PRINCIPAL COMPONENT ANALYSIS OF BINAURAL HRTF PAIRS

Georgios Marentakis
Department of Computer Science and Communication
Østfold University College
georgios.marentakis@hiof.no

ABSTRACT

Principal Component Analysis (PCA) has been often used for HRTF compression and individualization. Most commonly, when creating the PCA input matrix, each ear is handled as an independent observation which essentially doubles the available observations but also the principal component weights that need to be calculated in order to reconstruct the complete dataset. It would therefore be interesting to investigate the extent to which they can be handled jointly when creating the HRTF input matrix. Here, we explore one way to handle this possibility by comparing the standard method of handling ears in the PCA input matrix to a variation in which ears are handled jointly. We performed simulations using three different HRTF databases involving linear and logarithmic HRTF magnitude spectrum and calculated the number of components required to explain 90% variance and the measure of spectral distortion. Results show that the proposed approach is not as efficient in terms of the number of components required, however, spectral distortion is not affected as much with the alternative representation. Furthermore, the resulting components provide insight into how the HRTFs of the two ears are related.

1. INTRODUCTION

Head Related Transfer Functions (HRTFs) are receiving a lot of interest as they allow designers and engineers to create 3D audio using headphones [1]. Several applications can be found in virtual and augmented reality but also music reproduction. HRTFs are specific to individual users and need to be measured for all positions of interest in a relatively resource-intensive process [2]. HRTF models that support individualisation, compact representation, and transfer are therefore important so that HRTFs can reach out to all users.

Several compact HRTF models are based on decomposing HRTF sets upon a set of orthogonal basis functions. The resulting weights (or loadings) can be used for recreating HRTFs but also for performing interpolation and other operations. Importantly, such decompositions can also be used to reduce the dimensionality of HRTF sets and serve as a basis for efficient compression, individualization, and the investigation of numerical and perceptual properties. Principal Component Analysis (PCA) e.g., [3, 4, 5, 6], the Spherical Harmonic Transform e.g., [7, 8, 9], but also Independent Component Analysis [10] have been used for this purpose. More recent approaches focus on deep learning [11].

This article focuses on applying PCA for modelling HRTFs. Motivated by the fact that HRTFs show certain similarities across ears it investigates whether both ears can be treated jointly in PCA. An approach for such joint representation is proposed and evaluated in comparison to the method used in the state of the art which treats the HRTF of each ears as an independent observation. The aim is to investigate advantages and disadvantages of such a joint representation and its potential for application in PCA models of HRTFs.

2. BACKGROUND

2.1 HRTFs

The *head-related impulse response (HRIR)* $h(\theta, \phi, t)$ denotes the time domain impulse response for a sound originating at azimuth θ and elevation ϕ. The *head-related transfer function (HRTF)* $H(\theta, \phi, f)$ is the frequency domain representation of the HRIR. HRTFs are recorded using miniature microphones for a subject and source position of interest measured at or inside the ear-canal [2]. Frequently, HRTFs are diffuse-field equalized to exclude the ear canal resonance and measurement system response and called *directional transfer functions* (DTFs).

Monaural cues such as spectral peaks and notches in the magnitude HRTF spectrum colour the sound and are used for elevation perception and front/back and up/down discrimination [12, 13]. They are typically found between 4 and 16 kHz and are the result of the superposition of sound entering the ear canal directly with sound entering after reflection by the outer ear. For example, a prominent 1-octave notch centered between 6 and 11 kHz changes systematically with the vertical source location [14].

Whereas HRTFs incorporate the effects of the whole body, *pinna-related transfer functions (PRTFs)* isolate the contribution of the pinna from the rest of the body. They can be calculated by applying a 1 ms right window at the beginning of the HRIR signal in order to eliminate reflections by torso and shoulders [15]. Such functions are helpful when relating features in the magnitude spectrum to particular anthropometric dimensions. Spectral features below 3 kHz are mainly produced by head diffraction and torso reflections [16].

2.2 Principal Component Analysis

Principal Component Analysis is normally applied on a two-dimensional matrix, with columns defining the independent variables and rows containing observations. PCA can be calculated directly using the eigendecomposition of the sample covariance matrix $\mathbf{C_Y}$ of the observations or using the Singular Value Decomposition [17]. The sample covariance matrix $\mathbf{C_Y}$ of a set of observations $\mathbf{Y}$ with M rows of observations and N columns of variables corresponding to a random vector is defined as

$$\mathbf{C_Y} = \mathbf{Y^T} \, \mathbf{Y} \, . \tag{1}$$

$\mathbf{C_Y}$ is as a symmetric, real-valued, square matrix. $\mathbf{Y}$ needs to be centered by subtracting the observation means. The eigenvectors of the covariance matrix $\mathbf{C_Y}$ are also called the principal components of $\mathbf{Y}$. Since $\mathbf{C_Y}$ is symmetric, it is also diagonalizable,

$$\mathbf{C_Y} = \mathbf{V} \, \mathbf{D} \, \mathbf{V}^{-1} \, , \tag{2}$$

with a diagonal matrix $\mathbf{D}$ ($m \times m$) containing the eigenvalues of $\mathbf{C_Y}$ and $\mathbf{V}$ as an orthonormal eigenvector matrix including the right eigenvectors as columns.

