# REPRESENTATION OF HEAD RELATED TRANSFER FUNCTIONS WITH PRINCIPAL COMPONENT ANALYSIS

Jaka Sodnik, Anton Umek, Rudolf Susnik, Goran Bobojevic and Saso Tomazic

University of Ljubljana, Faculty of Electrical Engineering, Slovenia, Europe

## Abstract

Head Related Transfer Functions (HRTFs) describe the changes in the sound wave as it propagates from a spatial sound source to the human eardrum. One possible representation of HRTF data is the use of Principal Component Analysis (PCA), which decomposes data to principal components and corresponding weights. We applied PCA to MIT Media Lab non-individualized HRTF library. The linear amplitudes of elevation 0º were decomposed to four principal components and four weights per amplitude. The aim of our experiment was to find some mathematical regularity in the weights for different azimuths. It has been established that the variation of each weight can be approximated with a suitable mathematical function (sinus functions or polynomial). Such representation of weight variation enables the reconstruction of HRTF amplitudes for non-measured positions and reduces the amount of necessary data. Our model was tested on 25 subjects and proved to be very efficient.

## Introduction

The human brain together with the sense of hearing is capable of determining the position of an arbitrary sound source with a high precision. All the factors that influence the localization of the spatial sound source are contained in Head Related Transfer Functions (HRTF), which describe the changes in the sound wave as it propagates from the sound source to the human eardrum [3]. HRTF presented in time domain are called Head Related Impulse Responses (HRIR), which can be measured directly as impulse responses of the appointed system (the propagation path between the sound source and the human eardrum) [1],[5]. If the impulse response is time limited, its samples can be used as Finite Impulse Response (FIR) filter coefficients. Such FIR filters can be used to generate virtual sound sources played through the headphones. In order to create a virtual sound source in arbitrary spatial positions, impulse responses are needed for all the required spatial positions.

Several experiments have already been carried out in order to find a simpler and more effective way to create spatial sound for headphones. Some researchers tried to substitute FIR filters with Infinite Impulse Response (IIR) filters, which contain fewer coefficients [7]. Others tried to model and analyze HRTFs with the common-acoustical-pole-zero model and defined the direction dependent information in HRTF spectra [6]. One possible way of modeling HRTF spectra is the use of resonators which have to be amplified or attenuated properly [4], [15].

A very effective way of describing HRTFs with less data was proposed by Martens [12], who applied Principal Component Analysis (PCA) on HRTFs. PCA describes the original data set with only a few orthogonal components and corresponding weights. Martens modeled the functions for 35 spatial positions with four principal components and weights. He noted a systematic variation of the weights for azimuth changes from left to right and from front to rear.

Kistler and Wightman [9], on the other hand, applied PCA on 10 people's HRTFs. Each set with the data for 256 spatial positions was decomposed to five principal components and corresponding weights. In their second experiment, they used hearing tests to try and validate their model, reconstructed from a different number of principal components (1-5). The input data for their analysis consisted of the log-magnitudes of HRTFs.

The third experiment, where they applied PCA to HRTFs, was the search for some common properties in HRTF data sets of large numbers of people [8],[11].

All authors reported very effective data compression as an important result of their experiments.

Our research focused on the study of PCA weight variations. Our main goal was to find some regularity in those variations for different spatial positions. The variations were described with simple mathematical functions and reconstruction of original data was performed. We focused on the sets of HRTFs at fixed elevations at the front side of the head, noting only the variation of the weights for different azimuths.

The input data of our analysis were the linear amplitudes at elevation 0º. MIT Media Lab HRTF library was used in the experiment [5]. The second phase included the reconstruction of the data with our proposed model and its validation with several localization tests. The test subjects were asked to determine virtual sound source positions for spatial sounds played through headphones. The test environment was developed in Matlab programming language and it contained a special navigation panel which enabled the playback of spatial sound at a specific position by clicking on the selected coordinates.

## Head Related Impulse Response (HRIR) library

The HRIR library used in our experiment was the MIT Media Lab public library [5]. It includes the measurements for 710 spatial positions. The impulse responses were measured using the Kemar artificial head. Each impulse response consists of 128 samples and can be used as FIR filter to generate spatial sound. For each position, two filters are needed - for the left and right ear separately. In our experiment, we used measurements at the elevation 0º. At elevation 0º, 37 impulse responses are available for the front side of the head (azimuths from -90º to 90º).

