Resolution enhancement of a general HRTF library

Jaka Sodnik, Rudolf Susnik and Saso Tomazic

Faculty of Electrical Engineering, University of Ljubljana, Trzaska 25, Ljubljana, Slovenia

ABSTRACT

Head Related Transfer Function (HRTF) libraries enable the generation of virtual acoustic sources in arbitrary spatial positions. In experiments with a large number of test subjects, general or non-personified HRTF libraries are used, measured with human-shaped dolls. The resolution of azimuth in the frontal area is much better than in lateral areas. To improve resolution in lateral areas, Principal Component Analysis (PCA) on 37 HRTF magnitudes at elevation 0^0 was performed. The 1st and, in part, the 2nd weight were found to be monotonic functions of azimuth, with greater slope at an azimuth of around 0^0 and smaller in lateral areas. To enhance resolution in lateral areas, the weights were linearly interpolated and the slope was increased to become constant over all azimuths in the median plane (from -90^0 to 90^0) and reconstructed HRTFs using those modified weights. Localization tests performed with original and modified HRTFs showed that resolution in the frontal area remains the same, while there is a noticeable improvement of resolution in the lateral areas (azimuths $\pm 90^0$ and $\pm 45^0$).

INTRODUCTION

Perception of spatial sound is very important for orientation in space, driving a car, communicating with a group of people, etc. Imagine watching a sci-fi movie at the cinema with a modern, surround-sound speaker system. It is quite easy to understand how to create spatial sound with six, seven or more speakers at different positions in space. Listening to spatial sound through headphones is even more fascinating. The creation of spatial sound in headphones or the synthesis of so-called virtual acoustic sources at different spatial positions can be done with the aid of Head Related Impulse Responses (HRIRs) or their Fourier pair Head Related Transfer Functions (HRTFs) (Cheng & Wakefield 2001). HRIRs describe the influence of acoustic path on the source signal as it propagates from the speaker to the listener's eardrums. There are three very important cues for the perception of acoustic source position:

- inter-aural time difference (ITD)
- inter-aural level difference (ILD)
- transformation of spectral content of acoustic signals along their propagation path

All three factors are contained in HRIRs and therefore enable regeneration of spatial sound in headphones. HRIRs are usually recorded in a noiseless anechoic room for each listener individually (to obtain personified HRIR libraries). Two small microphones are put in the ears and the impulse responses are recorded. The speaker is moved around the room to measure different spatial positions relative to the listener.

Measuring impulse responses for separate listeners is quite difficult and takes a lot of time and patience. Therefore, HRIRs are often measured with human-shaped dolls so that general or non-personified HRIR libraries can be obtained. General libraries are usually less accurate but more practical for acoustic tests with a large number of test subjects.

Previous localization tests with different general HRIR libraries showed a significant drop of accuracy compared to personified HRIR libraries (Bronkhorst 1995), (Wenzel et al. 1993), especially in localization of acoustic source elevation. Azimuth localization tests did not prove to be as poor (Sodnik et al. 2004a). Tests with a general MIT HRTF library showed that the highest resolution for azimuth perception is in the central area (azimuth 0^0), a little lower in the extreme left or right areas (azimuth $\pm 90^0$) and lowest in the middle (azimuth $\pm 45^0$) (Sodnik, Susnik & Tomazic 2005). Here resolution is defined to be the ability of human listeners to separate sounds from two virtual sources in near proximity; for example, if the listener can distinguish two sources at azimuths 35^0 and 45^0, the resolution of the library at central azimuth 45^0 is 10^0.

In order to explain the purpose of the current research, some background is required. Another interesting and potential application where spatial sound could be used is so-called acoustic imaging of space, where acoustic imaging is the description of a visual image with sound. Visual information that can be perceived through the eyes contains a large amount of information that could be, in this case, perceived through the ears. Therefore an efficient way of coding this information is required, and spatial sound extends the options available. Positions of different objects in an acoustic image can be appropriately described with different spatial sounds. Our paper deals with the left-right direction, or azimuth, of those options. In order to make the application work for many different users, a general, or non-personified, HRTF library is used. The aim of our research is to improve the azimuth resolution of such a library.

