Compression efficiency and Signal Distortion of common PCA bases for HRTF Modelling

Georgios Marentakis (Γεώργιος Μαρεντάκης)[1] and **Josef Hölzl**[2]

[1]*Faculty of Computer Sciences*, **Østfold University College**, Halden, Norway
[2]*Institute of Electronic Music and Acoustics*, **University of Music and Performing Arts Graz**, Graz, Austria

ABSTRACT

Principal Component Analysis (PCA) has been often used for HRTF compression and individualization. However, there is significant variation in how the input matrix on which PCA is applied is constructed. Here, we study the effect of choices on the selection of independent variables, the domain in which impulse responses are represented, the HRTF database used, and possible smoothing on the compression efficiency and the reconstruction quality of the resulting PCA model. Several findings replicate well across different databases. Results point to a benefit for signal compared to space PCA and for using minimum-phase HRIRs or HRTFs. Smoothing HRTFs leads to an increase in compression efficiency and a reduction in spectral distortion and using HRTFs with logarithmic magnitude leads to lower spectral distortion compared to linear.

1. INTRODUCTION

Head Related Transfer Functions (HRTFs) allow designers and engineers to create 3D audio using headphones [1] with applications in virtual and augmented reality. HRTF models that support individualisation, compact representation, and transfer are important as HRTFs are relatively long filters that are specific to individual users and need to be measured for all positions of interest in a relatively resource-intensive process [2].

A compact HRTF model can be reached by decomposing an HRTF set upon a set of orthogonal basis functions and obtaining the related weights (or loadings). Such decompositions can be used to reduce the, typically high, dimensionality of HRTF sets and serve as a basis for compression, individualization, and the investigation of their numerical and perceptual properties. Most often, Principal Component Analysis (PCA) e.g., [3–6] and the Spherical Harmonic Transform e.g., [7–9] have been used for this purpose. More recent approaches focus on deep learning [10].

This article focuses on using PCA for HRTF modelling. It is motivated by the fact that HRTF functions have been arranged in markedly different ways for PCA processing in the literature and aims to investigate the extent to which such differences may affect the compression efficiency and representation ability of the obtained model.

Copyright: © 2021 the Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

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 measured at or inside the ear-canal. 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 [2], most commonly, on a dense sampling grid. Frequently, HRTFs are diffuse-field equalized to exclude the ear canal resonance and measurement system response and come in the form of *directional transfer functions* (DTFs).

Binaural cues encoded in the *interaural transfer function*, $ITF = H_L(\theta, \phi, f)/H_R(\theta, \phi, f)$ help localize sounds in the horizontal plane. Monaural cues in the magnitude HRTF spectrum are used for elevation perception and front/back and up/down discrimination [11, 12]. These are spectral peaks and notches between 4 and 16 kHz that are mainly effected by the shape of the outer ear. For example, a prominent 1-octave notch centered between 6 and 11 kHz changes systematically with the vertical source location [13].

Whereas HRTFs incorporate the effects of the whole body, *pinna-related transfer functions (PRTFs)* indicate only the contribution of the pinna and reduce the dependence with respect to azimuth. 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 [14] and then transformed into frequency domain. 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 [15].

2.2 Principal Component Analysis

Principal Component Analysis is normally applied onto 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 [16]. 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 $\mathbf{S} = diag(s_1, \ldots, s_n)$, 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.

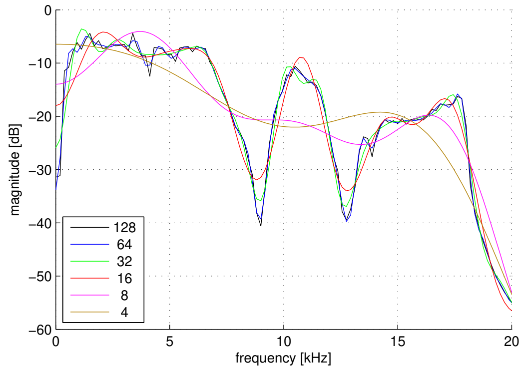


Figure 1. Spectral differences between unprocessed (128 coefficients) and smoothed DTF magnitude spectrum (with 64, 32, 16, 8 and 4 coefficients) in ARI database.

2.3 Modelling HRTFs using PCA

Head-related transfer function sets are mutli-dimensional and typically include the 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, 14, 17].

Most of the studies use enough components so as 90% of the variance in the data is explained [3, 15, 18]. 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, 19].

