DIRECTIONAL INFORMATION IN HEAD RELATED TRANSFER FUNCTIONS

Jaka Sodnik and Saso Tomazic

University of Ljubljana, Faculty of Electrical Engineering
Slovenia, Europe

ABSTRACT

The synthesis of spatial sound which is played through the headphones can be done by the use of Head Related Transfer Functions (HRTF). The latter describe the changes in the sound wave as it propagates from a spatial sound source to the human eardrum. The functions include many different factors which influence the accuracy of spatial sound source perception. The most important factors are Inter-aural time difference (ITD), Inter-aural level difference (ILD) and the spectral changes for different spatial positions. The main goal of our research was determining with the aid of Principal Component Analysis (PCA) the spectral variation important for the perception of direction. We concentrated on the azimuth perception in the horizontal plain. PCA was applied to HRTF linear amplitude spectra, thus excluding the phase spectrum and ITD. The analysis assembled all spectrum variations important for the azimuth perception in the first and partially second PCA weight variation. The hypothesis was tested on and confirmed by 15 test subjects who preformed the virtual sound sources localization tests.

1. INTRODUCTION

The human brain with its sense of hearing is able to determine the spatial sound source position with a high accuracy. The accuracy depends on many factors which have been studied and defined very thoroughly. Some of these factors are more important for the azimuth perception and some for the elevation perception. All these factors are contained in Head Related Transfer Functions (HRTF). In the time domain, HRTFs can be represented as Head Related Impulse Responses (HRIR). They describe the changes in the sound wave as it propagates from a spatial sound source to the human eardrum. These impulse responses can be used as FIR filters with a given number of coefficients. By filtering an arbitrary sound source, spatial sound can be generated and played through the headphones. To be able to generate an arbitrary spatial sound, HRIRs have to be measured for all the necessary spatial positions.

HRTFs include all the key factors important for spatial sound reproduction. Two most important factors for the azimuth perception are Inter-aural time difference (ITD) and Inter-aural level difference (ILD). Their influence depends on the frequency of the given sound. ITD is more important for the frequencies from 0-1400 Hz and ILD is more important for the frequencies above 1400 Hz. Each sound wave from an arbitrary space position reaches the left and the right eardrum at a different time. This difference in time is called ITD [10]. The amplitude difference in the sound wave at both eardrums is called ILD.

ITD depends on the source distance, head and shoulders shape and also on the frequency. Kuhn [10] measured and modeled ITD for different frequency bands. On the other hand, the elevation perception depends mostly on spectral contents of HRTFs. The most important factors are the amplifications and attenuations of different frequency bands.

The purpose of our research was the determination of the segment of HRTF spectral variation which is important for azimuth perception. We studied the linear amplitude spectra derived from MIT Media Lab impulse responses [1]. The linear amplitudes did not contain any information about the phase spectra, which eliminated the ITD factor from the analysis. Based on our past research we applied Principal Component Analysis (PCA) [5] as the mathematical tool for our calculations. PCA decomposed all amplitude spectra to only a few basic functions and their corresponding weights. We concentrated mostly on PCA weights variations for different azimuths. Each weight and its variation are important for the reconstruction of the original data in a different way. The variations of the most important weights are modeled with simple monotonic functions. The less important weights are described with more complicated and non-monotonic functions. Our hypothesis anticipates that only the monotonic variations of weights contain the information about the direction. The reason for this lies in the bijection of the monotonic function, which enables a uniform determination of directions. If a particular function is not monotonic and is therefore injective, each weight value belongs to more than one direction. Considering our hypothesis, HRTF amplitude values can be reconstructed using only monotonic PCA weights, while the others can be approximated by using their mean value.