In some previous work, Principal Component Analysis (PCA) proved to be a very effective method to establish those HRTF components important for accurate azimuth perception. It has been shown that only a few of the largest principal components (PCs) with corresponding weights can sufficiently describe those variations in HRTFs necessary for azimuth perception. The first weight as a function of azimuth is practically identical to ITD as a function of azimuth (Kistler & Wightman 1992). The interesting fact is that only the first two weights are almost monotonic functions of azimuth, while other weights are all non-monotonic. The hypothesis proposed here anticipates that only the monotonic variations of weights contain the information about direction (Sodnik & Tomazic 2004b). 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 that fact, the first two weights were modified to become completely monotonic along all azimuths from -90^0 to +90^0 and their slope increased. The impulse responses were reconstructed and resolution measurements at different central azimuths were performed. The resolution actually improves, especially at central azimuths of ±45^0 and ±90^0.

METHODS

MIT Media Lab HRTF library

In the experiment, the MIT Media Lab general HRTF library (Gardner & Martin 1994) was used, originally consisting of measurements at 710 positions in space. As the experiment was concentrating on perception of azimuth, only functions at elevation 0^0 were used. Therefore, 37 impulse responses (azimuths from -90^0 to +90^0) were dealt with, each consisting of 512 samples. The sample frequency of the MIT library was 44.1 kHz. All impulse responses were filtered with the inverse filter of the Optimus 7 Pro speaker used in the HRTF measurements. Information on ITD was calculated from impulse responses using cross-correlation (Oppenheim & Schafer 1989), and it is almost a linear function of azimuth. Fourier transforms of the original data were calculated to obtain HRTF magnitudes, each consisting of 257 samples.

Principal Component Analysis (PCA) of HRTF

PCA can be effectively described as a simple, non-parametric method of extracting relevant information from confusing data sets (Shlens et al. 2003). PCA reduces a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it.

The input to PCA is the matrix **X**, consisting of zero-mean HRTF magnitudes at a chosen elevation and at varying azimuths:

$$\mathbf{X} = \left[\mathbf{H}(\varphi_1) - \overline{\mathbf{H}}, \ldots, \mathbf{H}(\varphi_n) - \overline{\mathbf{H}}, \ldots, \mathbf{H}(\varphi_N) - \overline{\mathbf{H}} \right] \tag{1}$$

where $\mathbf{H}(\varphi_n)$ is a column vector containing M samples of HRTF magnitude at azimuth φ_n and $\overline{\mathbf{H}}$ is a column vector containing mean values of samples at a chosen elevation. The vectors $\mathbf{H}(\varphi_n)$ can be expressed in a new coordinate system with base vectors $\mathbf{pc}_k$:

$$\mathbf{H}(\varphi_n) = \sum_{k=1}^{M} \mathbf{pc}_k \cdot w_k(\varphi_n) + \overline{\mathbf{H}} \tag{2}$$

where $\mathbf{pc}_k$ are the principal components (PCs), i.e., orthonormal eigenvectors of the covariance matrix $\mathbf{C}$

$$C = \frac{1}{N-1} \mathbf{X} \cdot \mathbf{X}^T \tag{3}$$

ordered according to the decreasing values of their respective eigenvalues. PCs represent the directions with the largest variances, i.e., the directions in which the input data change the most. Weights $w_k(\varphi_n)$ are functions of the azimuth φ_n and they describe the contribution of the particular PC to the reconstruction.

PCA has already been applied to different HRTF libraries by several authors (Martens 1987), (Sodnik et al. 2003), (Sodnik et al. 2004c). They tried to replace the entire HRTF set with the minimum principal components (PCs) possible. Kistler and Wightman (1992, pp. 1637-1647) reported that the variations of HRTFs at different spatial positions can be described with only 4 PCs. They applied PCA on log-magnitudes of HRTF and performed minimum phase reconstruction. On the other hand, Middlebrooks and Green (1992, pp. 597-599) reported that PCAs of different personalized HRTF libraries point out some common properties and main differences of many test subjects.

Linearization of weights

In the current work PCA was performed on just one specific part of the general HRTF library, on linear magnitudes of 37 azimuths (from -90^0 to +90^0) at elevation 0^0. The first three base vectors $\mathbf{pc}_k$ are illustrated in Fig. 1:

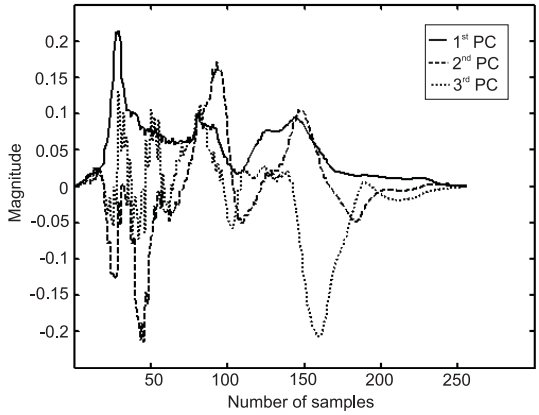

Figure 1. First three principal components

Of particular interest here is the variation of weights as functions of azimuth: $w_k(\varphi_n)$. The first three weights are shown in Fig. 2:

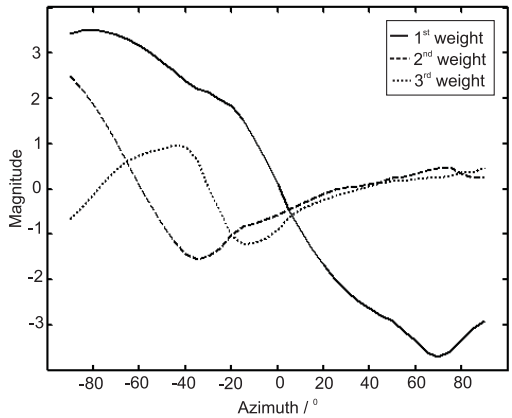

Figure 2. First three weights

It has been established that in PCA with such input data, $\mathbf{pc}_1$ and $\mathbf{pc}_2$ with $w_1(\varphi_n)$ and $w_2(\varphi_n)$ sufficiently describe those variations in HRTFs necessary for azimuth perception (Sodnik & Tomazic 2004b). As mentioned in the introduction, it is thought that the reason for this lies in the monotony of $w_1(\varphi_n)$ and the first part of $w_2(\varphi_n)$.

VERIFICATION OF A NEW MODEL WITH LINEAR WEIGHTS

Since monotony of $w_1(\varphi_n)$ and $w_2(\varphi_n)$ is important for correct azimuth perception, these two weights were modified to become monotonic across all azimuths (from -90^0 to +90^0). A linear approximation for $w_1(\varphi_n)$ and $w_2(\varphi_n)$ was made, as shown in Fig. 3.

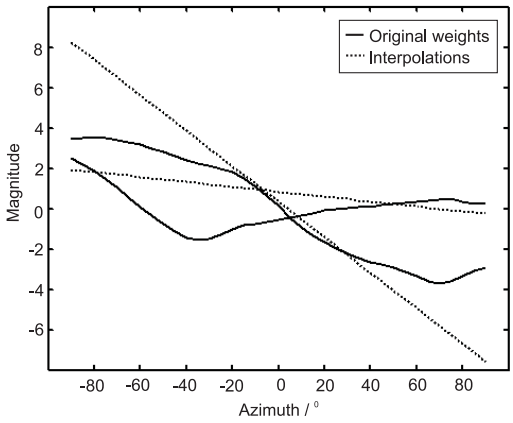

Figure 3. Linearization of the first two weights

Two different models were reconstructed in order to verify the correctness of the hypothesis. In the first model, called **PC12**, HRTFs using only $\mathbf{pc}_1$ and $\mathbf{pc}_2$ and linear approximations of $w_1(\varphi_n)$ and $w_2(\varphi_n)$ were reconstructed. All other PCs were discarded. In the second model, **CMPL**, HRTFs using all PCs and weights were reconstructed, but $w_1(\varphi_n)$ and $w_2(\varphi_n)$ were substituted with their linear approximations. In both cases impulse responses for 37 azimuths were obtained with minimum phase reconstruction. Additionally, information on ITD was added, as it has been proved to be very important for correct azimuth perception.

The two models together with the original HRTF library were submitted to resolution measurement tests at five central azimuths separately: 0^0, ±45^0 and ±90^0 as shown in Fig. 4:

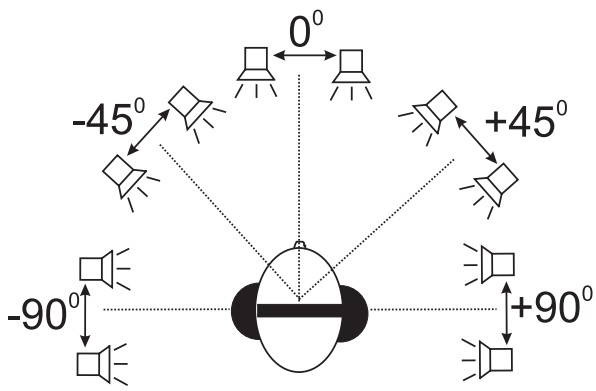

Figure 5. Resolution measurement areas

The resolution measurement procedure consisted of playing a test signal from two close virtual acoustic sources alternately. Resolution at each central azimuth defines the minimum detectable distance (angle) between two acoustic sources.

