Journal of Sound and Vibration 329 (2010) 4093–4106

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

A hybrid algorithm for selecting head-related transfer function based on similarity of anthropometric structures

Xiang-Yang Zeng *, Shu-Guang Wang, Li-Ping Gao

Institute of Environmental Engineering, P.B. 58, Northwestern Polytechnical University, Xi'an 710072, China

ARTICLE INFO

Article history:
Received 14 April 2009
Received in revised form
23 March 2010
Accepted 29 March 2010
Handling Editor: K. Shin

ABSTRACT

As the basic data for virtual auditory technology, head-related transfer function (HRTF) has many applications in the areas of room acoustic modeling, spatial hearing and multimedia. How to individualize HRTF fast and effectively has become an opening problem at present. Based on the similarity and relativity of anthropometric structures, a hybrid HRTF customization algorithm, which has combined the method of principal component analysis (PCA), multiple linear regression (MLR) and database matching (DM), has been presented in this paper. The HRTFs selected by both the best match and the worst match have been applied into obtaining binaurally auralized sounds, which are then used for subjective listening experiments and the results are compared. For the area in the horizontal plane, the localization results have shown that the selection of HRTFs can enhance the localization accuracy and can also abate the problem of front–back confusion.

1. Introduction

Defined as "the ratio of the Fourier transform of the signal at the listener's eardrum to that at the center of the listener's head with the listener absent", head-related transfer function (HRTF) and its corresponding impulse response, head-related impulse response (HRIR), are essential components of many approaches to binaurally based spatial audio synthesis. They describe the changes in the sound wave as it propagates from a spatial sound source to the human eardrums [1]. The acquisition of accurate binaural HRTFs is crucial to the generation of 3D sound. Because of the individual difference of anthropometric shape and size, HRTF varies with frequencies, directions and subjects [2]. Unfortunately, it is difficult and time consuming to measure HRTF, and is not feasible to obtain the binaural HRTFs for an arbitrary listener. Therefore, how to individualize HRTF fast and effectively becomes an opening problem.

In the recent years, more and more researchers have concentrated on the individualization of HRTF. Besides the direct measuring method, calculating by boundary element method (BEM) is also a way that can obtain good accuracy [3]. However, these two types of methods are very time consuming and are difficult to implement in the study of auralization. Some researchers tried to find the relationship between the anthropometric structures of the subject and the corresponding HRTFs, from which some simpler ways could be found to predict the personalized HRTFs. Several types of prediction algorithms have been brought forward, such as database matching [4], principal component analysis [5], structure modeling [6,7] and other statistical approaches [8,9]. Some of these approaches are faster while others are more accurate [10]. However, in our study on real-time auralization for room acoustic modeling, the binaural information is

* Corresponding author.
E-mail address: zenggxy@nwpu.edu.cn (X.-Y. Zeng).

0022-460X/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2010.03.031

necessary for an arbitrary listener. To save computation time we need an improved individualization approach, which is effective as well as accurate. Since the direct measurement for each listener is quite time consuming and not practical, we hope to provide well-performing HRTFs of a listener based on just a few measurements on him. The basis of our method includes two aspects: one is that there is similarity in various persons' anthropometric structures; another is that only part of the anthropometric parameters are crucial to the spatial hearing.

For simplicity and effectiveness of database matching method, we think it can be improved as a practical approach for our applications in auralization. To select HRTFs for any listener from a measured database, we present a hybrid algorithm that combines the method of principal component analysis, multiple linear regression analysis and database matching. First, we use principal component analysis to decompose HRTFs and extract the characteristic parameters. Then we use multiple linear regression to analyze the relationship between HRTFs and the anthropometric parameters and to find the reference parameters. Last, we use database matching algorithm to find the closest HRTFs for the listener. The algorithm has been tested by subjective localization experiments using auralized sounds produced by the convolution of dry sound and the selected HRTFs. In our current research, only the horizontal plane data are used for database matching and listening tests. The detailed algorithm will be presented in the next section.

2. Methodology

Since there are a number of anthropometric parameters related to human's hearing, the core of the hybrid algorithm is to find the most crucial anthropometric parameters (as reference parameters). Once they have been found, the main job that needs to be done in applications is to measure them and use them for database matching in order to find the best matched HRTFs.

The structure of the algorithm is shown in Fig. 1. Firstly, the direction transfer function (DTF) can be calculated from the measured HRTFs. Secondly, the characteristic parameters can be extracted using principal component analysis. Thirdly, correlation analysis and multiple linear regression analysis are used to find the relationship between the characteristic parameters and the anthropometric parameters. Finally, the significant ones of these anthropometric parameters are chosen as the reference parameters for database matching.

2.1. Principal component analysis on direction transfer function

Principal component analysis (PCA) is a statistical method based on the Gaussian distribution of random variable, which has been applied into the analysis of HRTF [11–13]. In this paper the PCA method is not applied directly on HRTF, but on the direction transfer function (DTF), which is obtained by subtracting the mean of log-magnitude response of HRTF from each log-magnitude response. According to Wightman's report [12], mean HRTF log-magnitude function includes not only the subject-dependent and direction-independent spectral features shared by all HRTFs recorded from an individual ear, but also the measurement artifacts such as the spectral notches caused by standing waves. Therefore, extracting the DTF from HRTF can effectively eliminate the component that is sensitive to the position of the microphone in the standing wave pattern of the ear canal and retained the directional component.

The HRTF data we have used are derived from the CIPIC database, which includes HRIRs of 45 subjects at 25 different azimuths and 50 different elevations [14]. We select 35 subjects and turn their horizontal plane HRIRs into HRTFs as the data source. Each HRTF ranged from 0 to 22.05 kHz and has $N=100$ discrete points.