2. DETERMINATION OF ITD AND LINEAR AMPLITUDES

The basic MIT Media Lab HRTF library contains from 1 to 37 measurements at different elevations. We used the set of 37 impulse responses for the horizontal plain (elevation -20^0). This set includes the measurements for

0-7803-8560-8/04/$20.00©2004IEEE

azimuths from -90^0 to 90^0. These impulse responses are very much alike and differ only in the specific time delay (ITD), and can therefore be handled as a minimum phase system. ITD can be determined or eliminated by calculating the cross correlation (1) between particular responses [2].

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

The amplitude spectra can be calculated from the absolute part of the Fourier transform of the impulse responses. The phase spectra needed for the reconstruction can be derived from the amplitude spectra using Hilbert transform [3].

3. PRINCIPAL COMPONENT ANALYSIS

PCA is a statistical method based on the Gaussian distribution of random variable. According to Shlens [4], it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA reduces a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it.

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

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \quad (2)$$

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

$$C_X = \frac{1}{n-1} XX^T \quad (3)$$

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 [5].

$$P = \left[p_1, p_2, \ldots, p_j \right] \quad (4)$$

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

$$w_i = P^T \cdot x_i \qquad (5)$$

PCA has already been applied to HRTFs by Martens [7], Kistler and Wightman [6], and also Middlebrooks [9]. All the authors pointed out substantial data reduction, as this method allows the description of all HRTF data with merely 4-7 basic functions and their corresponding weights. The PCA of personalized HRTFs efficiently pointed out some common properties and some main differences in HRTFs for various test subjects.

We applied PCA on MIT Media Lab library of non-personalized functions. The series of 37 amplitude spectra were decomposed to four common basic functions and 37 times 4 weights (for 37 azimuth positions). That means we took into consideration only the four basic eigenvectors of the covariance matrix. The number four was chosen due to the minimal square error (the difference between original and reconstructed amplitude spectra).

Not all four basic functions are equally important for the reconstruction. The first basic function or the first eigenvector represents approximately 75% of the variation of the complete HRTF set. All other functions represent less than 10% of variation each [6]. Consequently, if we take into account only the first basic function and its corresponding weight, we can reconstruct the data with 75% accuracy. Using all four basic functions, the data can be reconstructed with a 90% accuracy. The reconstruction accuracy also depends on the PCA weights variation for different azimuths. The PCA weights determine the contribution of each basic function to each source direction (figure 2-5: solid lines).

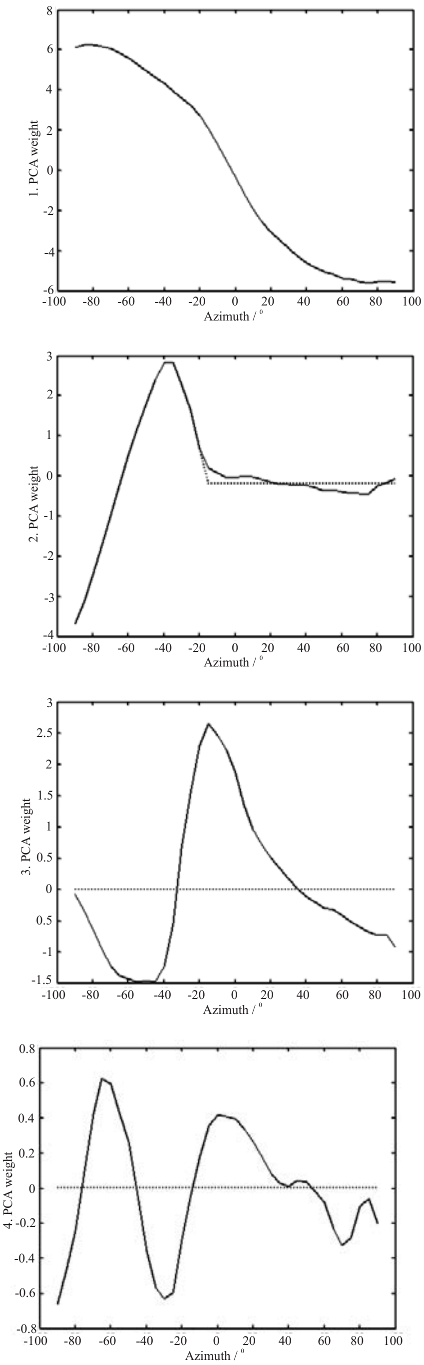
Our hypothesis presupposes that the human ear or human brain can not separate two instances with the same contribution of the two basic functions. That means that each basic function contains the direction-dependent information only for those azimuths where its corresponding weight is monotonic and bijective. The review of all PCA weights shows that only the first and partially the second weight (only for specific azimuths) can be described with a monotonic function. The third and fourth weights are non-monotonic and do not contain any directional information.