Eigenvectors and eigenvalues may also be obtained through the singular value decomposition (SVD), using which $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \mathbf{U} \, \mathbf{S} \, \mathbf{V}^T \, , \tag{3}$$

where $\mathbf{U}$ are ($m{\times}n$) and $\mathbf{V^T}$ ($n{\times}n$) orthogonal matrices including *left* and *right* eigenvectors $\mathbf{u}_k$ and $\mathbf{v}_k$, respectively. $\mathbf{S}$ ($n \times n$) is a diagonal matrix with nonzero non-negative diagonal elements, so that $\mathbf{S} = diag(s_1, \, ... \, , s_n)$, also known as *singular values*. Note that

$$\mathbf{Y^T} \, \mathbf{Y} = (\mathbf{U} \, \mathbf{S} \, \mathbf{V}^T)^T \, (\mathbf{U} \, \mathbf{S} \, \mathbf{V}^T) = \mathbf{V} \, \mathbf{S}^2 \, \mathbf{V}^T \, , \tag{4}$$

Consequently, the square root of the eigenvalues of $\mathbf{YY^T}$ are the singular values (s_k) of $\mathbf{Y}$. The original centered data $\mathbf{Y}$ set can be transformed to the new basis by projecting it on the eigenvector basis $\mathbf{V}$ to obtain the principal component weight (PCW) (or score) matrix $\mathbf{W}$ which can be used for reconstruction.

$$\mathbf{W} = \mathbf{Y} \, \mathbf{V} \text{ and } \mathbf{Y} = \mathbf{W} \, \mathbf{V}^{-1} \tag{5}$$

Assuming that the matrix $\mathbf{Y}$ has a rank r, it follows that $s_k > 0$ for $1 \leq k \leq r$ and $s_k = 0$ for $(r + 1) \leq k \leq n$ and one can neglect eigenvalues that are very close to zero to reduce the dimensionality. $\mathbf{Y}$ can thus be approximated by reducing the number of eigenvectors involved in the reconstruction.

$$\mathbf{Y}^l = \sum_{k=1}^{l} \mathbf{u}_k \, s_k \, \mathbf{v}_k^T \, + \bar{\mathbf{Y}} \, , \tag{6}$$

l is commonly chosen by calculating the number of components required to explain, say 90%, of the variance. The variance explained by l components is given by:

$$var(l) = \frac{\sum_{k=1}^{l} s_k}{\sum_{k=1}^{N} s_k} \cdot 100 \; [\%] \, , \tag{7}$$

where s_k is the k^{th} singular value, l is the number of a particular PC and N is the total number of components.

2.3 Modelling HRTFs using PCA

Head-related transfer function sets are multi-dimensional and typically include the recorded impulse response for each subject number, direction of sound incidence, and ear. To proceed with PCA, the dataset needs to be placed into a 2D input matrix before calculating principal components and associated weights. Subsequently, the number of required principal components is determined depending on the application [6, 18, 15]. Most of the studies use enough components so as 90% of the variance in the data is explained [19, 3, 16]. Promising results have been obtained when evaluating sound localization with HRTFs that have been reconstructed with a limited set of components subject to the aforementioned variance constraint [3, 20, 10]. Principal components obtained from different HTRF datasets are consistent as long as the number of measurement directions and subjects is reasonable [21]. This invariance is more evident for the first 3-5 components explaining a large amount of variance, as higher order components associated with smaller variance reflect specificities that might not be shared across datasets [22, 3].

In the literature, several approaches into applying PCA for HRTFs can be found, which vary depending on the *domain* in which ear transfer functions are represented and the *structure* of the input matrix. PCA has been applied on HRIRs [23, 24, 25, 16, 5, 26], minimum-phase HRIRs [16, 5, 26], and on HRTFs of linear [4, 27] or logarithmic magnitude [19, 28, 3, 22, 18, 29], or the real and imaginary part of the complex spectrum. Typically, signal bins are considered as the independent variables, however, more recently spatial directions have also been used as independent variables in what has been called spatial PCA [6].