Suppose the kth azimuth HRTF data from horizontal plane of the ith subject is $\mathbf{H}_{i,k}(f)$, $i=1,2,\ldots,35$, $k=1,2,\ldots,25$, then all HRTFs data can be described as a i^*k column matrix $[\mathbf{H}]_{i,k}$:

$$[\mathbf{H}]_{i,k}=[\mathbf{H}_{1,1}(f)\mathbf{H}_{1,2}(f)\cdots\mathbf{H}_{1,25}(f)\cdots\mathbf{H}_{35,1}(f)\mathbf{H}_{35,2}(f)\cdots\mathbf{H}_{35,25}(f)] \tag{1}$$

To calculate the DTF, we can compute the mean value of log-HRTFs from each direction:

$$\mathbf{H}_{\mathrm{av,log}}(f)=\frac{20}{i^*k}\sum_{i=1}^{35}\sum_{k=1}^{25}lg|\mathbf{H}_i(f)|=\frac{20}{i^*k}\sum_{i=1}^{35}\sum_{k=1}^{25}\begin{bmatrix}|\mathbf{H}_i(\theta_k,f_1)|\\|\mathbf{H}_i(\theta_k,f_2)|\\\vdots\\|\mathbf{H}_i(\theta_k,f_N)|\end{bmatrix} \tag{2}$$

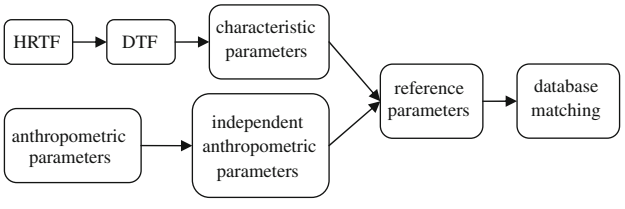

Fig. 1. Workflow of the hybrid algorithm.

Here, $\mathbf{H}_{av,log}(f)$ is the N by 1 matrix. Subtracting the mean value of each log-magnitude response of HRTFs we can obtain the corresponding DTFs, which represent primarily direction-dependent spectral effects:

$$\mathbf{H}_{\Delta}(\theta_k,f_N)=20lg|\mathbf{H}_i(\theta_k,f_N)|-\frac{20}{i^*k}\sum_{i=1}^{35}\sum_{k=1}^{25}lg|\mathbf{H}_i(f)|,\quad i=1,2,\ldots,35,\ k=1,2,\ldots,25 \tag{3}$$

Similarly, DTFs from all M spatial directions can be described as an N by 1 matrix $[\mathbf{H}_{\lambda}]_{N\times M}$. Then we can calculate the covariance matrix $[\mathbf{R}]$:

$$[\mathbf{R}]_{N\times N}=\frac{1}{M}[\mathbf{H}_{\lambda}]_{N\times M}[\mathbf{H}_{\lambda}]^{+}_{M\times N} \tag{4}$$

where $[\mathbf{R}]$ is the N by N Hermite matrix, and its eigenvalue is real. Its eigenvector is extracted and arranged as the eigenvalue reduced-order $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_Q$. Then the front Q eigenvector is taken out as the base-vectors (which are described as PCs) and to build a matrix:

$$[\mathbf{D}]_{N\times Q}=[\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_Q] \tag{5}$$

Because $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_Q$ are orthogonal, by the use of these base-vectors we can decompose $[\mathbf{H}_{\lambda}]_{N\times M}$, and then get the weight matrix accordingly

$$[\mathbf{W}]_{Q\times M}=[\mathbf{D}]^{+}_{Q\times N}[\mathbf{H}_{\lambda}]_{N\times M} \tag{6}$$

Finally, we can predict all of the spatial directions of DTFs as

$$[\hat{\mathbf{H}}_{\lambda}]_{N\times M}=[\mathbf{D}]_{N\times Q}[\mathbf{W}]_{Q\times M} \tag{7}$$

The more the PCs are used, the better the accuracy we can get.

Fig. 2 shows the results of principal component analysis on a left ear's DTF. From the figure we can see that using 8 PCs we can reconstruct more than 90% accuracy of the original DTF. In this paper we use 8 PCs to predict DTF and they can explain more than 90% of the total variability.

2.2. Multiple linear regression analysis

Suppose in the spatial direction θ, the relation between PCs and the corresponding anthropometric measurement is

$$\mathbf{w}_{\theta}=\mathbf{X}\mathbf{B}_{\theta}+\mathbf{E}_{\theta} \tag{8}$$

where $\mathbf{w}_{\theta}$ represents the weight vector of the DTFs in the spatial direction θ; $\mathbf{B}_{\theta}$ is the regression coefficients matrix; $\mathbf{X}$ is the anthropometric measurements matrix; and $\mathbf{E}_{\theta}$ is the estimation errors matrix.

Once we have obtained the anthropometric measurement matrix and the corresponding weight vector, the regression coefficients matrix $\mathbf{B}_{\theta}$ can be predicted as follows:

$$\mathbf{B}_{\theta}=(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{w}_{\theta} \tag{9}$$

As the above equation described, the regression coefficients matrix $\mathbf{B}_{\theta}$ depends on the anthropometric measurements matrix $\mathbf{X}$ and the weight vector $\mathbf{w}_{\theta}$. Since there are usually a number of anthropometric parameters in a measured database, it is obviously unadvisable and unreasonable to introduce all of them into the MLR model because there are different correlations between them and the DTFs. Some useful information might be concealed by the unnecessary

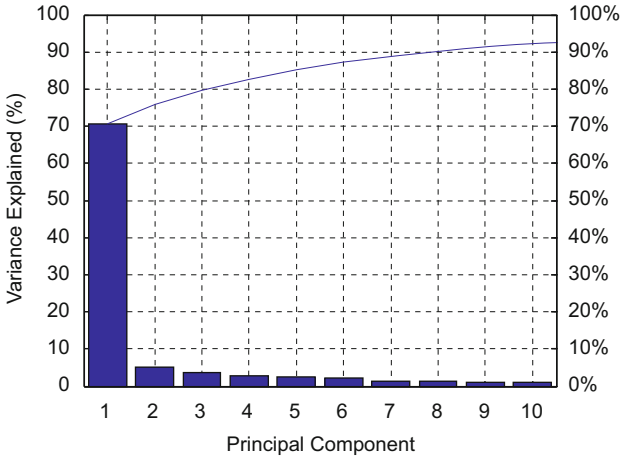

Fig. 2. Results of principal component analysis of DTF.

parameters, which will lead to a worse regression model. Thus, we have to make a choice from them. Furthermore, eliminating the unnecessary anthropometric parameters can directly alleviate the complexity of the system; thus we can alleviate the workload of the individualization process.

2.3. Selection of anthropometric parameters

We can take three steps to select the anthropometric parameters as reference parameters. First, a large number of correlation analyses are done on different anthropometric parameters. For the two different parameters that have large linear correlation coefficients, we can reserve the one that has more significant influence. Then, in order to delete the anthropometric parameters that have smaller correlation coefficient, the correlation analysis is applied to the remaining anthropometric parameters and DTFs. Third, the selected anthropometric parameters are applied into the multiple linear regression model in which F-test and backward selection at significant level $\alpha=0.05$ will be applied to delete the insignificant parameters .

