Visualization and principal component analysis from a head-related transfer function database

Alex Tung; *atung323@gmail.com*
Department of Electrical and Computer Engineering, University of California, San Diego

ABSTRACT

Binaural localization can be characterized by head-related transfer functions (HRTFs), which are stored as collections of filters that correspond to discrete angles of incidence. Of the efforts to characterize and compress these datasets, a common method used throughout literature is principal component analysis (PCA). In this work, PCA is conducted on the CIPIC HRTF database, accompanied with a brief discussion on the efficacy of projecting filter sets onto a reduced subspace.

INTRODUCTION

Humans' ability to localize acoustic stimuli arise from the interactions between the stimulus, acoustic environment, and the body, and an a priori knowledge of common stimuli. Of these interactions, significant effort has been spent characterizing acoustic interactions with the body, namely the diffraction and absorption of acoustic wavefronts around the head, torso, and ear pinnae. The difference between this baffled propagation and free-field propagation is known as a head-related transfer function (HRTF hereafter). These transfer functions vary with respect to the incident direction of the sound source per individual, and serve as binaural cues for humans to discern the direction from which a sound originates.

An individual set of HRTFs is directly measured by placing microphones in each ear canal of a subject (known as a blocked-meatus measurement). A broadband stimulus is played from a loudspeaker placed at a known angle of incidence toward the subject, and a frequency response is collected at the microphones. The measurement is repeated at a range of angles by rotating either the subject or the loudspeaker. This resulting set of head-related impulse responses (HRIRs[i]) is then transformed into a set of HRTFs in frequency domain with the discrete Fourier Transform (DFT).

Measuring HRTFs is useful in that a recorded acoustic stimulus can be perceived to originate from a specific direction using the filter associated with that direction – either by convolving the stimulus with the time domain HRIR or by multiplying the DFT-transformed stimulus with the HRTF and applying the inverse DFT. This ability for sound sources to be perceptibly externalized is known as a virtual auditory display, and is especially useful for synthesizing realistic acoustic environments for simulation applications, such as virtual reality and surround-sound entertainment through headphones.

The conventional method of collecting HRTF data, however, is costly and cumbersome, as it requires access to an environment with a reasonable level of diffuse-field acoustic absorption and a long period of time for separate impulse responses to be collected for stimuli from different directions. Choosing to query fewer angles may reduce time at the cost of omitting important spectral features that are otherwise hard to interpolate.

In addition, the collected filters result in a large dataset. Several efforts have been made to simplify and reduce different aspects of HRTF characterization and storage. For individual filters, it has been shown that an HRIR can be characterized as a linear filter and a constant interaural time difference, then implemented as a cascade of a minimum-phase FIR filter and an all-pass delay[1]; the process of storing each HRIR as such has become convention. There have also been efforts to derive generic HRTFs and resynthesize individual filter sets using anthropomorphic parameters correlated with different spectral features[2,3,4]. For the set as a whole, some investigations on correlations between filters in a set have been conducted to evaluate the variability of different databases[5,6].

Of the mentioned works, the technique of principal component analysis (PCA) is often used. This is a mathematical technique of reducing the dimensionality of a dataset by eigendecomposition and matrix projection.

[i] HRIRs and HRTFs correspond to the same filter in the frequency domain and time domain, respectively. Although there are different methods of deriving HRIRs by transforming responses from filtered broadband noise, multi-tone signals and maximum-length sequences, they will all be referred as HRIRs. Because the HRTFs and HRIRs are related by the DFT, the two terms will be used interchangeably when it is not crucial to distinguish between the two.

In this work, PCA will be performed and explained in the context of HRTFs similarly to Kistler et al.[ii], while using the CIPIC database, a popular public-domain set of measured HRTFs and anthropomorphic data from UC Davis[7].

DATABASE EXPLORATION

The CIPIC database contains HRIRs[iii] of length 200 samples and left- and right-ear onset times from different directions for each subject. A range of azimuth angles $\theta \in [-80, +80]$ and elevation angles $\phi \in [-45, +230.625]$ is non-uniformly covered[iv]. Although a graphical interface is provided from UC Davis to visualize and explore the data, it does not return the filters needed for the statistical analysis in this work, so custom scripts were written to visualize and manipulate the filters. The interaural time difference (ITD) and the total interaural intensity difference (IID) are calculated as follows:

$$T_{ITD,k} = T_{OnR,k} - T_{OnL,k}, \quad k \in [1{,}1250],$$

$$I_{IID,k} = \sum_{t=1}^{200} h_k(t)^2$$

In the equations, $h_k(t)$ is the measured HRIR and $T_{OnL,k}$, $T_{OnR,k}$ are the left- and right-ear onset times at the k^{th} direction. For the angles queried in this dataset, each subject contains $25 * 50 = 1250$ HRIRs. Because the HRIRs are real-valued, negative frequencies are truncated after transforming to HRTFs with a 256-point DFT:

$$HRIR_k \xrightarrow{DFT,256} HRTF_k; \qquad HRTF_k \in R^{128}$$

These quantities for a sample subject are visualized in Figure 1. From this, it is apparent that the ipsilateral sound sources are the highest in intensity across azimuths. It can also be seen that the ITD increases from left to right and varies minimally with respect to elevation.

Figure 2a shows the frontal HRTF ($\theta = \phi = 0$) for the same subject. For the rest of this work relating to PCA, the focus will be on the DFT-transformed HRTFs, which do not include the ITD constants.

PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is the process of transforming a matrix of correlated features into a linear combination of orthogonal basis vectors, or principal components. This technique can employ matrix

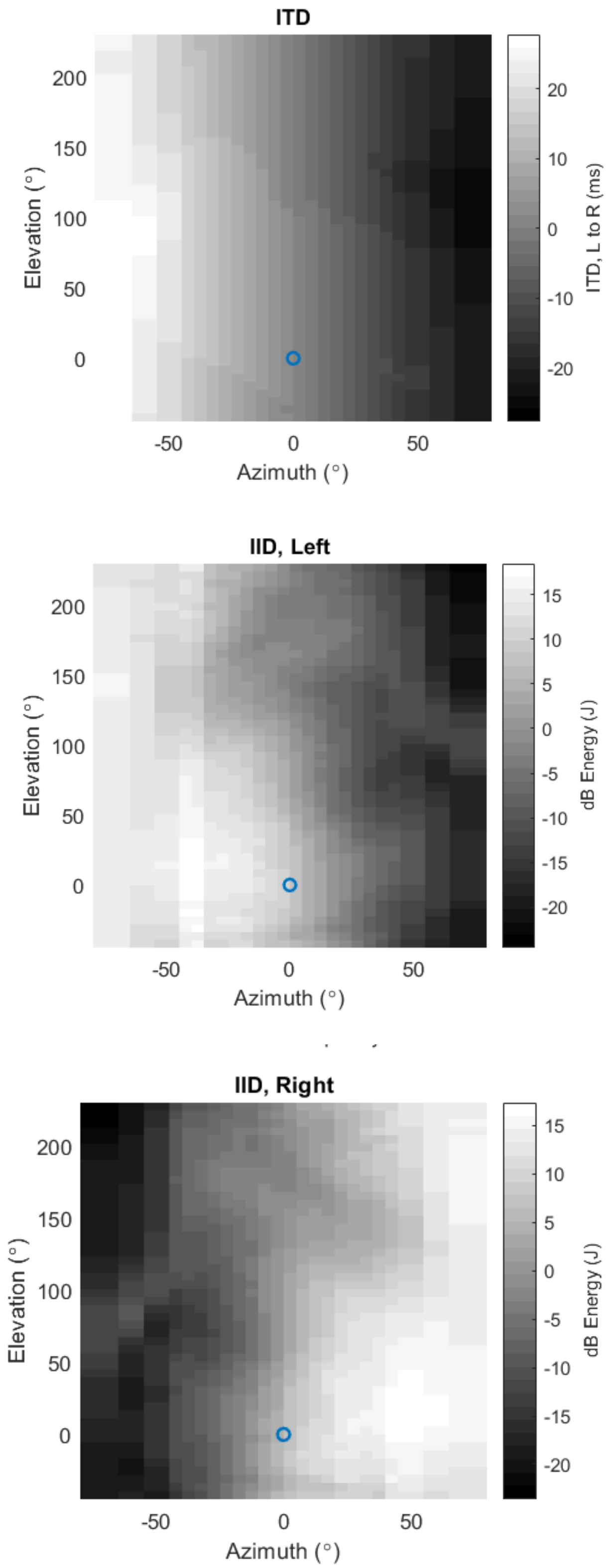

Fig. 1a (top): Sample ITD with respect to azimuth and elevation.
Fig. 1b (middle): Sample Left IID
Fig. 1c (bottom): Sample Right IID. The circular marker represents the frontal HRTF, where the ITD is close to 0.