Few studies have attempted a direct comparison of the aforementioned approaches. Leung and Carlile [20] investigated the PCA compression efficiency and suggested that the optimal format for PCA decomposition in terms of compression is the linear amplitude form in frequency domain. Takane *et al* [30] extended the work of Liang *et al* [31] and compared four data representations: HRIR, complex spectrum HRTF, linear spectral magnitude HRTF, and log-spectral magnitude HRTF. Sample amplitude (or frequency bin magnitude) appear on input matrix columns and comparison is done based on explained variance, signal distortion, and signal-to-distortion ratio using the KEMAR HATS database [32]. The results confirmed an advantage in using representations in the frequency domain but are somewhat inconclusive otherwise. Marentakis *et al* [33] compared different ways to structure the input matrix using numerical simulations, using three publicly available HRTF datasets with human subjects. They varied *domain* as in HRIRs with direction-dependent time delays, minimum-phase HRIRs, and HRTFs using linear or logarithmic magnitude and the *structure* of the PCA input matrix as in signal and spatial PCA. Simulations also investigated the extent to which spectral smoothing affected the results. The results pointed to a relative inefficiency of representing the transfer functions in the time-domain as representing transfer functions in the frequency domain required fewer components to explain 90% of the variance in the data set and resulted in lower spectral distortion. However, the improvement in compression efficiency and spectral distortion was not as marked as long as minimum-phase HRIRs were used. Signal PCA was more efficient in terms of the number of components required to represent 90% of the variance and resulted in lower spectral distortion for the LISTEN and CIPIC databases. For the ARI database, spectral distortion was still smaller for signal compared to spatial PCA, however, fewer components were required for explaining 90% of the variance in the dataset for spatial PCA. Finally, linear and logarithmic representations resulted in smaller differences and an advantage for logarithmic representation was observed if both compression efficiency and signal distortion are considered, especially if transfer functions are smoothed.

The way the signals from the left and right ears enter the PCA input matrix has not received much attention in literature. Sometimes only one ear is modelled and the second one is considered to be symmetric and its transfer function is estimated based on the modelled one [16, 5]. Alternatively, it can be attempted to use PCA to explain the variability across the two ears. This can be done by using the time/frequency signals from the second ear as observations in rows [28, 3] in the PCA input matrix.

2.4 Summary and Research Questions

The literature review shows that although certain differences in constructing the PCA input matrix have been investigated [20, 30, 33], studies have not focused on how HRTF measurement pairs from sounds from a single direction at the two ears are handled.

The common approach of including the ears as observations in the PCA input matrix leads to a different weight for each ear for each principal component. This leads to a good precision when modelling HRTFs, it does, however, double the number of weights that need to be stored but also adjusted when HRTFs are modified or individualized using the HRTF model. Our research question is therefore how it can be made possible to reduce the need to model ears independently in PCA models applied to HRTFs. The way we examine this here is quite simple and consists in including the signal (or spatial) bins of the second ear as variables in the columns of the PCA input matrix. In this way, one can adjust the HRTFs of both ears using a single weight for each principal component used. Such an approach has not been investigated earlier. It is therefore interesting to investigate how well it works for HRTF modelling. However, as the number of columns doubles, this approach inadvertently increases the variable to observation ratio and may therefore affect the compression efficiency and spectral distortion associated with the HRTF base. To examine this possibility, we perform here a number of simulations in which the two approaches are compared using the number of PCs required to model 90% of the signal variance and the spectral distortion of the resulting reconstructed HRTFs.

3. NUMERICAL EVALUATION

In the evaluation, several parameters were varied jointly with the way the ears entered in the PCA input matrix: the HRTF database, the *structure* of the input matrix, the *domain* in which the signal was represented, and how ears were handled (the *ear mode*).