The HRTF library was calculated from the HRIR library using the Fourier transform. The spectrum of each function consisted of 64 samples. The set of input data for PCA was presented with the matrix of 64x37 samples (for 37 different azimuths).

MIT Media Lab library, used in the experiment, consisted of non-personified or general HRTFs. The consequence was somewhat higher inaccuracy of elevation localization and strong front-back confusion [15]. As mentioned before, we used functions at fixed elevations and for the front side only.

## Principal Component Analysis (PCA)

PCA is, according to Calvo [2], "a useful technique for reducing the dimensionality of datasets for compression or recognition purposes". It is also, according to Shlens [14], "a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to simplified dynamics that often underlie it. "

Input data S can be interpreted as follows. Each row of S corresponds to all measurements of a particular type ($S_i$).

$$S = \begin{bmatrix} S_1 \\ \vdots \\ S_m \end{bmatrix} \qquad (1)$$

Each column of S corresponds to a set of measurements from a particular trial. Covariance matrix K from input data S can therefore be defined as:

$$K_S = \frac{1}{n-1} SS^T \qquad (2)$$

In order to find the directions of maximum variance, the eigenvalues of autocorrelation matrix and the corresponding eigenvectors are calculated. To extract most of the variation, linear transformation matrix P is prepared from biggest eigenvectors.

$$PC = \lfloor pc_1, pc_2, ..., pc_j \rfloor \qquad (3)$$

The elements of PC are the so-called principal components of S. The projection of the original data S into eigenspace is necessary to obtain the matrix of appropriate weights:

$$w_i = PC^T \cdot S_i \qquad (4)$$

In our experiment, the input matrix S consisted of the linear amplitudes of HRTF (64 x 37samples). Phase spectra were not included in the analysis, but were derived in the course of the reconstruction from the amplitude spectra using Hilbert transform [13]. Based on the least mean square error and subjective criteria when listening to reconstructed signals, we decided to use four largest eigenvalues or eigenvectors ($pc_1 - pc_4$). That means that all amplitudes at any given elevation can be described with four PCs for all azimuths and four weights ($w_1 - w_4$) per azimuth. During the reconstruction, each of the four weights is multiplied by a corresponding principal component.

$$S = \sum_{i=1}^{37} pc_1 \cdot w_{i1} + pc_2 \cdot w_{i2} + pc_3 \cdot w_{i3} + pc_4 \cdot w_{i4} \qquad (5)$$

In this way, the $w$ matrix specifies the contribution of each $pc$ to the reconstruction of data values.

## Mathematical formulation of PCA weight variations

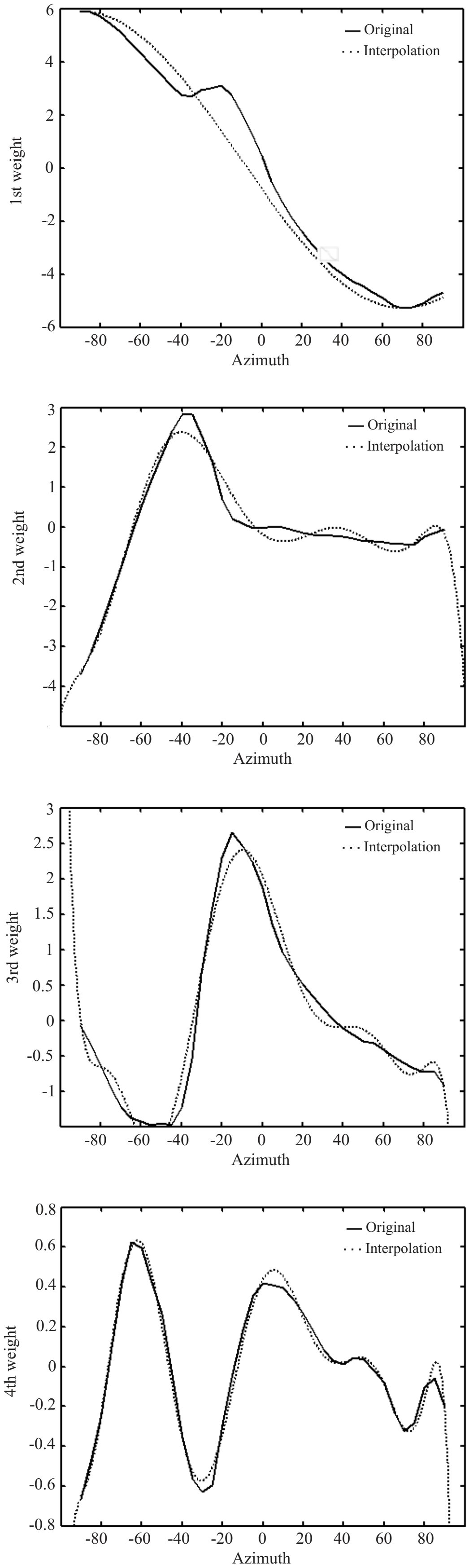
As explained before we aimed to find some hidden mathematical regularities in HRTF data which can not be seen in the original set. Principal components ($pc_1 - pc_4$) which were derived from HRTFs are orthogonal and do not have anything in common. Therefore we focused on weight ($w_1 - w_4$) variations, which have simple courses and can be described or substituted by simple mathematical functions.