The process of F-test is as follows [15]. Considering the linear regression model

$$\mathbf{Y}=\mathbf{X}\beta+\boldsymbol{\varepsilon} \tag{10}$$

where $\mathbf{X}$ is an n by p full rank matrix of known constant, $\mathbf{Y}$ is an n-vector of response, β is a p-vector of unknown parameter, and $\boldsymbol{\varepsilon}$ is an n by 1 unobservable error with a normal distribution $N(0,\ \sigma^2 I_n)$. We assume that a hypothesis in model (10) is given by H_0:$\mathbf{A}\beta=c$, where $\mathbf{A}$ is a known q by p matrix with rank q and $\mathbf{c}$ is a q by 1 vector. The usual test for hypothesis H_0 is F-test.

F-test is equivalent to the likelihood ratio test. The test statistic F-statistic is given by

$$F=\frac{1}{q\hat{\sigma}^2}(\mathbf{c}-\mathbf{A}\hat{\beta})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{c}-\mathbf{A}\hat{\beta}) \tag{11}$$

where $\hat{\beta}=(X'X)^{-1}X'Y$ is a least-squares estimate of β, and $\sigma^2=\mathbf{Y}'(\mathbf{I}_n-\mathbf{P}_x)\mathbf{Y}/(n-p)$ is an unbiased estimator of σ^2 in model (10), where $\mathbf{P}_x=\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. When H_0 holds the F-statistic equation (11) distributes as an F distribution with degrees of freedom q and $(n-p)$.

When the F-statistic has been calculated, we can use backward selection to select the significant parameters. First, all of the parameters are put in the regression model, then we can calculate F-statistic of each parameter and compare them to the one at the significant level $\alpha=0.05$. After deleting the most insignificant one, we can put the rest into the model again and delete the insignificant one. The process is repeated until all of the rest parameters are significant at the significant level.

After all of the three steps of the selection process, the anthropometric parameters remained can be used as the reference parameters in the following database matching algorithm.

2.4. Database matching

Database matching depends on the similarity of anthropometric structures between the listener and the subject from the database. It aims to select the most appropriate HRTFs from the database for a listener. Matching is performed separately for the left and the right ears, which sometimes leads to the selection of the left and right HRTFs belonging to two different database subjects [16]. Suppose the measured value is $\hat{d}_i$ and the database value is d_i. Then the parameter error is calculated as follows:

$$e_i=\frac{(\hat{d}_i-d_i)}{\mathrm{var}(d_i)} \tag{12}$$

The total error is calculated by Eq. (13) and the subject that corresponds to the minimal total error E is selected as the closest match

$$E=\sum {e_i}^2 \tag{13}$$

In the above matching process, the listeners' anthropometric parameters of ears are measured from digital photographs and other parameters are measured directly by a ruler. Fig. 3 shows the measurement of an ear.

3. Experimental research

3.1. Selection of HRTF

The CIPIC database we used contains HRIRs of 45 subjects and 43 of them have the measured anthropometric parameters, along with some other information about the subjects [17]. The anthropometric information in the database consists of 27 measurements per subject—17 for the head and the torso (x_1–x_{17}) and 10 for the pinna (d_1–d_8, θ_1, θ_2). Considering there may be a relation among the anthropometric structures and spatial directions, we select 35 subjects that have completed anthropometric parameters, using their horizontal plane data as the database for matching.

Fig. 3. Measurement of anthropometric parameters.

Table 1
Processing steps for the selection of HRTF.

Processing procedure	Results
Principal component analysis	Weight vector
Correlation analysis between different anthropometric parameters	d_2, d_3, d_5, d_6, d_8, θ_2, θ_2, $x_1 \sim x_5$, x_7, x_9, x_{12}, x_{14}
Correlation analysis between weight vector and remaining anthropometric parameters	d_2, d_3, d_5, d_6, d_8, $x_1 \sim x_5$, x_7, x_9, x_{12}, x_{14}
Multiple linear regression	d_2, d_3, d_5, d_6, $x_1 \sim x_5$, x_7, x_9, x_{12}, x_{14}
Database matching	Selected HRTFs

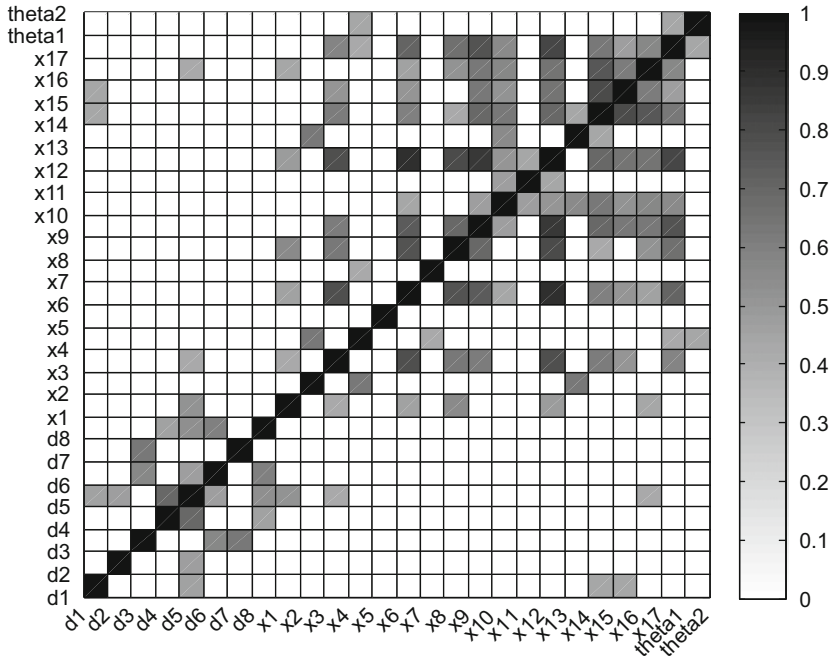

Fig. 4. Correlation coefficients of the anthropometric parameters.

As described in the above sections, we can acquire the reference parameters by PCA, CA and MLR, and then measure the corresponding parameters of the listeners. Using database matching the closest subject's HRTFs can be selected as the individualized HRTFs. The process is shown in Table 1.

