* (Guangzhou, China, 2006), vol. 2, pp. 1829–1832.
- [5] J. C. Middlebrooks, E. A. Macpherson, and Z. A. Onsan, *The Journal of the Acoustical Society of America* **108**, 3088 (2000).
- [6] B. U. Seeber and H. Fastl, in *International Conference on Auditory Display (ICAD)* (Boston, MA, USA, 2003).
- [7] S. Hwang, Y. Park, and Y.-s. Park, *Acta Acustica united with Acustica* **94**, 965 (2008).

- [8] K. H. Shin and Y. Park, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* **91**, 345 (2008).
- [9] K. Yamamoto and T. Igarashi, *ACM Transactions on Graphics* **36**, 1 (2017).
- [10] P. Majdak, M. J. Goupell, and B. Laback, *Attention, Perception, & Psychophysics* **72**, 454 (2010).
- [11] R. Bomhardt, M. de la Fuente Klein, and J. Fels, in *172nd Meeting of the Acoustical Society of America* (Honolulu, HI, USA, 2016), vol. 29, p. 050002.
- [12] C. Guezenoc and R. Séguier, *The Journal of the Acoustical Society of America* (2020 (submitted, preprint available here: <https://hal.archives-ouvertes.fr/hal-02506173>)).
- [13] S. Ghorbal, R. Segquier, and X. Bonjour, *US Patent* 16/300,044 (2019).
- [14] F. Asano, Y. Suzuki, and T. Sone, *The Journal of the Acoustical Society of America* **88**, 159 (1990).
- [15] H. Ziegelwanger, W. Kreuzer, and P. Majdak, in *22st International Congress on Sound and Vibration* (Florence, Italy, 2015).
- [16] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics (Springer-Verlag, New York, 2002), 2nd ed.
- [17] R. Baumgartner, P. Majdak, and B. Laback, *The Journal of the Acoustical Society of America* **136**, 791 (2014).