[ii] Kistler performed the same PCA on a HRTF database – although the operation was restricted to a much narrower band of frequencies to decrease dataset variability and achieve greater set reduction.

[iii] Sampled at F_S = 44.1 kHz
[iv] Azimuth $\theta = \{-80, -65, -55, -45{:}\,5{:}\,45, 55, 65, 80\}$, Elevation $\phi = \{-45{:}\,5.625{:}\,45\}$

eigendecomposition, where for a square matrix $A \in R^{NxN}$, there exists the vector $v \in R^N$ scalar $\lambda \in R$ such that

$$Av = \lambda v$$

These vectors and their corresponding scalars are respectively defined as eigenvectors and eigenvalues, and the set can be wholly expressed as the following:

$$AV = V\Lambda$$

$$V = [v_1, v_2, \dots, v_N], \qquad \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_N \end{bmatrix}$$

Although PCA is often used to make predictive models from a matrix of observations, it can also be used to reduce the dimensionality of a large dataset by expressing it in terms of its "strong" components, highly correlated to the highest sources of variability in the set.

To ensure that PCA operates on the most variable parts of the dataset, several means are subtracted[v] from the left and right HRTFs to obtain directional transfer functions (DTF and DTF2)[vi]:

$$DTF_{L,k} = HRTF_{L,k} - \mu_L \qquad DTF2_{L,k} = DTF_{L,k} - \mu_G$$

$$DTF_{R,k} = HRTF_{R,k} - \mu_R \qquad DTF2_{R,k} = DTF_{R,k} - \mu_G$$

Where μ_L and μ_R are the means of each HRTF across all angles (unique per-subject, per-ear), and μ_G is the grand mean of DTFs across all angles, all subjects, and both ears. Figure 2 depicts the process of subtracting the means from the left side of a sample HRTF.

The frequency covariance matrix is then calculated for each location, across all b = 45 subjects, from which the mean covariance matrix was calculated across all locations. Figure 3 shows the covariance matrix and normalized correlation matrix (for ease of visualization).

$$\Sigma_k = \left[DTF2_{k,1}, DTF2_{k,2}, \dots, DTF2_{k,b}\right]^T, \quad b = 45$$

$$\mathbf{\Sigma} = \frac{1}{1250} * \sum_{k=1}^{1250} \Sigma_k, \quad \mathbf{\Sigma} \in R^{128 \times 128}$$

The eigenvalue and eigenvector matrices P and σ are then computed for $\mathbf{\Sigma}$ (in this case, 128 exist):

$$\mathbf{\Sigma} P = P\sigma, P = [p_1, p_2, \dots, p_{128}]$$

The matrix P is a linear transformation that maps the discrete frequency space (R^{128}) into a space of weights corresponding to the 128 orthogonal eigenvectors. Because the eigenvalues along the diagonal of σ is sorted in

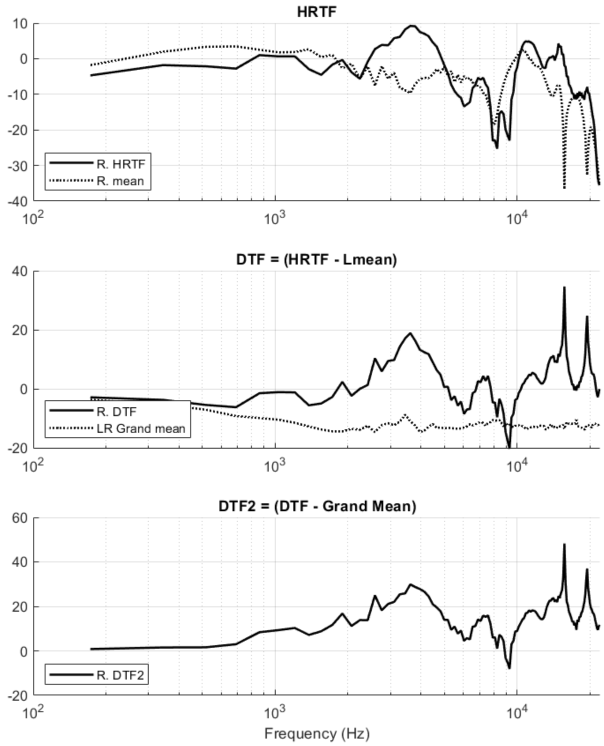

Figure 2a (top): Sample HRTF and subject-ear mean
Figure 2b (middle): Sample DTF and grand mean
Figure 2c (bottom): Sample DTF2

ascending order when computing in MATLAB, the corresponding eigenvectors in P are also sorted in ascending order of "contribution" to the frequency covariance of the dataset.