Fig. 4 shows the gray image of correlation coefficients larger than 0.6 (theta1, theta2 represent θ_1, θ_2 separately in the figure). From the figure we can find that there are large correlations between some parameters, for example, pinna height d_5 and neck width x_6. Considering the pinna height has more obvious influence on the scattering and reflection of incoming

sound, we delete x_6 to reduce the number of variables. In this way, we can select d_2, d_3, d_5, d_6, d_8, θ_1, θ_2, x_1–x_5, x_7, x_9, x_{12}, x_{14}. After this, the correlation analysis is used for the weight vector and for the remaining anthropometric parameters. Then, the insignificant anthropometric parameters (here, θ_1, θ_2) are deleted. By using the multiple linear regression model we can delete the less significant parameters and choose d_2, d_3, d_5, d_6, x_1–x_5, x_7, x_9, x_{12} and x_{14} (representing cymba concha height, cavum concha width, pinna height, pinna width, head width, head height, head depth, pinna offset down, pinna offset back, neck height, torso top width, shoulder width and height, respectively) as the reference parameters. Finally, we use database matching to select the closest match for each listener and use the closest one's HRTFs as those of the listener.

3.2. Sound localization experiment

When a listener hears a sound filtered by HRTFs measured from his/her own ears, an immersive "virtual acoustic environment" results. The listener feels that the sound appears to originate from well-designed directions in the 3D space surrounding the listener. Thus the individualization accuracy can be evaluated by sound localization experiments. For comparing, we also take the worst match (corresponding to the maximize E in Eq. (13)) of HRTFs for each subject.

Nine subjects aged from 23 to 27 with normal hearing took part in our experiments. The testing sounds are made by convoluting a 0.25 s burst of white noise and the selected HRIRs. During the tests, the sounds are repeated 8 times, with 0.25 s of silence between repetitions. In total 22 target azimuths are selected from the horizontal plane (see Fig. 5). Then the testing sound from a randomly chosen azimuth is played back through the headphone. After this, the subjects are asked to judge the azimuth of the testing sound and write it down by degree on the diagram. When a listener has accomplished this process, a new testing azimuth by the worst matched HRTFs will be selected to repeat the process. The process was repeated until all azimuths by the best and worst matched HRTFs are achieved. The localization experiment continues until both the best and worst matched HRTFs are repeated 9 times. Between repetitions there is at least a 30 min rest.

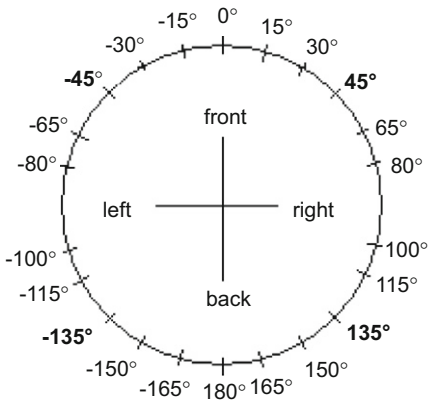

Fig. 5. Twenty-two target azimuths in the horizontal plane.

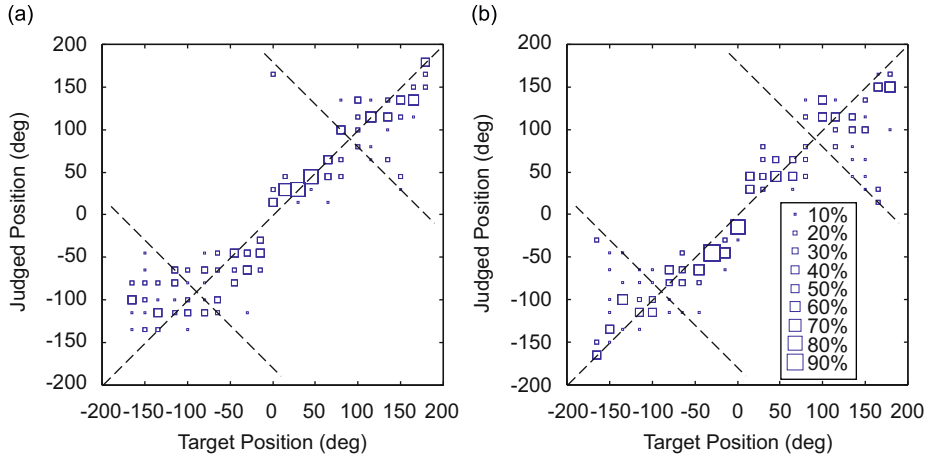

Fig. 6. Localization results of subject 1: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

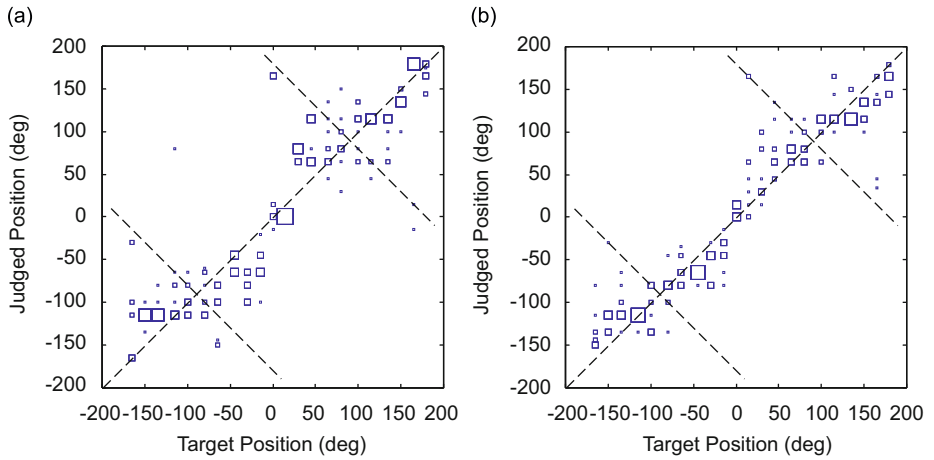

Fig. 7. Localization results of subject 2: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