The test signals were 2 second-long, repeating sequences (1 second of white noise and 1second of silence).

Measurements were taken with Acer TravelMate 4000 notebook computer with a Digigram VXpocket 440 sound card and Sennheiser HD270 headphones in a quiet place with an ambient noise level of 38dB-42dB (measured with a Lutron SL-4012 sound level meter). The Sennheiser HD270s are studio headphones with excellent attenuation of ambient noise (-10dB to -15dB). Prior to the measurements, the impulse responses of the headphones were measured to create an inverse filter and eliminate their influence on the measurement results.

15 volunteers participated in the resolution measurements; none of them reported any hearing problems.

Test signals were calculated and prepared in the Matlab programming language and converted to Windows Audio Video Format – WAV files. The testing console, with a simple administration panel developed in MS Visual Basic 6, enabled simple playback of different WAV files and automatic logging. The computer played back test signals from two proximate sources and slowly increased the difference in their azimuth from 0^0 to 40^0. The test signal was played three times alternately from each source before the difference was increased (every 6 seconds). Test subjects were asked to press a button at the moment when they were able to differentiate two directions of the source sounds. The computer logged the current difference in azimuth and number of required repetitions.

MEASUREMENT RESULTS AND DISCUSSION

The measurement results are dealt with separately for three areas with different central azimuths. Results at central azimuth +90^0 and -90^0 are combined, as well as results at central azimuths +45^0 and -45^0.

Average resolutions are provided with confidence intervals:

$$RES = \bar{x} \pm s \tag{4}$$

for each HRTF library at different central areas. The average of all test subjects is calculated as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5}$$

and

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{6}$$

is the dispersion.

Average resolutions with confidence intervals are presented in Table 1 and Fig. 5:

Table 1. Average resolutions with confidence intervals

Area	HRTF library					
	Original		PC12		CMPL	
	RES / ^{0}	s / ^{0}	RES / ^{0}	s / ^{0}	RES / ^{0}	s / ^{0}
0^0	5.45	1.51	7.73	3.44	5.45	1.51
$\pm45^0$	6.82	3.37	8.64	3.23	5.91	3.02
$\pm90^0$	5.91	2.02	8.18	2.52	5.45	1.51

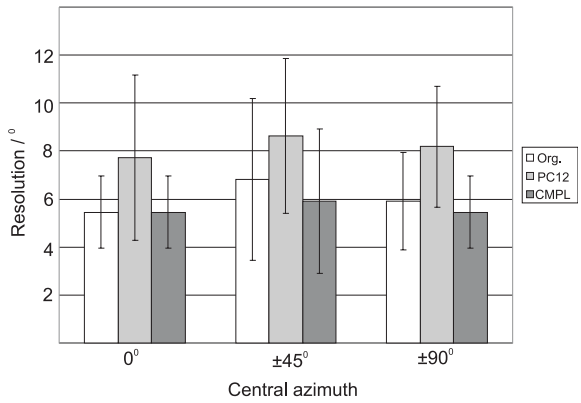

Figure 5. Average resolutions with confidence intervals

The highest column means the lowest resolution and vice versa.

The average number of repetitions or the reaction time for detection of differences in azimuth at each area is presented in Table 2:

Table 2. Number of repetitions at different azimuths

Area	HRTF library		
	Original	PC12	CMPL
0^0	1.73	2.12	1.41
$\pm45^0$	2.27	2.38	1.11
$\pm90^0$	2.07	2.85	1.07

As seen in Table 1 and Figure 4, the highest average resolutions were measured with the CMPL model. The main result of our experiment is the comparison of the original and CMPL models. The improvement of performance can be seen in three different aspects. First of all, average resolution increases at all central azimuths, especially those in lateral areas. The second advantage of the CMPL model is the reduction of confidence intervals, which means that results do not differ much from test subject to test subject. The third important fact is the reduction of reaction time (Table 2). All test subjects were able to differentiate between two close acoustic sources faster.

According to these results, the PC12 model decreases resolution compared to the original library.