The variation of the first and most important weight $w_1$, which is derived from the largest eigenvalue, can be accurately described with a half period of the sinus function (the amplitude and the phase):

$$w_1(az) = 5{,}10 \cdot \sin(az) + 0{,}212$$
$$15(-90^0) \leq az \leq 51(90^0) \qquad (6)$$

The suitable sinus function can be determined using a special algorithm which is based on the least mean square error between the original and reconstructed data.

The variations of the other three $w$ do not have such a simple course, but can nevertheless be described with a lot less coefficients. We decided to use the polynomial approximation with different polynomial degrees. Interestingly, each further $w$ variation requires a higher polynomial degree. The degrees of specific polynomial can be found by using the least mean square error criteria




Figures 1-4. PCA weights at elevation 0º and azimuth from -90º to 90º

(the $w_2$ requires the 8th degree, the $w_3$ requires the 9th degree and $w_4$ requires the 10th polynomial degree).

$$w_2(az) = -2{,}7\cdot 10^{-9} az^8 + 3{,}8\cdot 10^{-7} az^7 - 2{,}2\cdot 10^{-5} az^6 + \\ + 6{,}4\cdot 10^{-4} az^5 + 0{,}0094\cdot az^4 + 0{,}061\cdot az^3 - 0{,}11\cdot az^2 + \\ + 0{,}46\cdot az - 4 \qquad (7)$$

$$w_3(az) = -4{,}2\cdot 10^{-10} az^9 + 7{,}3\cdot 10^{-8} az^8 - 5{,}3\cdot 10^{-6} az^7 + \\ + 2{,}1\cdot 10^{-4} az^6 - 0{,}0048\cdot az^5 + 0{,}063\cdot az^4 - 0{,}48\cdot az^3 + \\ + 1{,}9\cdot az^2 - 3{,}8\cdot az + 2{,}3 \qquad (8)$$

$$w_4(az) = -1\cdot 10^{-11} az^{10} + 3{,}4\cdot 10^{-9} az^9 - 2{,}7\cdot 10^{-7} az^8 + \\ + 1{,}1\cdot 10^{-5} az^7 - -2{,}9\cdot 10^{-4} az^6 + 0{,}0045\cdot az^5 - 0{,}04\cdot az^4 + \\ + 0{,}18\cdot az^3 - 0{,}38\cdot az^2 + 0{,}48\cdot az - 0{,}89 \qquad (9)$$

All polynomials are defined on the interval: $1(-90^0) \le az \le 37(90^0)$

Interpolations for the elevation 0º are shown in Figures 1-4.