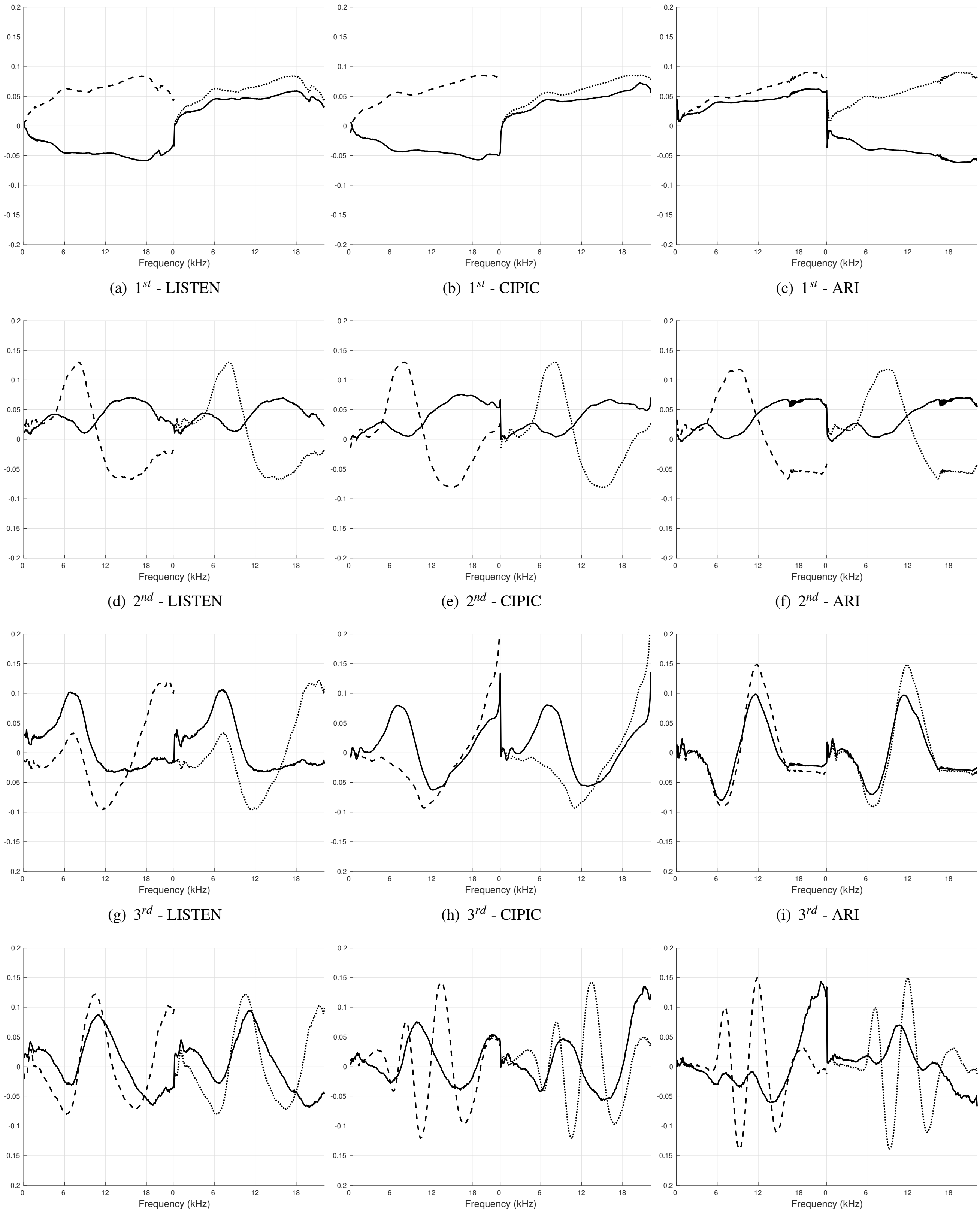
Three open access HRTF databases were used: the Acoustics Research Institute (ARI) HRTF database [34], the LISTEN database from the Institut de Recherche et Coordination Acoustique/Musique [35] and the HRTF database from the University of California at Davis (CIPIC) [36]. ARI contains HRIRs of 256 samples measured at 1550 sound locations and the first 80 subjects were used here. CIPIC includes HRIRs of 200 samples from 45 subjects measured at 1250 sound locations. The LISTEN database HRIRs of 512 samples from 50 subjects and 187 positions.

HRTFs were represented in the frequency domain and both the linear and the logarithmic amplitude representation was used in the simulation. The PCA input matrix was structured both as is observed in the signal as well as in the spatial structures described earlier. Finally, ears were handled either as observations (in the rows) or as variables (in the columns) of the PCA input matrix. The first approach will be called *ear-mode* 1 (EM1) and the second *ear-mode* 2 (EM2).

Compression efficiency was evaluated by examining the number of components required to explain 90% of the variance in the input data and by estimating the error in the reconstruction accuracy of the original HRTF set. HRTF reconstruction was evaluated in the frequency domain using the *Spectral Distortion* (SD). For an arbitrary subject s and sound incidence from at θ,ϕ, SD it is calculated by:

$$\mathrm{SD}(s,\theta,\phi) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[20\log_{10}\frac{|H(s,\theta,\phi,f_j)|}{|\hat{H}(s,\theta,\phi,f_j)|}\right]^2} \qquad (8)$$

where $H(s,\theta,\phi,f_j)$ and $\hat{H}(s,\theta,\phi,f_j)$ are measured and estimated HRTF logarithmic magnitudes respectively, and f_j refers to the frequency index, and N is the total number of frequency bins used in the calculation. The synthesized signal is more similar to the measured one when a small SD is obtained. According to [37], the spectral distortion of a reconstructed HRTF should not be greater than 5.7 dB. To measure spectral distortion, the number of PCs used in reconstruction was manipulated from one to all PCs in five steps and the signal distortion was estimated. Simulations were performed in MATLAB®.