The course of amplification (i.e. ILD) of the original HRTF data proves to be very similar to the course of the first and most important PCA weight. The cross-correlation between them is higher than 0.95. That confirms the fact that ILD, which is the second most important factor for the azimuth perception (as ITD was not included in analysis), is collected in first PCA weight variation. The first variation is also the only one that is monotonic for all given azimuths.

This claim can be tested by using one amplitude spectrum and creating a set of spectra by changing only the attenuation. If we apply PCA on that set of functions, only one eigenvector is non-zero. The course of the corresponding weight is exactly the same as that of the attenuation chosen at the beginning of the experiment.

In our case, some directional information is also gathered in the monotonic part of the second PCA weight (only for azimuths from -90^0 to -20^0).

Considering these facts, the reconstruction of original data set can be done by using only the monotonic weight functions. The first PCA weight and the monotonic part of second PCA weight have to be multiplied precisely by the corresponding basic functions, while the remaining part of the second, third and fourth weight can be substituted with their mean value. The new courses of all four weights are presented in the following figures (dotted lines):



Figures 1-4: Four PCA weights for all azimuths

In this way it is possible to reconstruct the amplitude spectra for all 37 azimuths.

4. TESTING ENVIRONMENT AND PROCEDURE

Testing environment was developed in Matlab 6.5 and Visual Basic 6.0 programming languages. PCA was preformed in Matlab by preparing three different function sets and saving them as WAV files. The first test set consisted of the sound recordings for all azimuths from -90^0 to 90^0 using the original MIT Media Lab impulse responses. The second test set comprised the PCA model with a complete reconstruction using all four basic functions and corresponding weights. The third test model was our proposed PCA model taking into account only the monotonic parts of PCA weights. The test sound in all three cases was a 100 ms long sequence of white noise which had proven to be very efficient for virtual sounds localization tests [8].

The test environment was developed in Visual Basic programming language, which enabled the playing of WAV files. It contained a special control panel with the azimuth and elevation coordinates. The program randomly chose a position in the horizontal plain as a reference sound source. The test subjects played the spatial sound coming from the reference point by clicking on the mouse button. The coordinates of the second virtual sound source were determined by the current mouse cursor position. This second sound source was played at constant intervals. The objective of the test was for the test subjects to move the mouse cursor to the reference position of the sound source chosen by the program. The task should be done as quickly as possible and with as few clicks as possible (as each click played the sound from the reference position).

The tests were preformed with 15 test subjects with normal sight and hearing. Each test subject repeated the test with all three models for five times (15 trials all together). The localization time was recorded for each trial.

5. RESULTS

The test subjects were asked to localize the virtual sound source five times for each test set of functions. Before actually performing the experiment, each test subject had some time to adjust to the spatial sound in the headphones. As the consequence of the use of non-individualized HRTFs strong front-back confusion was present.