PCA reconstruction with just linear $w_1(\varphi_n)$ and $w_2(\varphi_n)$ decreases localization performance in all areas, since the average resolution of the PC12 model is the worst. This was unexpected, since previous experiments showed that only $w_1(\varphi_n)$ and $w_2(\varphi_n)$ are necessary for correct azimuth perception.

The side effect reported by test subjects was the perception of a change of elevation. As mentioned before only HRTF data at elevation 0^0 was used in the experiments. At central azimuths of $\pm90^0$ some virtual acoustic sources appeared at different elevations. It is believed that the reason for this lies in the linearization of $w_1(\varphi_n)$ and $w_2(\varphi_n)$. Because $w_1(\varphi_n)$ as a function of azimuth is proportional to ILD, it is infered that linearization of ITD causes irregularities in elevation perception. This is not disturbing because in this instance correct elevation perception is considered to be irrelevant.

CONCLUSION

The results of the experiments show that perception resolution of general HRTF libraries can be improved. The experiment also confirms the significance of the 1st and 2nd PCA weights, as well as their monotonity for azimuth perception. With linear substitution of their non-monotonic parts and increase of slope significant improvement in azimuth localization was achieved. Improvement can be seen as higher resolution in lateral areas (central azimuths $\pm45^0$ and $\pm90^0$), smaller confidence intervals and shorter response time.

The interesting and unexpected result is the decreased effectiveness of the PC12 model compared to the CMPL model. Previous tests with similar HRTF models have shown no difference in azimuth localization performance when reconstructing data with only the first two or all PCs.

General HRTFs are very suitable for applications in which a large number of different users require spatial sound. The simple modification described in this paper offers significant improvement of general HRTF library resolution, especially in lateral areas. In the future, it would be interesting to compare the effectiveness of a modified general HRTF library with a personified HRTF library.

ACKNOWLEDGMENTS

This work was supported by Ministry of High School and Science of Slovenia within a project Algorithms and optimization methods in telecommunications.

REFERENCES

Bronkhorst, AW 1995, 'Localization of real and virtual sound sources', *Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2542-2553.

Cheng, CI & Wakefield, GH 2001, '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*, vol. 49, no. 4, pp. 231-249.

Gardner, B & Martin, K 1994, 'HRTF Measurements of a KEMAR Dummy-Head Microphone', *MIT Media Lab Perceptual Computing – Technical Report #280*.

Kistler, DJ & Wightman, FL 1992, 'A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction', *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637-1647.

Martens, WL 1987, 'Principal components analysis and resynthesis of spectral cues to perceived direction',

Proceedings of the International Computer Music Conference, pp. 274-281.

Middlebrooks JC & Green, DM 1992, 'Observations on a principal components analysis of head-related transfer functions', *Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 597-599.

Oppenheim, AV & Schafer, RW 1989, *Discrete-Time Signal Processing*, Prentice-Hall International (UK) Limited, London, pp. 230-250.

Shlens, J, *A tutorial on principal component analysis: Derivation, Discussion and Singular Value Decomposition*, Retrieved: March 2003, from http://www.snl.salk.edu/~shlens/pub/notes/pca.pdf.

Sodnik, J, Susnik R & Tomazic, S 2003, 'Approximation of Head Related Transfer Functions', *HCI International 2003: adjunct proceedings*, pp. 99-100.

Sodnik, J, Susnik, R & Tomazic, S 2005, 'Acoustic signal localization through the use of Head Related Transfer Functions', *Journal of Systemics, Cybernetics and Informatics*, vol. 2, no. 6.

Sodnik, J, Susnik, R, Bobojevic, G & Tomazic S 2004a, 'Smerna ločljivost navideznih izvorov zvoka pri človeku', *Electrotechnical Review*, vol. 71, no. 3, pp. 121-127.

Sodnik, J, Tomazic, S 2004b, 'Directional information in head related transfer functions', *Tencon 2004 : conference proceedings, Analog and Digital Techniques in Electrical Engineering*, Chiang Mai, Thailand, Venue, Lotus Hotel Pang Kaew, pp. 100-103.

Sodnik, J, Umek, A, Susnik, R, Bobojevic, G & Tomazic S 2004c, 'Representation of head related transfer functions with principal component analysis', *Proc. of the Annual Conference of the Australian Acoustical Society*, Gold Coast, Australia, pp. 603-607.

Wenzel, EM, Arruda, M & Kistler, D, Wightman, FL 1993, Localization using nonindividualized head-related transfer functions', *Journal of the Acoustical Society of America,* vol. 94, no. 1, pp. 111-123.