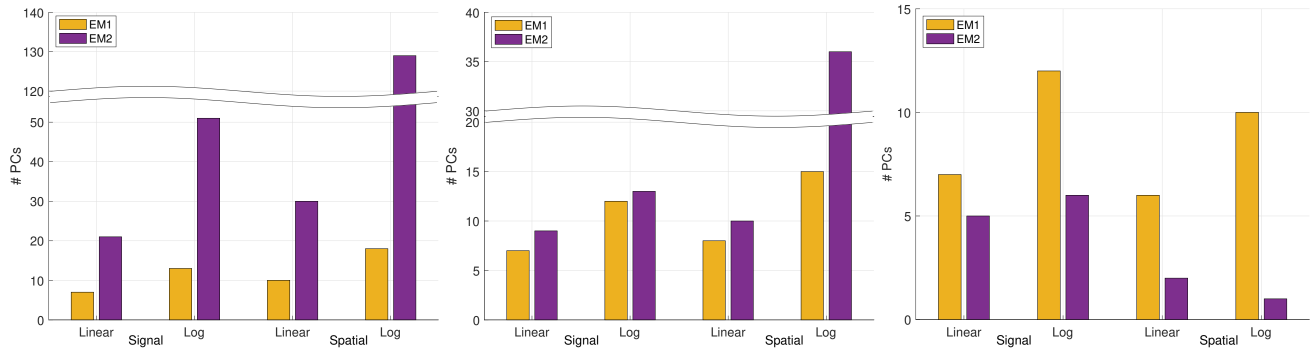
(a) 1^{st} - LISTEN (b) 1^{st} - CIPIC (c) 1^{st} - ARI

(d) 2^{nd} - LISTEN (e) 2^{nd} - CIPIC (f) 2^{nd} - ARI

(g) 3^{rd} - LISTEN (h) 3^{rd} - CIPIC (i) 3^{rd} - ARI

(j) 4^{th} - LISTEN (k) 5^{th} - LISTEN (l) 6^{th} - LISTEN

Figure 1. The first to third principal components for all three databases for EM1 and EM2 in the signal structure log-HRTF magnitude for all three HRTF databases as well as the 4th to 6th components for the LISTEN database. Solid lines show EM2 PCs, dashed lines EM1 PCs and dotted lines repeat EM1 to enable comparison with EM2. PCs for EM2 are twice as long due to the doubling of columns in the PCA input matrix.



(a) LISTEN (b) CIPIC (c) ARI

Figure 2. The number of PCs required to explain 90% of variance for the domains, structures, and ear modes used in the simulations for the three HRTF databases.

4. RESULTS

The results of the simulation are presented next starting from some comments on the resulting principal components as well as the number of components required to explain 90% of the input matrix variance and following up with the spectral distortion results.

4.1 Resulting PCs

In Figure 1, the emerging principal components for the signal *structure* log HRTF magnitude can be observed. It can be seen that the components emerging out of EM2 which handles ears jointly in the PCA input matrix bear several similarities but also important differences in comparison to the components obtained using the traditional EM1 approach. Quite interesting is the first component that is typically associated with head shadow. It can be seen that jointly including ears results in a certain symmetry or mirroring of the component for the one ear. This is a consequence of the fact that signal will be louder in the ipsilateral in comparison to the contralateral ear and would normally manifest in the principal component weights in which a similar symmetry would appear to reflect this property [21]. We see that forcing PCA to deal with both ears using a single weight results in incorporating this in the principal component rather than the loadings. It is also interesting to observe that this appears in all three databases examined here.

Observing the 2^{nd} principal component we see a similar tendency. Here, the shape of the principal component for each ear is similar in the joint EM2 model which is consistent with the observation that this component encoding contrast between high and low elevation sounds [21]. However, some differences in the shape of the PC appear compared to the EM1 model. In particular, the magnitude of the spectral peaks and notches is smaller. Furthermore, notches and peaks in EM1 become peaks and notches in EM2 while their center frequencies are somewhat shifted mainly towards lower frequencies. As reversing affects both peaks and notches, it does not represent a significant difference as it can be resolved easily. A similar tendency appears for the 3^{rd} component in which the location of peaks and notches is shifted and again inversions appear. Furthermore, the results are relatively consistent across databases for the first and second components but less so for the third.

From the fourth component onwards, the differences in the obtained components between ear modes become more marked and less easy to explain. Even though an inversion of the PC for each ear can be observed for the fourth component, the location of peaks and notches and their bandwidth also changes compared to the EM1 model. Still, one can observe that EM2 leads to incorporating possible symmetries or asymmetries with respect to the ears in the PCA model basis functions.