Structuring the PCA input matrix: Studies in the literature differ in the HRTF set used, the domain the signal is represented in, and in the way they are transformed into a 2D matrix for PCA.

Some studies apply PCA on HRIRs [5, 15, 20–23]. This is appealing as the time-domain signals maintain delay and phase information and can easily be windowed to isolate the effects of pinna, head, or shoulder. Other studies use minimum-phase HRIRs [5, 15, 23] which do not include direction-dependent delays. Several other studies use HRTF magnitude [3, 4, 17, 24–27] when forming the PCA input matrix and minimum-phase is used when transforming in the time domain. Minimum-phase is used based on the assumption that the original signal phase can be discarded and replaced by a direction-dependent delay.

In the case of HRTFs, PCA has been applied to both linear [4, 28] and logarithmic magnitude spectrum [3, 17, 18, 25, 26, 29] A further difference originates in the number of points that are used in the DTF to estimate the frequency spectrum. More recent studies use the complex spectrum as input to PCA [30, 31].

Most commonly, signal amplitude (time-domain representations) or spectral magnitude (frequency-domain representations) are used as variables in columns and subjects and directions of incidence as observations (rows). Recently, an alternative model has been proposed [6] which uses spatial directions as variables in columns and signal amplitude or spectral magnitude for each subject in rows.

The way the signals from the left and right ears enter the PCA input matrix has also been treated in different ways in literature. Sometimes only one ear is modelled and the second one is considered to be symmetric and therefore duplicated by the modelled one [5, 15]. Alternatively, it can be attempted to use PCA to explain the variability across the two ears. This can be done either by using the time/frequency signals from the second ear as observations in rows [3, 29] in the PCA input matrix.

PCA has not been always performed on the complete set of sound directions in the dataset. Decomposition has been applied on the whole database [3,25], smaller subsets, such as the median [5, 23] or horizontal plane [22, 32], and on single sound directions [15]. The latter approach yields a different set of principal components for each direction, which may not be optimal from a compression perspective. However, as the variability due to direction is not present such an approach allows to focus on individual differences caused by subjects' anthropometry for smaller sets.

A final difference is the HRTF database used to perform the analysis. In general, principal components obtained from different HTRF datasets are consistent as long as the number of measurement directions and subjects is reasonable. This invariance is more evident for components explaining a large amount of variance, as components of smaller variance reflect specificities that might not be shared across datasets. Middlebrooks and Green [26] were among the first who compared basis vectors calculated from their own measurement data (8 subjects, 360 positions) with an existing database by Kistler and Wightman [3] (10 subjects, 265 positions) and indeed confirmed a high correlation between the components, which however decreased with rising principal component order number.

2.4 Summary and Research Questions

The literature review shows that the differences in constructing the PCA input matrix relate to the domain used (time or frequency), the representation (linear or logarithmic magnitude spectrum), the use of minimum-phase HRIRs, the handling of the two ears, and the number of directions analyzed. Given the complexity of modelling HRTFs, it is reasonable to ask what is the impact of choices for the aforementioned parameters on the compression efficiency and the reconstruction potential of the obtained PCA basis. Quite reasonably, researchers would favor an alignment that can represent and re-synthesize the HRTF dataset with the lowest possible number of components and smallest distortion.