For a reduction to r dimensions, retaining the r trailing eigenvectors and omitting all other values yields a reduced projection matrix P_{red} that transforms each DTF2 into a lower dimensional subset of weights $w \in R^r \in R^{128}$. It is these weights that are stored in place of the original HRTFs.

$$P_{red} = [p_{128-r+1}, p_{128-r+2}, \dots, p_{127}, p_{128}] \in R^{128 \times r}$$

To resynthesize the 128-length HRTF, the set of weights is left-multiplied by the projection matrix P_{red} and added to the grand mean and individual per-ear means:

$$HRTF_{red,L,k} = P_{red} * w_{L,k} + \mu_G + \mu_L$$

$$HRTF_{red,R,k} = P_{red} * w_{R,k} + \mu_G + \mu_R$$

Figure 4 shows the five strongest principal components, and Figure 5 shows a sample HRTF, resynthesized with different numbers of principal components (bases). By inspection, the complex peaks and valleys of the sample HRTF are sufficiently synthesized for $r > 30$. This is corroborated in Figure 6 by the mean square

[v] HRTFs are expressed in a logarithmic decibel scale and subtracted. This corresponds to division on a linear scale.

[vi] DTF was originally coined by D. Kistler's paper.

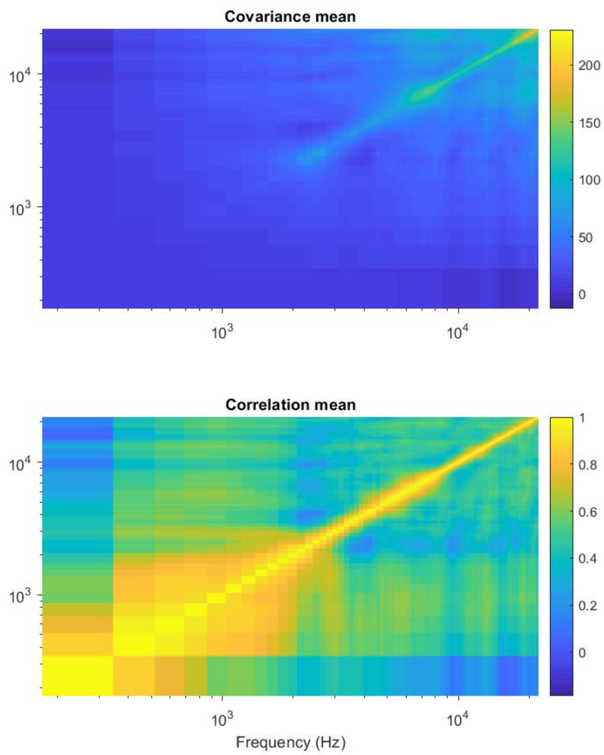

Figure 3a (top): Covariance matrix values
Figure 3b (bottom): Correlation matrix values

error and L2 norms calculated between the original sample HRTF and its approximations: both loss functions drop below 20dB when reconstructing with more than 40 bases, ultimately converging to a zero-loss, perfect reconstruction condition at $r = 128$. We will posit that a reduced projection matrix of 30 bases is sufficient.

CONCLUSION

A broad-band PCA was conducted on an HRTF database by factoring out individual and grand means. For the case of the CIPIC database, 1250 HRIRs of 200 samples each were stored for each of 45 subjects, resulting in of 11,250,000 floating point numbers[vii]. If a reduced projection matrix of 30 components is used in place of the HRIRs and the various means are stored, the total size was reduced to 1,702,988 floating point numbers[viii], or a reduction of about 84.9%.

In future works, analyses can be conducted on the correlation between the existing principal components and subjective listening test results to further identify the features that are perceptually significant, rather than computing the PCA based solely on variance (e.g. omitting

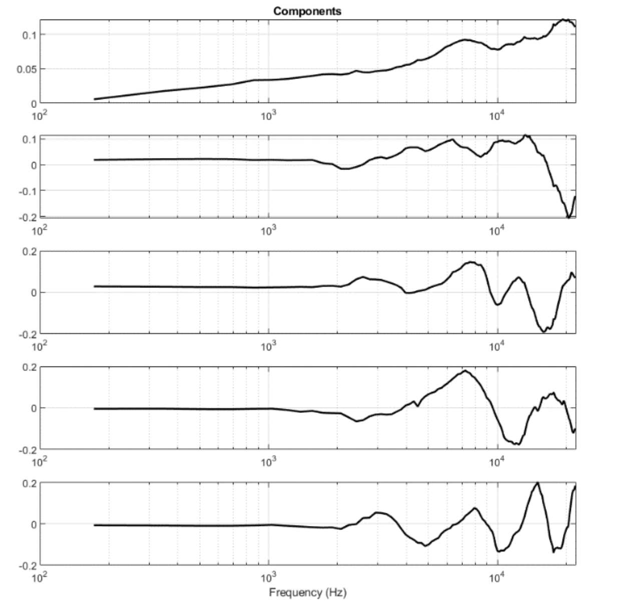

Figure 4: Strongest 5 principal components, in ascending order from top to bottom.