These observations also generally also hold for PCs obtained based on the linear HRTF magnitude (not shown here), However, the agreement between components for the two ear modes is smaller. The components obtained for the spatial *structure* show similar tendencies, however, they are less clear cut and are not discussed here. Overall, it can be seen that EM2 yields principal components that are reasonable and can relate to physical properties of sound propagation to the two ears. Without loss of generality components may be inverted. If this is considered, the components appear to be relatively consistent across databases and agree well with the ones obtained for EM1, especially for lower order.

4.2 Compression Efficiency

By observing Figure 2, it can be seen that the way ears are treated in the input matrix affects compression efficiency. For the LISTEN and CIPIC databases, in particular, treating both ear signals as independent variables (as in EM2) results in an increase in the number of components required for explaining 90% of the variance in the dataset. The increase is bigger when the frequency response is expressed in dB in comparison to the linear *domain* and for the spatial in comparison to the signal *structure*. The result is quite the opposite when considering the ARI database as placing the ears as columns in the PCA input matrix turns out to result in a decrease in the number of components required to represent 90% of the variance and this is higher for the logarithmic compared to the linear *domain* and for the signal compared to the spatial *structure*.

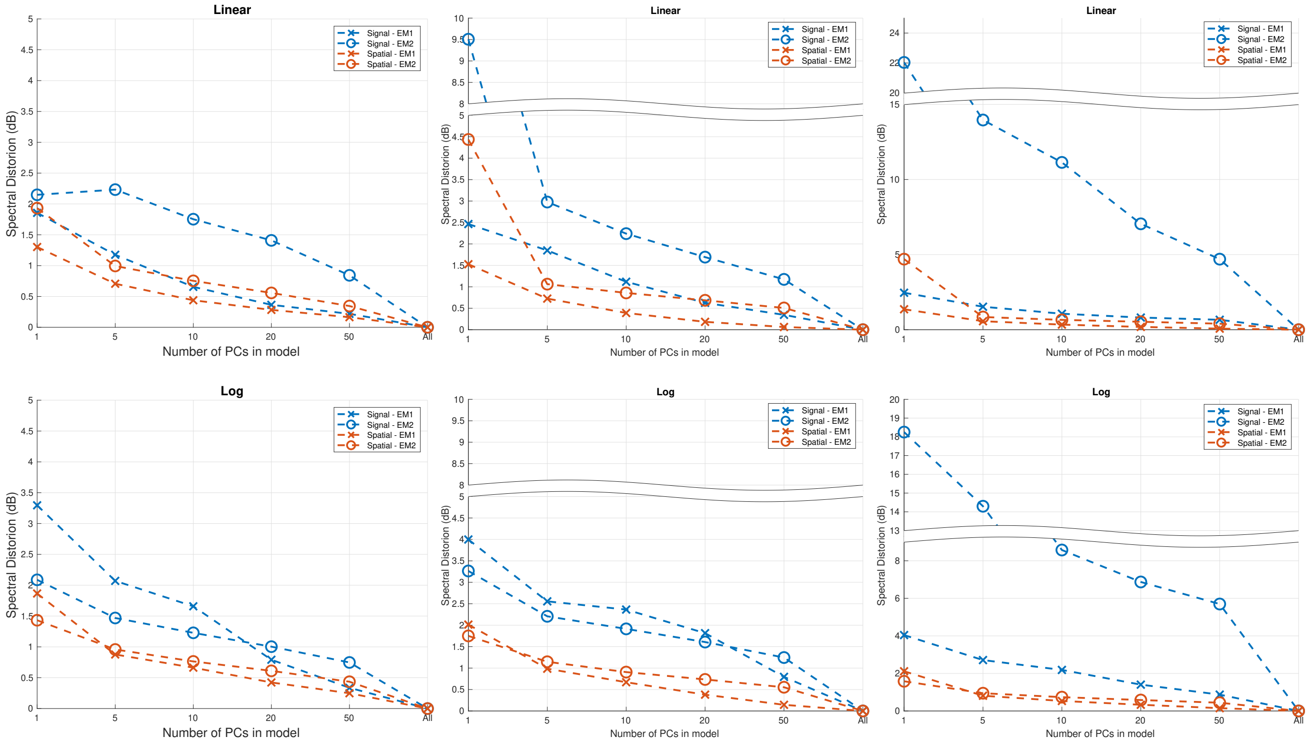


Figure 3. Spectral Distortion upon reconstruction averaged among subjects, and directions in the conditions examined in the simulations.

4.3 Spectral Distortion

By observing Figure 3, it can be seen that spectral distortion results are more complex and depend on whether HRTFs are represented in the linear or the logarithmic *domain*.