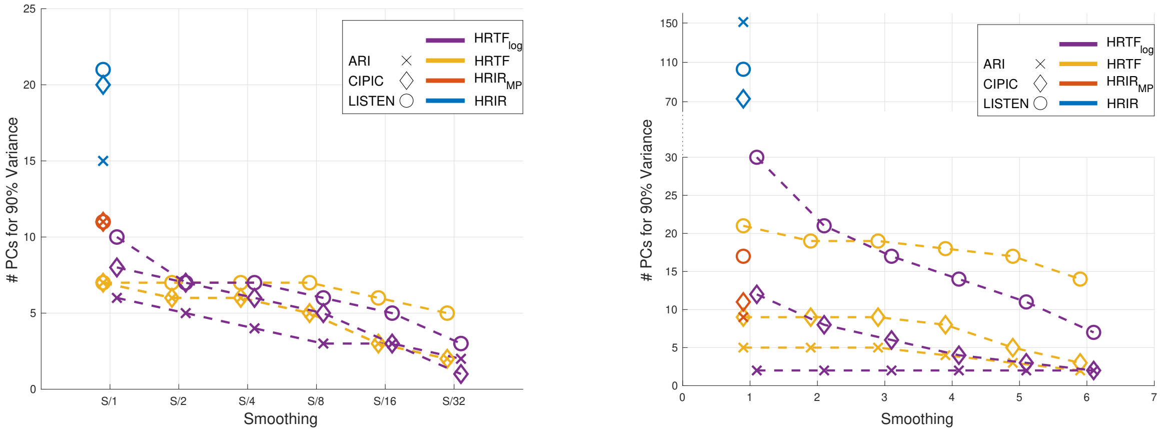
Despite the obvious benefit in identifying an optimal PCA basis, few studies have attempted a direct comparison. Leung and Carlile [19] investigated the PCA compression efficiency and came to the conclusion that the optimal format for PCA decomposition in terms of compression is the linear amplitude form in frequency domain. They used an HRTF dataset of 393 directions. They found that 5 PCs are required for explaining 90% variance when linear magnitude is used; there were less than the number required with logarithmic magnitude. However, the number of subjects or the structure of the PCA input matrix is not clearly described. Takane *et al* [33] extend the work of Liang *et al* [34] and compare 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 input structures are evaluated based on explained variance, signal distortion, and signal-to-distortion ratio using the KEMAR HATS database [35]. The results confirm an advantage in using representations in the frequency domain but are somewhat inconclusive otherwise. The HRIR database used in this study is this of a dummy head and does not include several or real subjects. Furthermore, the structure of the input matrix structure is not varied to include spatial PCA. Another parameter that has not been considered is the extend to which HRTFs were smoothed. It has been shown that mild spectral smoothing does not affect localization accuracy after reconstruction [36]. Smoothing may have a positive effect on the compression efficiency as perceptually-irrelevant details of HRTF magnitude are smoothed out [37, 38]. For this reason, the smoothing factor is worth including as a parameter in simulations. Overall, a more systematic investigation on the impact of setting up the PCA input matrix on compression efficiency and reconstruction accuracy is attempted here.

3. NUMERICAL EVALUATION

The parameters varied in the evaluation were: HRTF database, the structure of the input matrix, the domain in which the signal was represented, and extent to which HRTFs have been smoothed as explained below. 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:

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

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 [39], 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. When HRIRs were used, original and reconstructed HRIRs were



(a) Signal PCA

(b) Spatial PCA

Figure 2. The number of PCs required to explain 90% of variance for signal and spatial PCA in the examined cases. HRIRs were not smoothed and are presented as single points up left on each plot.

transformed in the frequency domain in order to estimate the spectral distortion. The calculation extended over the entire frequency range. Simulations were performed in MATLAB®.

3.1 HRTF Databases

Three open access HRTF databases were used: the Acoustics Research Institute (ARI) HRTF database [40], the LISTEN database from the Institut de Recherche et Coordination Acoustique/Musique [41] and the HRTF database from the University of California at Davis (CIPIC) [42]. 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. Two subjects were excluded from calculations because impulse responses were not measured for all sound directions. In addition, subject ID 1034 in the LISTEN database resulted in outlying weights and was excluded from the dataset.

3.2 PCA Input Matrix Structure

The first structure (*Signal PCA*) follows a common pattern that has been also used by Kistler and Wightman [3, 29]. Here, signal bins (in frequency or time domain) are the independent variables in columns, while replications for the different subjects and measurement directions are observations in rows. This leads to an input matrix with (subjects × sound directions) rows and (signal bins) columns. The number of rows is doubled if both ears are included and the resulting principal component weights (PCWs) for each subject can be used to recreate the HRIR or HRTF for both ears of each subject and for all directions.