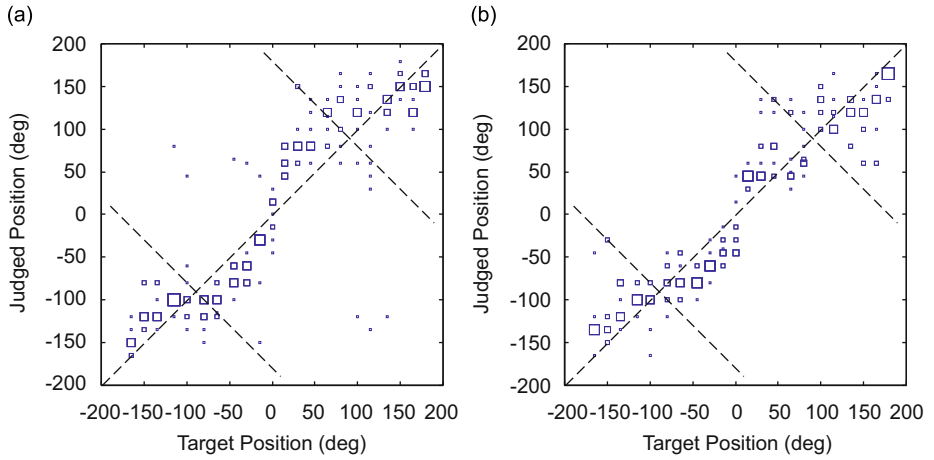

Fig. 8. Localization results of subject 3: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

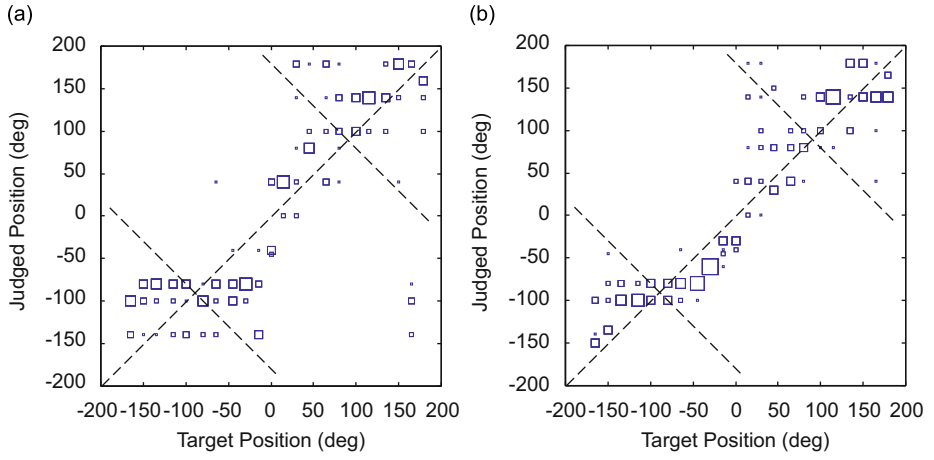

Fig. 9. Localization results of subject 4: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

Figs. 6–14 show the results of the localization experiments of all the 9 subjects by using the worst and best matched HRTFs. Each square in the plot represents the subject's individual location judgment for each target location. As illustrated in the legend, the size of the square represents the number of judgments at that location. For example, the smallest symbol

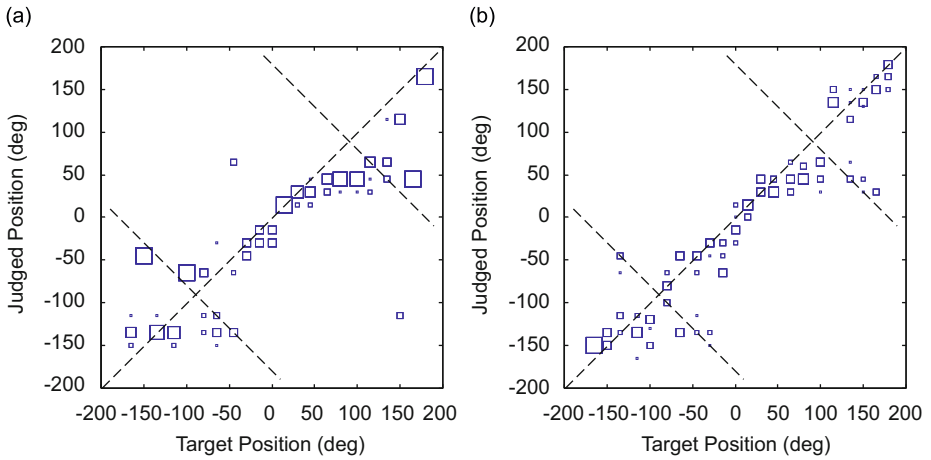

Fig. 10. Localization results of subject 5: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

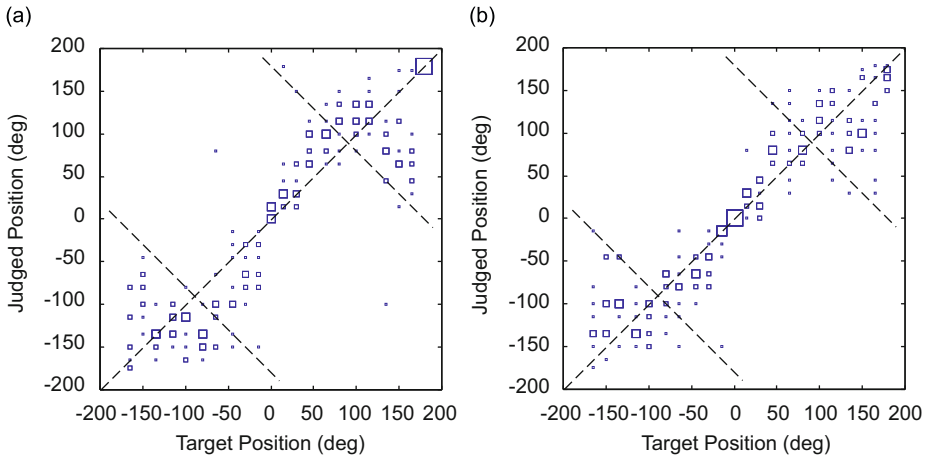

Fig. 11. Localization results of subject 6: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

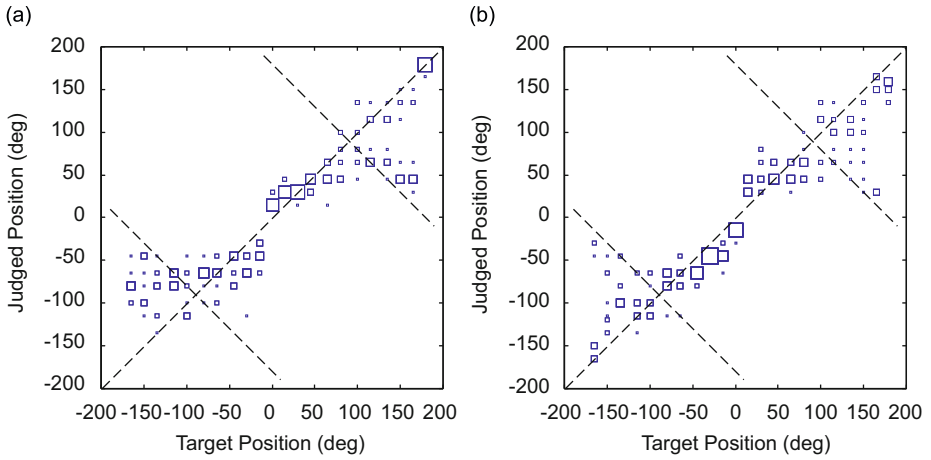

Fig. 12. Localization results of subject 7: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