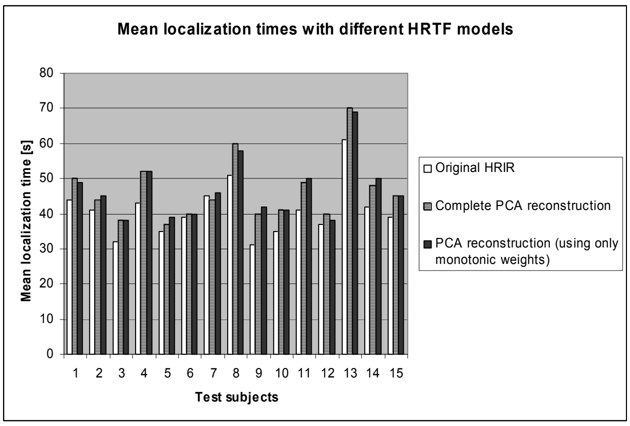


Figure 5: Comparison of three localization tests

This phenomenon had already been explained before [11] and did not disturb our test procedure.

Tab. 1 and Fig. 5 present the mean localization time (the mean time of all five trials) for each HRTF model and for all test subjects.

Test subjects	Original HRIR [s]	Complete PCA reconstruction [s]	PCA reconstruction (using only monotonic weights) [s]
1	44	50	49
2	41	44	45
3	32	38	38
4	43	52	52
5	35	37	39
6	39	40	40
7	45	44	46
8	51	60	58
9	31	40	42
10	35	41	41
11	41	49	50
12	37	40	38
13	61	70	69
14	42	48	50
15	39	45	45

Table 1: Measurement results for all three HRTF models

As can be seen from the figure 6, the shortest localization times were achieved with the use of the original HRIR (white columns). The localization times using PCA models are somewhat longer, regardless whether we included the PCA weights (dashed columns) or just the monotonic PCA weights (black columns).

6. CONCLUSION

The results presented above confirm the effectiveness of the proposed model. The decomposition of the original data by PCA contributes to some loss of accuracy which can be seen when comparing the first measurement result with the other two. On the other hand, PCA effectively exposes all the direction depended information in the first and partially second PCA weight. The rest of the unnecessary information or variation can be easily eliminated and replaced with its mean value. The effectiveness of such procedure can be proven by comparing the measurement results of the second and the third HRTF set. The mean localization time is practically the same when using the complete PCA reconstruction or just the monotonic weights.

In the future we are planning to measure our own personalized HRIR and repeat the analysis. We expect the same effectiveness of PCA in separating the direction-dependent from direction-independent information in the HRTF set.

7. ACKNOWLEDGMENTS

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

8. REFERENCES

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

[2] S. Tomažič, S. Leonardis, “Digitalni signali in sistemi”, Fakulteta za elektrotehniko, Ljubljana, ch. 1, pp. 16-17, 2004.

[3] A.V. Oppenheim, R.W. Schafer, “Discrete-Time Signal Processing”, Prentice-Hall International (UK) Limited, London, ch. 5, pp. 230-250, 1989.

[4] J. Shlens, “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

[5] J. Sodnik, R. Susnik, S. Tomazic, „Approximation of Head Related Transfer Functions”, HCI International 2003: adjunct proceedings, pp. 99-100, June 2003.

[6] D.J. Kistler, F.L. Wightman, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction”, J. Acoust. Soc. Am., vol. 91, no. 3, pp. 1637-1647, March 1992.

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

[8] R. Susnik, J. Sodnik, S. Tomazic, “Sound source choise in HRTF acoustic imaging”, HCI International 2003 : adjunct proceedings, pp. 101-102, June 2003.

[9] J.C. Middlebrooks and D.M. Green, “Observations on a principal components analysis of head-related transfer fuinctions”, J. Acoust. Soc. Am., vol. 92, no. 1, pp. 597-599, July 1992.

[10] G.F. Kuhn, “Model for the interaural time differences in the azimuthal plane”, J. Acoust. Soc. Am., vol. 62, no. 1, pp. 157-167, July 1977.

[11] A.W. Bronkhorst, “Localization of real and virtual sound sources”, J. Acoust. Soc. Am., vol. 98, no. 5, pp. 2542-2553, November 1995.