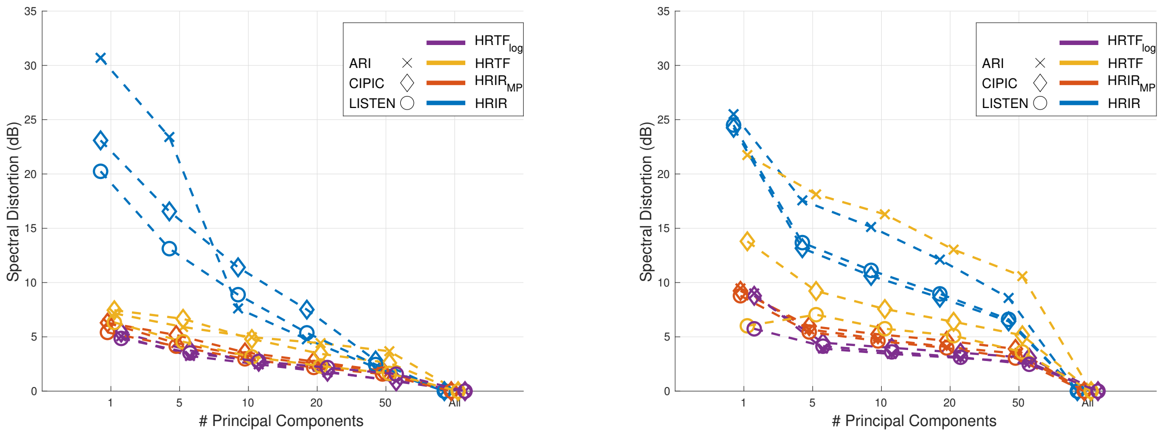
The second structure (*Spatial PCA*) has (subjects × signal bins) rows and (sound directions) columns and was also used by Xie [6, 25]. It lists each sound direction as independent variable in the matrix columns while frequency or time samples from the head-related functions of all subjects are placed as observations in rows. The number of rows is doubled if both ears are included. It has been called spatial PCA because analyzed directions are independent variables placed in columns. The resulting weights can be used to recreate each frequency or time bin for a given position, ear, and subject. In the simulations, HRTFs from both ears were entered as observations in rows for both input structures.

3.3 Signal Domain

For each of the input structures, four signal representations were tested. The first two were in the time domain: the HRIR and the minimum-phase HRIR. The minimum-phase HRIR was used because it allows to remove the direction-dependent initial delay and phase and may reduce the number of components required to represent the signal. Direction-dependent delay can be added after reconstruction in case such a representation is used for compression or individualization. The latter two were in the frequency domain: the magnitude spectrum in either linear or logarithmic amplitude, a common difference in studies applying PCA to HRTFs.

3.4 Smoothing

Smoothing was done by taking the logarithm of the HRTF spectrum, performing FFT, and limiting the number of the Fourier coefficients used to recreate the spectrum, a lowpass filtering operation. Even as few as 16 coefficients within a spectrum of 512 coefficients, a smoothing factor of 1/32 for IR length of 1024 samples, were found to yield satisfactory localization [36]. An example of the output of the smoothing process is shown in Figure 1. Smoothing was only applied when constructing PCA bases using HRTFs and not when using HRIRs.



(a) Signal PCA

(b) Spatial PCA

Figure 3. Spectral Distorion upon reconstruction averaged among subjects, ears, and directions for unsmoothed head related functions in the conditions examined in the simulations.

4. RESULTS

The results of the simulation ar epresented next starting from the number of components required to explain 90% of the input matrix variance and following up with the spectral distortion results.

4.1 Compression Efficiency

By observing Figure 2, it can be seen that the input matrix structure has a considerable impact on compression efficiency. For the CIPIC and LISTEN databases, most efficient is signal PCA followed by spatial PCA. This is consistent across smoothing factors. For the ARI database, spatial PCA using either linear or log HRTF magnitude yields a compression efficiency that is higher than signal PCA.

Quite clearly, the HRIR representation requires most components, irrespective of whether signal or spatial PCA is performed. Taking the minimum-phase HRIR results in a significant reduction in the number of PCs required to explain 90% of the variance which makes using minimum-phase HRIRs comparable to PCA using unsmoothed HRTFs in terms of compression efficiency. This result is consistent across the databases examined here.

The number of PCs required by frequency domain representations is reduced significantly due to the application of spectral smoothing. Each time the Fourier coefficients used in spectral reconstruction are halved, a significant reduction in the number of components required to explain 90% variance is observed.

The impact of a linear or logarithmic magnitude representation in the frequency domain is not as clear-cut. For unsmoothed HRTFs, a small advantage for linear amplitude representation is registered for the LISTEN and CIPIC database for both signal and space PCA. As long as smoothing is applied, the situation is reversed and the logarithmic representation results in a smaller number of required components. For the ARI database, a small advantage for logarithmic magnitude representation appears which remains consistent as smoothing is applied.