# Data reconstruction and model testing

A set of original HRIR can be dealt with as a minimum-phase system [9] on condition that time delays between separate impulse responses are eliminated. These delays describe the time difference between the sound wave impact to the left and to the right eardrum. This phenomenon is called Inter-aural Time Difference (ITD) [10] and it is a very important clue for the determination of the azimuth of virtual sound source position. In the frequency space, the ITD information is stored as phase. The latter was not taken into account in the course of our experiment, although it is vital for the reconstruction of the data. Before transforming HRIR into HRTF, ITD was calculated with the use of the cross-correlation function [13]:

$$R_{xy} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]y[n] \qquad (10)$$

ITD course for azimuths from -90º to 90º at elevation 0º is shown on Fig. 5:

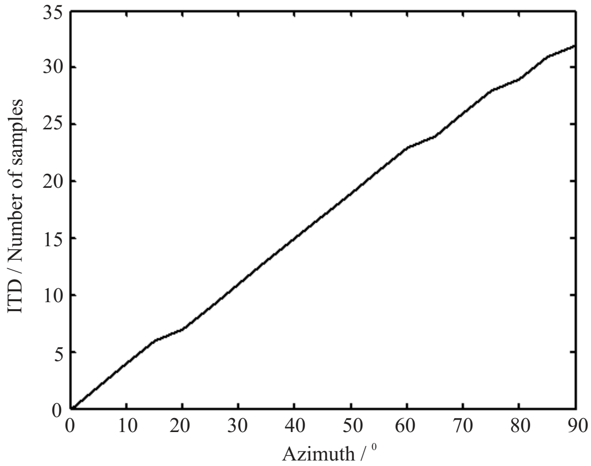



Figure 5. Interaural time difference for horizontal plane

After the reconstruction of the amplitude spectra based on functional presentations of $w_1 - w_4$, the impulse responses were calculated using the inverse Fourier transform and minimum-phase reconstruction.

The mathematical formulation of PCA weights enables the reading of specific values with arbitrary precision. In this way, new HRTFs can be constructed for non-measured spatial positions. The result is not only the representation of data with fewer coefficients, but also an effective interpolation mechanism. Beside interpolating HRTF amplitude, ITD must also be interpolated for each new HRTF. It can be interpolated by using simple linear interpolation, as the ITD course is almost linear (Fig. 5).

The new data was tested on 25 subjects. Using the Matlab programming language, we developed a test environment enabling the generation of spatial sound played through the headphones. The sound source positions could be chosen with a mouse click on a special navigation panel. The test consisted of measuring the average localization time needed for the determination of a specific azimuth by the test subjects. Each azimuth had to be indicated with the precision of 5º. All test subjects performed five localization tests. All tests were carried out both with the original HRIR data and the data reconstructed from $w$ functions and *PCs*.

## Localization test results

The results of our test on the test subjects demonstrate a high accuracy of our model. The ability of spatial sound source localization is almost as good as with the use of the original HRIR data. Tab. 1 and Fig. 6 represent mean localization time in seconds. Each number is the average of five measurements for each test subject. The most important result is approximately the same localization time when using our model if compared to the original HRIR data set (2nd and 3rd column in Tab. 1).

Table 1. Results of localization tests

| | **Mean localization time** | |
|---|---|---|
| **Test subject** | **Our PCA model** | **Orig. HRIR data set** |
| 1 | 5,4 | 5,4 |
| 2 | 3,2 | 3,3 |
| 3 | 3,7 | 3,9 |
| 4 | 4,6 | 4,3 |
| 5 | 9,2 | 8,5 |
| 6 | 6,3 | 6,3 |
| 7 | 5,4 | 6,1 |
| 8 | 6 | 5,8 |
| 9 | 12,3 | 12,4 |
| 10 | 4,4 | 4,4 |
| 11 | 6,1 | 6,2 |
| 12 | 5,4 | 5,3 |
| 13 | 7,3 | 7,1 |
| 14 | 10,2 | 10,4 |
| 15 | 8,2 | 8,1 |
| 16 | 7,1 | 7,1 |
| 17 | 6,2 | 6,2 |
| 18 | 5,4 | 5,5 |
| 19 | 8,9 | 9,1 |
| 20 | 4,2 | 4,0 |
| 21 | 3,8 | 3,9 |
| 22 | 5,2 | 5,5 |
| 23 | 9,4 | 9,1 |
| 24 | 10,1 | 11,0 |
| 25 | 7,4 | 7,2 |

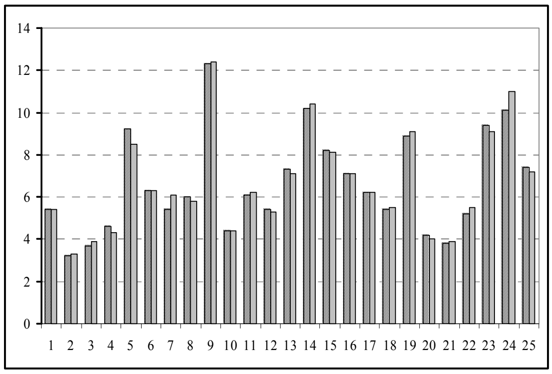



Figure 6. Comparison of test results for PCA model and original HRIR

A quick overview of the results presented above shows much variance in mean localization times for different test subjects. Some of them oriented well in the virtual sound world and performed their localization test very quickly. Others were somewhat confused and needed more time for localization.