For the linear HRTF magnitude *domain*, there is a tendency for higher spectral distortion when ears are placed in the matrix columns as variables (EM2), for all three databases. However, the difference is highest when the number of components used to recreate the HRTFs is low and spectral distortion for the two ear modes becomes quite similar when the number of components used is increased. This tendency is most visible for the spatial *structure* for which spectral distortion for the two ear modes becomes very similar as soon as the number of components is increased. These observations are replicated for the three HRTF databases.

On the other hand, for the logarithmic HRTF magnitude *domain*, EM2 tends to result in lower signal distortion for the *spatial* structure. Again the difference in spectral distortion for the two models becomes smaller as the number of components used to recreate the HRTF dataset is increased. When the *structure* is spatial the difference between the two ear modes is in general quite small. These observations are replicated for the three HRTF databases.

5. DISCUSSION

Treating both ears in a joint way in the PCA input matrix led to interesting results and viable principal components which encoded physical properties of sound propagation to the two ears. Forcing the PCA model to use a single weight for both ears led to symmetries and asymmetries due to sound propagation to the two ears to be encoded in the principal components themselves rather than the principal component weights. This is particularly evident in the case of the first principal component in the signal *structure* which models the effect of head shadowing and is inverted for one ear relative to the other. This allows modelling the effect of head shadowing as a function of azimuth by switching between positive and negative values in the principal component weights. There was good agreement in the shape of the principal components obtained from EM1 and EM2 up to the third principal component after which deviations became noticeable which implies that different qualitative aspects are encoded in the components of higher order. A more detailed investigation is required to comment on this aspect and to explain the variations obtained in the spatial PCA *structure* approach.

The results indicate that EM2 results in plausible components with significant explanatory potential which relate well to the components obtained using EM1. Furthermore, there were several scenarios in which EM2 provided less signal distortion without increasing significantly the number of components required for representing the dataset with sufficient detail. It appears that this approach is best suited to HRTF datasets with large number of observations as this can counteract the doubling of independent variables required by this approach. As mentioned earlier, EM2 results in halving the weights that need to be estimated or adjusted for HRTF modelling and individualization. Even if slightly fewer components may be necessary for explaining the same amount of variance as with EM1, halving the number of weights may be quite significant improvement for a number of applications, most notably individualization. In such a case, adjusting a single weight results in manipulating the HRTFs of both ears instead of just one, which is quite important for both numerical and user controlled applications.

The number of components required to explain 90% of the variance was lower for EM1 for all configurations simulated in the LISTEN and CIPIC databases with higher differences for the logarithmic *domain* and the spatial *structure*. The situation was reversed for the ARI database for which compression efficiency was highest for EM2. The difference in the number of components required to explain 90% of the variance among the databases used here was small for signal PCA compared to spatial PCA. This difference may attributed to the large number of observations (subjects and spatial directions) in the ARI dataset compared to the LISTEN and CIPIC datasets which led to good encoding of components despite the lower variable to observation ratio of the EM2 structure. Concerning spectral distortion, for linear *domain* an advantage for EM1 is registered when few components are used to reconstruct the dataset. The difference in spectral distortion becomes smaller as long as more components are employed in HRTF reconstruction for both the signal and spatial *structure*. For the logarithmic *domain*, an advantage for EM2 is registered which again becomes smaller as long as the number of components used in the reconstruction of HRTFs is increased.

6. CONCLUSION

We presented a study that compared a novel approach for incorporating jointly data from both ears in the PCA input matrix to that standard approach of treating the two ears as independent observations. The comparison was done for HRTFs with linear or logarithmic amplitude placed in the PCA using either a signal or a spatial *structure* using three publicly available HRTF databases. The results showed that the introduced approach led to viable principal components with explanatory value. Furthermore, in certain configurations EM2 was similar or better than EM1 when it comes to the number of components required to explain 90% variance as well as spectral distortion. Further investigation is therefore required to understand better the potential of the introduced approach for HRTF modelling and individualization.

7. REFERENCES

[1] D. R. Begault, *3-D sound for virtual reality and multimedia*. Morgan Kaufmann Pub, 1994.

[2] S. Li and J. Peissig, "Measurement of head-related transfer functions: A review," *Applied Sciences*, vol. 10, no. 14, p. 5014, 2020.

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

[4] J. Qian and D. A. Eddins, "The role of spectral modulation cues in virtual sound localization," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 302–314, 2008.

[5] S. Hwang, Y. Park, and Y.-s. Park, "Modeling and customization of head-related impulse responses based on general basis functions in time domain," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 965–980, 2008.

[6] M. Zhang, Z. Ge, T. Liu, X. Wu, and T. Qu, "Modeling of individual hrtfs based on spatial principal component analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 785–797, 2020.

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

[8] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Modal expansion of hrtfs: Continuous representation in frequency-range-angle," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 285–288.

[9] ——, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2347–2357, 2010.