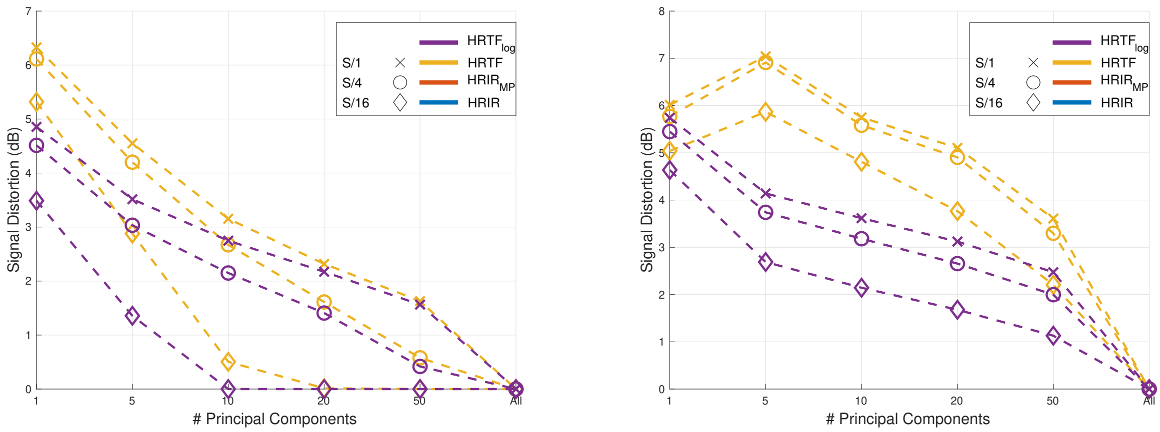
The number of components required to explain 90% of the variance is consistent across databases for the signal PCA. However, it varies considerably when spatial PCA is considered and the number of required components is doubled for ARI to CIPIC and then the LISTEN database.

4.2 Spectral Distortion

By observing Figure 3, it can be seen that spectral distortion results are in agreement with the compression efficiency observations. Spectral distortion was highest when HRIRs were used in the input matrix. Spectral distortion was reduced significantly when minimum-phase HRIRs or HRTFs were used. This result is consistent across databases.

For signal PCA, signal distortion is lowest when minimum-phase HRIRs and the logarithmic HRTFs are used in the PCA input matrix and falls below 5 dB as soon as 5 components are used for reconstruction. Signal PCA with with linear amplitude HRTFs result in an overall higher spectral distortion. This result is consistent across databases.

For spatial PCA, again PCA with minimum-phase HRIRs or logarithmic magnitude HRTFs result in the lowest spectral distortion which again falls below 5 dB as long as at least 5 components are used for reconstruction. Spectral distortion is highest for the ARI database for the linear magnitude HRTFs and the HRIRs compared to the rest but the differences among databases were smaller for minimum-phase HRIRs and log-magnitude HRTFs. Interestingly, achieving a spectral distortion below 5 dB requires a higher number of components than the one required to explain 90% of the variance in the ARI database for the spatial PCA. By observing Figure 4, it can be seen that smoothing does seem to reduce spectral distortion. This effect was consistent across databases input structures and appeared both for linear and log-magnitude HRTFs.



(a) Signal PCA

(b) Spatial PCA

Figure 4. Spectral Distorion upon reconstruction averaged across subjects, ears, and sound directions for smoothed head related functions taken from the LISTEN database.

5. DISCUSSION

In this study, we investigated the impact of choices in the design of the input matrix used to analyze head related impulse responses or head related transfer functions using principal component analysis. The impact of matrix structure (signal or spatial PCA) and the signal domain (time or frequency) was manipulated. For the time-domain HRIRs both raw and minimum-phase HRIRs were compared, while for the frequency domain linear and logarithmic amplitude was compared. Finally, for frequency domain representations, the impact of spectral smoothing was also considered. Three different HRTF databases were used in the analysis. The number of components required to explain 90% of the variance and the spectral distortion upon reconstruction were used as objective measures for the purpose of comparison.

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. The variable to observation ratio for spatial PCA was 0.27 for ARI, 0.15 for CIPIC, and 0.015 for LISTEN, while for signal PCA it was 0.013 for LISTEN 0.0009 for CIPIC, and 0.0005 for ARI. The signal PCA configurations have a better variable to observation ratio which may explain the better consistency of the results across databases for signal compared to spatial PCA.