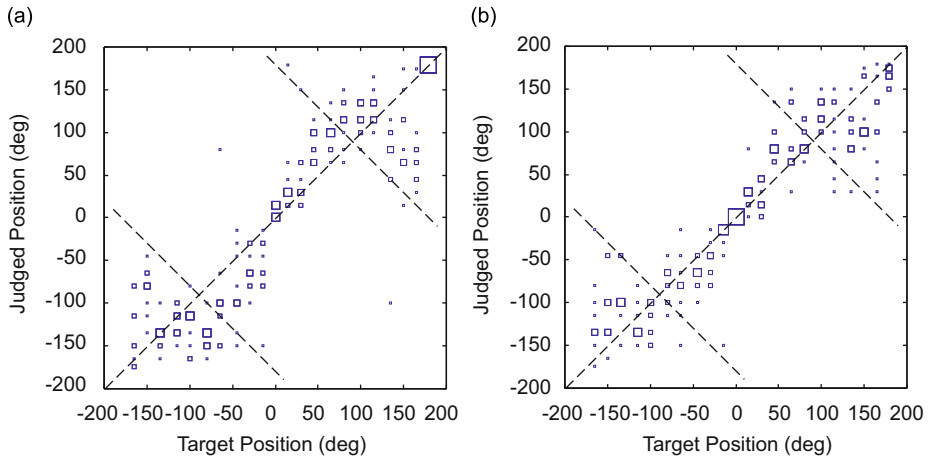

Fig. 13. Localization results of subject 8: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

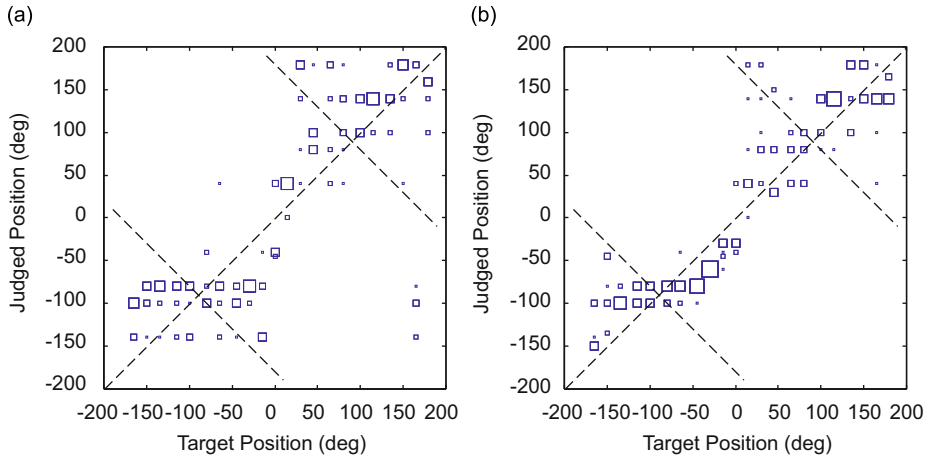

Fig. 14. Localization results of subject 9: (a) using the worst matched HRTFs and (b) using the best matched HRTFs.

in the legend illustrates the case when 10% of the judgments occur in that particular response location. Note that the scale is the same in these plots.

Perfect correlation between target position and response judgment corresponds to a diagonal slope of +1.0 on these graphs, and this means that the subject's judgments are exactly the same as the real directions. On the other hand, two short negative diagonals (slope of -1.0) running from target-response coordinates of -180, 0 to 0, -180 and 0, 180 to 180, 0 (-90, 0 to 0, -90 and 0, 90 to 90, 0 for elevation) correspond to front–back confusions, and the regions around these two lines correspond to the region where the different types of confusions would fall. To facilitate plotting, target and response positions in the left hemisphere are shown as negative numbers.

Unfortunately, we can find some left/right confusions in the figures, which seem abnormal. We have summed the total number of left/right confusions for all the cases, which shows that for best matched HRTFs there are no such confusions and for the worst matched HRTFs the confusion ratio is below 2%. This means the best match is really more reliable than the worst match. We think there might be two possible reasons for the existence of 2% confusions: one is the worst match because the results in the later part (Tables 4–6) have shown that the worst matched HRTFs do not perform as good as those of the best match; the other is the subject's carelessness since each listener had to test 396 spatial sounds in the experiment, which might make some of them feel tired.

As a supplement to the data of Figs. 6–14 and also for comparing, we have calculated the average angle of error, mean of the standard deviation of the error, inverse kappa (κ^{-1}), and the front–back confusion rate as Wightman did [18,19]. The results are listed in Table 2. The average angle of error is the mean of the unsigned angles between each judgment vector and the vector from the origin to the actual (or synthesized) target position. Standard deviation is the mean of the standard deviation of the absolute localization error. The κ^{-1} is a measure of the dispersion of the data. Because the difference between the real direction and that judged by the subject is usually large when the front–back confusion phenomenon

Table 2
Statistical results of the localization experiment.

Listener	HRTF data	Average angle (deg.)	Standard deviation (deg.)	κ^{-1}	Front–back confusion (%)
*subject*1	*worstmatch*	19	22	0.05	21
	bestmatch	18	14	0.03	15
*subject*2	*worstmatch*	23	29	0.09	23
	bestmatch	19	25	0.04	13
*subject*3	*worstmatch*	28	36	0.11	25
	bestmatch	22	15	0.04	16
*subject*4	*worstmatch*	37	46	0.10	33
	bestmatch	24	15	0.06	19
*subject*5	*worstmatch*	23	35	0.06	35
	bestmatch	15	13	0.02	18
*subject*6	*worstmatch*	24	26	0.09	34
	bestmatch	20	17	0.06	19
*subject*7	*worstmatch*	18	17	0.02	32
	bestmatch	18	13	0.02	18
*subject*8	*worstmatch*	24	26	0.09	35
	bestmatch	20	17	0.06	19
*subject*9	*worstmatch*	37	46	0.10	34
	bestmatch	24	15	0.06	19

Table 3
Statistical results for comparing.