[10] V. Larcher, O. Warusfel, J.-M. Jot, and J. Guyard, "Study and comparison of efficient methods for 3-d audio spatialization based on linear decomposition of hrtf data," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[11] R. Miccini and S. Spagnol, "Hrtf individualization using deep learning," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2020, pp. 390–395.

[12] J. Blauert, "Spatial hearing: The psychophysics of human sound localization," 1983.

[13] E. H. A. Langendijk and A. W. Bronkhorst, "Contribution of spectral cues to human sound localization," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1583–1596, 2002.

[14] B. Zonooz, E. Arani, K. P. Körding, P. R. Aalbers, T. Celikel, and A. J. Van Opstal, "Spectral weighting underlies perceived sound elevation," *Scientific reports*, vol. 9, no. 1, p. 1642, 2019.

[15] S. Spagnol, M. Geronazzo, and F. Avanzini, "Fitting pinna-related transfer functions to anthropometry for binaural sound rendering," in *2010 IEEE International Workshop on Multimedia Signal Processing*. IEEE, 2010, pp. 194–199.

[16] K. H. Shin and Y. Park, "Enhanced vertical perception through head-related impulse response customization based on pinna response tuning in the median plane," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 91, no. 1, pp. 345–356, 2008.

[17] I. T. Jolliffe, *Principal component analysis*. Springer Verlag, 2002.

[18] S. Xu, Z. Li, and G. Salvendy, "Identification of anthropometric measurements for individualization of head-related transfer functions," *Acta Acustica united with Acustica*, vol. 95, no. 1, pp. 168–177, 2009.

[19] W. L. Martens, "Principal components analysis and resynthesis of spectral cues to perceived direction," in *Proc. Int. Computer Music Conf., Champaine-Urbana, IL*, 1987.

[20] J. Leung and S. Carlile, "Pca compression of hrtfs and localization performance," in *International Workshop on the Principies and Applications of Spatial Hearing*, 2009.

[21] P. Mokhtari, H. Kato, H. Takemoto, R. Nishimura, S. Enomoto, S. Adachi, and T. Kitamura, "Further observations on a principal components analysis of head-related transfer functions," *Scientific reports*, vol. 9, no. 1, p. 7477, 2019.

[22] J. C. Middlebrooks and D. M. Green, "Observations on a principal components analysis of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 597–599, 1992.

[23] Z. Wu, F. H. Chan, F. Lam, and J. C. Chan, "A time domain binaural model based on spatial feature extraction for the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2211–2218, 1997.

[24] D. W. Grantham, J. A. Willhite, K. D. Frampton, and D. H. Ashmead, "Reduced order modeling of head related impulse responses for virtual acoustic displays," *The Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3116–3125, 2005.

[25] K. J. Fink and L. Ray, "Tuning principal component weights to individualize hrtfs," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 389–392.

[26] S. Hwang, Y. Park, and Y.-s. Park, "Customization of Spatially Continuous Head-Related Impulse Responses in the Median Plane," *Acta Acustica united with Acustica*, vol. 96, no. 2, pp. 351–363, Mar. 2010.

[27] M. Rothbucher, M. Durkovic, H. Shen, and K. Diepold, "Hrtf customization using multiway array analysis," in *2010 18th European Signal Processing Conference*. IEEE, 2010, pp. 229–233.

[28] F. Wightman and D. Kistler, "Localization of virtual sound sources synthesized from model hrtfs," in *Final Program and Paper Summaries 1991 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 1991, pp. 0_51–0_52.

[29] B.-S. Xie, "Recovery of individual head-related transfer functions from a small set of measurements," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 282–294, 2012.

[30] S. Takane, "Effect of domain selection for compact representation of spatial variation of head-related transfer function in all directions based on spatial principal components analysis," *Applied Acoustics*, vol. 101, pp. 64–77, 2016.

[31] Z. Liang, B. Xie, and X. Zhong, "Comparison of principal components analysis on linear and logarithmic magnitude of head-related transfer functions," in *2009 2nd International Congress on Image and Signal Processing*. IEEE, 2009, pp. 1–5.

[32] W. G. Gardner and K. D. Martin, "Hrtf measurements of a kemar," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[33] G. Marentakis and J. Hölzl, "Compression efficiency and signal distortion of common pca bases for hrtf modelling," in *Proceedings of the 18th Sound and Music Computing Conference (SMC)*, 2021, pp. 60–67.

[34] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger *et al.*, "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions," in *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.

[35] O. Warusfel. (2003) Listen hrtf database. [Online]. Available: http://recherche.ircam.fr/equipes/salles/listen

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

[37] T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolating head related transfer functions in the median plane," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452)*. IEEE, 1999, pp. 167–170.