For the LISTEN and CIPIC databases, fewer components were required to represent 90% of the input matrix variance and the resulting spectral distortion upon reconstruction was lower for signal PCA in comparison to spatial PCA. However, a lower number of components was required to account for 90% of the variance when analyzing the ARI database using spatial compared to signal PCA required, which was even lower when the log-magnitude spectrum was used. The number of components required for spatial PCA in the ARI database was consistently small even when the number of database locations used was reduced and the variable to observation ratio improved. However, the suggested number of components yielded increased spatial distortion upon reconstruction and would need to be increased to keep spectral distortion below 5 dB. Further incestigation is required to confirm if Spatial PCA can lead to an effective PCA basis and to explain the discrepancy among databases.

Overall, the raw HRIR representation was the most inefficient both in terms of compression efficiency and in terms of the resulting Signal Distortion upon reconstruction, in agreement with observations in the literature [33]. Removing the direction dependent delay and phase from the signals by taking the minimum-phase impulse response reduced the number of components and the spectral distortion upon reconstruction dramatically and made principal component analysis as efficient as with input matrices using spectral HRTFs.

A beneficial effect of smoothing for PCA using HRTFs was observed which improved compression efficiency and reduced spectral distortion. As smoothed HRTFs have been found to provide good sound localization, this may be a good option to consider in future applications of principal component analysis up to the point where localization is not affected and coloration does not appear [36]. It would be interesting to examine if a similar result would have been observed if the HRIRs were smoothed using a low pass filter in the time-domain but this was not investigated here.

Concerning the impact of a linear or logarithmic representation for PCA analysis based on spectral HRTF data the results are not as clear-cut. On the one hand, there is a tendency for lower number of components for representing 90% of the variance for the linear amplitude, as also mentioned by [19, 33, 38] but this advantage tends to be cancelled as long as smoothing is applied. Furthermore, the logarithmic representation leads to a lower spectral distortion. It appears therefore that the logarithmic representation may be more efficient if PCA is to be performed on HRTFs and both criteria are considered.

6. CONCLUSION

We presented a study that investigated the impact of HRTF database, input structure (signal or space), signal domain (time or frequency), and HRTF smoothing on the compression efficiency and the spectral distortion upon reconstruction when modelling HRTFs using Principal Component Analysis. The results of the numerical simulations show that signal PCA has a better compression efficiency (2/3 databases) and lower spectral distortion (3/3 databases) upon reconstruction. Furthermore, using HRIRs as input to PCA leads to worse compression efficiency and higher spectral distortion compared to HRTFs. Using minimum-phase HRIRs compensates for this discrepancy. Minimum-phase HRIRs lead to comparable compression efficiency and spectral distortion compared to HRTFs. Applying smoothing to HRTFs leads to an increase in compression efficiency and a reduction in spectral distortion for all databases used. Logarithmic magnitude leads to lowest spectral distortion when using HRTFs while compression efficiency depends on the database used.

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] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

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

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

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] G. Grindlay and M. A. O. Vasilescu, "A multilinear (tensor) framework for hrtf analysis and synthesis," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 1. IEEE, 2007, pp. I–161.

[25] 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.

[26] 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.

[27] J. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 439–452, 1995.

[28] 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.

[29] 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.

[30] O. A. Ramos and F. C. Tommasini, "Magnitude modelling of hrtf using principal component analysis applied to complex values," *Archives of Acoustics*, vol. 39, 2014.

[31] R. Bomhardt, H. Braren, and J. Fels, "Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions," in *Proceedings of Meetings on Acoustics 172ASA*, vol. 29, no. 1. Acoustical Society of America, 2016, p. 050007.

[32] K. J. Fink and L. Ray, "Individualization of head related transfer functions using principal component analysis," *Applied Acoustics*, vol. 87, pp. 162–173, 2015.

[33] 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.

[34] 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.

[35] 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.

[36] A. Kulkarni and H. S. Colburn, "Role of spectral detail in sound-source localization," *Nature*, vol. 396, no. 6713, pp. 747–749, 1998.

[37] J. Breebaart, F. Nater, and A. Kohlrausch, "Spectral and spatial parameter resolution requirements for parametric, filter-bank-based hrtf processing," *Journal of the Audio Engineering Society*, vol. 58, no. 3, pp. 126–140, 2010.

[38] J. Breebaart, "Effect of perceptually irrelevant variance in head-related transfer functions on principal component analysis," *The Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. EL1–EL6, 2013.

[39] 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.

[40] 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.

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

[42] 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.