Region	HRTF data	Angle of error (deg.)	κ^{-1}	Front-back confusion (%)
Front	Wightman	21.5	0.05	7
	Database matching	21.0	0.05	9
Side	Wightman	15.1	0.03	8
	Database matching	14.7	0.04	22
Back	Wightman	19.7	0.05	2
	Database matching	24.6	0.05	19

appears, similar to the algorithm in Ref. [18], we have corrected the reversals for statistical analysis, and list the front–back confusion out separately in Table 2. This can efficiently reduce the mean absolute error and its standard deviation but with little change of the front–back confusion rate. From the table we can see that at the best match condition the front–back confusion rates are 13–19% and the mean is 17%.

Because the test sounds that Wightman used in Refs. [18,19] also contain the elevation information, we use the middle elevation (elevation 0°, 18°) results of headphone conditions in Ref. [19] as the comparing data. Table 3 lists the mean absolute localization errors, inverse kappa and proportion of reversals for the two algorithms.

From Table 3 we can see that for inverse kappa the two algorithms have no significant difference. Regarding the mean absolute localization error and proportion of reversals, the database matching method we proposed does not seem as good as the Wightmans in Ref. [19] except for the front region. The possible reason is that the HRTF data in Ref. [19] are the measured data of the listener's own while the data we obtained are predicted. For the algorithm in Ref. [18], the front–back confusion rate of the headphone conditions is 29%, and the mean absolute localization errors and inverse kappa for the three regions are about 20–25° and 0.05–0.9, respectively (there is no accurate value in Ref. [18]; the value used here is read from the plot). In this case, the performance of the database matching method we proposed is better.

As for the efficiency of obtaining individualized HRTFs, the time-consuming process of our algorithm is to measure some anthropometric parameters, which will cost about 5–8 min for a listener. This is similar to some other algorithms, but because there is the parameter selection process in our algorithm, it can balance the accuracy and efficiency. Other jobs of the method can all be fulfilled by the computer and only cost several seconds of time. Compared with other individualization methods such as direct measurement and theoretical calculation, this improved database matching method has acceptable accuracy but is much quicker and does not need many experimental facilities. Therefore, if we can get the anthropometric parameters of any listener, they can be applied to the personalized auralization of sound fields in rooms.

Tables 4–6 show the regional averages for the average angle of error, κ^{-1} and standard deviation of error for the 9 subjects. 'Best' represents the best match condition and 'worst' represents the worst match condition. The front region azimuths range from -45 to 45°. The back region represents the azimuth ranging from -135 to -165 and from 135 to 180 degree. The remaining azimuths are contained in the side region. We use the Wilcoxon rank sum test to examine the mean localization error of all subjects at the best match and worst match conditions. The results have shown that at the significant level α=0.05 the mean localization error has a significant difference. As shown in Table 4, the average errors of

Table 4
Regional averages for the average angle of error.

Region	HRTF data	Average angle of error (deg.)								
		Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9
Front	*best*	19.68	20.40	26.42	28.89	11.98	16.35	19.68	16.35	28.89
	worst	21.03	33.65	33.10	35.08	13.97	21.83	16.74	21.83	35.08
Side	*best*	12.43	9.86	15.63	13.75	21.81	16.25	12.43	16.25	13.75
	worst	11.04	13.40	27.36	21.25	22.85	22.43	11.60	22.43	21.25
Back	*best*	21.67	27.14	23.33	30.00	10.95	27.78	22.94	27.78	30.00
	worst	26.98	24.37	24.84	56.98	30.87	28.33	25.79	28.33	56.98

Table 5
Regional averages for the κ^{-1}.

Region	HRTF data	κ^{-1}								
		Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9
Front	*best*	0.02	0.06	0.05	0.08	0.02	0.04	0.02	0.04	0.08
	worst	0.09	0.13	0.13	0.09	0.08	0.06	0.02	0.06	0.09
Side	*best*	0.02	0.03	0.04	0.03	0.02	0.05	0.02	0.05	0.03
	worst	0.03	0.07	0.15	0.10	0.01	0.09	0.03	0.09	0.10
Back	*best*	0.03	0.04	0.03	0.08	0.01	0.10	0.03	0.10	0.08
	worst	0.03	0.07	0.06	0.10	0.09	0.11	0.02	0.11	0.10

Table 6
Regional averages for the standard deviation of error.

Region	HRTF data	Standard deviation of error (deg.)								
		Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9
Front	*best*	15.70	8.44	12.99	8.55	9.40	11.77	8.44	11.77	8.55
	worst	18.20	17.30	19.23	7.61	13.09	15.00	9.38	15.00	7.61
Side	*best*	11.74	9.87	12.38	6.63	9.70	13.58	9.87	13.58	6.63
	worst	17.28	11.99	38.06	18.42	7.81	19.49	11.82	19.49	18.42
Back	*best*	23.68	13.68	13.36	13.78	7.83	16.35	12.25	16.35	13.78
	worst	19.22	12.42	22.47	32.71	20.27	23.18	11.28	23.18	32.71

the best match condition are always smaller than those at the worst match condition, and this means that for the Wilcoxon rank sum test at the significant level α=0.05 the localization performance of best match condition is significantly better than that of the worst match condition.

From Tables 4–6 we can also see that for most subjects, the regional averages for the average angle of error and κ^{-1} have a better performance in the best match condition than that in the worst match condition. For subjects 1 and 7 the performance did not seem to change too much. In some regions (Sub 1 on the side region, Sub 2 on the back region and Sub 7 on the front and side regions), the worst matched results seem better than the best matched results, although the difference is not big. But from Table 2 we can see that even for these subjects, the front–back confusion rates decrease evidently when using the best matched HRTFs.

3.3. Analysis of ITD

Since ITD is an important cue for the horizontal plane localization behavior, we have calculated the ITDs of two arbitrary subjects in the database to test the feasibility of our method. Fig. 15 shows the difference between measured and predicted ITDs of the subjects.

From the figures we can see that for the two subjects, the predicted ITDs at best match condition are closer to the real value. The mean absolute errors of ITDs at best/worst matched condition are 0.0214/0.0582 and 0.0372/0.1127 ms, respectively. The mean absolute errors at best match condition are smaller and are less than 0.1 ms.