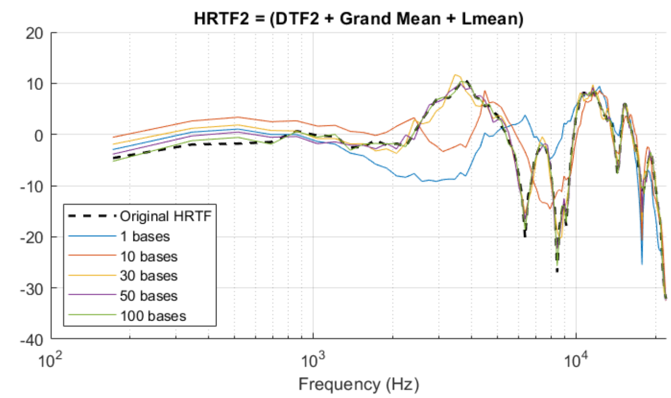

Figure 5: Reconstructed HRTFs compared to original

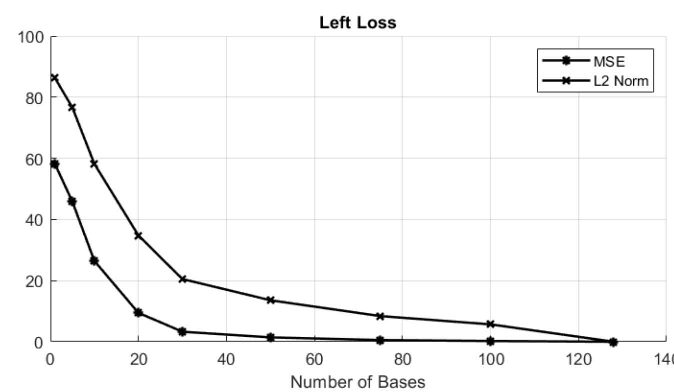

Figure 6: Mean square error and L2-norm (Euclidean) loss functions

[vii] (200 samples) * (1250 angles) * (45 subjects) = 11,250,000

[viii] ((30 samples * 1250 angles) + (128 μ_L samples) + (128 μ_R samples)) * 45 subjects + 128 μ_G samples + (128*30 P_{red} samples) = 1,702,988

frequencies in which differences are virtually imperceptible, as in Kistler's work). In addition, the loss functions can be quantitatively interpreted by comparing the error decibels to other real-world factors, such as signal-to-noise ratio and subjective results with non-individualized HRTFs. Reports on the implementation of this PCA reduction in a larger binaural playback system could also be conducted, where performance trade-offs between the memory accesses of smaller datasets and the increased processing of reconstructing HRTFs can be calculated and measured in real-time audio processing environments.

REFERENCES

[1] Kistler, D. J. and Wightman, F. L., "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, p. 1637 (1992).
[2] Nishino, T., Inoue, N., Takeda, K., et al., "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, pp. 897-908 (2007).
[3] Xie, B., Zhong, X., Dan, R., et al., "Head-related transfer function database and its analyses," *Science in China Series G: Physics, Mechanics & Astronomy*, vol. 50 (2007).
[4] Wang, L., Zeng, X., and Ma, X., "Advancement of Individualized Head-Related Transfer Functions (HRTFs) in Perceiving the Spatialization Cues: Case Study for an Integrated HRTF Individualization Method," *Appl. Sci.*, 2019.
[5] Andreopoulou, A., Roginska, A., and Bello, J., "Observing the Clustering Tendencies of Head Related Transfer Function Databases," *131st Audio Engineering Society Convention,* 2011.
[6] Andreopoulou, A., Roginska, A., and Mohanraj, H., "Analysis of of the Spectral Variations in Repeated Head-Related Transfer Function Measurements," *International Conference on Auditory Display*, 2013.
[7] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, "The CIPIC HRTF Database," *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24, 2001.