Generally, a relatively strong front-back confusion was present. It is the consequence of the use of non-personified HRTFs.

## Discussion and conclusion

The weight variations described with suitable functions allow a very effective presentation of the HRTF data with fewer coefficients. They also enable the generation of spatial sound for arbitrary non-measured positions, as each weight variation is described with a continuous function whose value can be read for any azimuth position.

In the future, it would be useful to test our model on individualized HRTF data. We expect similar functions for the description of the weight variations.

## Acknowledgements

This work has been supported by the Ministry of Education, Science and Sport of Slovenia within the program: Algorithms and optimization methods in telecommunications.

## References

[1] Algazi VR, Duda RO, Thompson DM and Avendano C, “The CIPIC HRTF Database”, Proceedings 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, Mohonk Mountain House, New Paltz, NY, 99-102, 2001.

[2] Calvo RA and Partridge M and Jabri MA, “A Comparative Study of Principal Component Analysis Techniques”, Australian Conference in Neural Networks. Brisbane, 1998.

[3] Cheng CI and Wakefield GH. “Introduction to head-related transfer functions (HRTF's): representations of HRTF's in time, frequency, and space (invited tutorial)”. Journal of the Audio Engineering Society, 49(4):231-249, 2001.

[4] Ferguson BS, Bogner RE, Warwryk S, “A bottle model for Head-Related Transfer Functions”, Proceedings ICASSP98, VI, 3533-3536, USA, 1998.

[5] Gardner B and Martin K, “HRTF Measurements of a KEMAR Dummy-Head Microphone”, MIT Media Lab Perceptual Computing - Technical Report \#280,1994.

[6] Haneda Y, Makino S, Kaneda Y and Kitawaki N, “Common-Acoustical-Pole and Zero Modeling of Head-Related Transfer Functioons”, IEEE Transactions on speech and audio processing, 7(2):188-196, 1999.

[7] Hasegawa H, Kasuga M, Matsumoto S and Koike A, “Simply Ralization of Sound Localization Using HRFT Approximated by IIR Filter”, IEICE Trans. Fundamentals, E83-A(6): 973-978, 2000.

[8] Jin C, Leong P, Leung J, Corderoy A and Carlile S, “Enabling individualized virtual auditory space using morphological measurements”, Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing), 235-238, December, 2000.

[9] Kistler DJ, Wightman FL, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction”, Journal of the Acoustical Society of America, 91(3):1637-1647, 1992.

[10] Kuhn GF, “Model for the interaural time differences in the azimuthal plane”, Journal of the Acoustical Society of America, 62(1):157-167, 1977.

[11] Leung J, Jin C, Carlile S, “An efficient method for the rendering of high fidelity virtual auditory space”, Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing), 216-219 December, 2000.

[12] Martens WL, “Principal components analysis and resynthesis of spectral cues to perceived direction”, Proceedings of the International Computer Music Conference, 274-281, 1987.

[13] Oppenheim AV, Schafer RW, “Discrete-Time Signal Processing”, Prentice-Hall International (UK) Limited, London, 1989, 5:230-250.

[14] Shlens J, “A tutorial on principal component analysis: Derivation, Discussion and Singular Value Decomposition”, Version 1, March 2003, http://www.snl.salk.edu/~shlens/pub/notes/pca.pdf

[15] Susnik R, Sodnik J, Umek A and Tomazic S, “Spatial sound generation using HRTF created by the use of recursive filter”, Proceedings of EUROCON 2003, Ljubljana, Slovenia , 1:449-453, 2003.

[16] Wenzel EM, Arruda M and Kistler D, Wightman FL, “Localization using nonindividualized head-related transfer functions”, Journal of the Acoustical Society of America, 94(1):111-123, 1993.