According to the above figures and all the data describing the relationship between ITDs and localization directions, we can roughly estimate that the ITD difference of 0.1 ms will lead to the interval of about 15° for source localization.

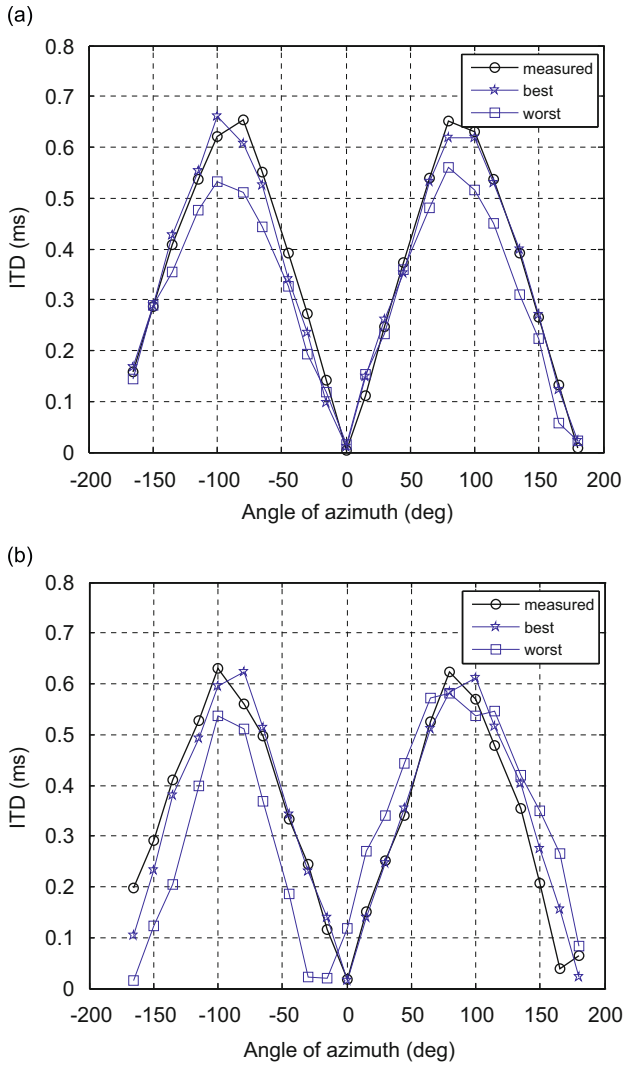

Fig. 15. Measured and predicted ITDs: (a) sub003 and (b) sub065.

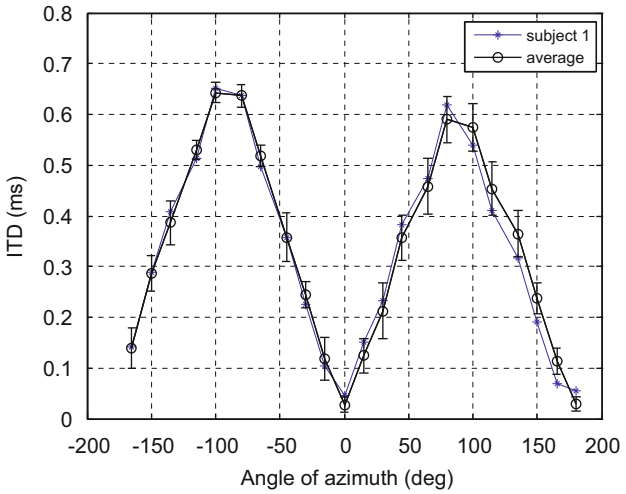

Fig. 16. ITD of the subjects participated in the localization test.

Table 7
Mean absolute error of best/worst ITD.

Subject	Best matched ID		Mean error of ITD (ms)	Worst matched ID		Mean error of ITD (ms)
	Left	Right		Left	Right	
Sub003	Sub044	Sub044	0.0214	Sub018	Sub018	0.0582
Sub010	Sub048	Sub048	0.0215	Sub003	Sub162	0.0616
Sub018	Sub040	Sub156	0.0795	Sub050	Sub050	0.0578
Sub020	Sub058	Sub154	0.0988	Sub126	Sub018	0.1053
Sub027	Sub152	Sub058	0.0577	Sub003	Sub162	0.0401
Sub028	Sub152	Sub027	0.0515	Sub126	Sub135	0.0557
Sub033	Sub59	Sub058	0.0559	Sub126	Sub018	0.0951
Sub040	Sub126	Sub126	0.0398	Sub003	Sub050	0.0664
Sub044	Sub124	Sub147	0.0459	Sub018	Sub018	0.0497
Sub048	Sub061	Sub061	0.0207	Sub028	Sub028	0.0367
Sub050	Sub134	Sub003	0.055	Sub018	Sub018	0.0578
Sub051	Sub134	Sub134	0.0341	Sub018	Sub018	0.0264
Sub058	Sub065	Sub027	0.0966	Sub018	Sub018	0.0441
Sub059	Sub152	Sub033	0.0611	Sub126	Sub148	0.1227
Sub060	Sub061	Sub061	0.0432	Sub050	Sub028	0.0664
Sub061	Sub048	Sub048	0.0207	Sub050	Sub028	0.0863
Sub065	Sub147	Sub147	0.0372	Sub126	Sub018	0.1127
Sub119	Sub048	Sub133	0.0241	Sub028	Sub028	0.0299
Sub124	Sub127	Sub119	0.0475	Sub018	Sub018	0.0341
Sub126	Sub040	Sub040	0.0398	Sub050	Sub050	0.0532
Sub127	Sub134	Sub134	0.029	Sub028	Sub018	0.0383
Sub131	Sub135	Sub135	0.0255	Sub033	Sub028	0.0822
Sub133	Sub119	Sub119	0.0287	Sub018	Sub028	0.0316
Sub134	Sub127	Sub058	0.0571	Sub018	Sub018	0.0469
Sub135	Sub131	Sub131	0.0255	Sub033	Sub028	0.0879
Sub137	Sub119	Sub065	0.0535	Sub033	Sub028	0.097
Sub147	Sub065	Sub134	0.0694	Sub018	Sub018	0.0432
Sub148	Sub048	Sub048	0.0317	Sub033	Sub028	0.0834
Sub152	Sub027	Sub027	0.0381	Sub126	Sub162	0.1073
Sub153	Sub027	Sub033	0.0609	Sub003	Sub044	0.0773
Sub154	Sub027	Sub152	0.0736	Sub003	Sub044	0.0934
Sub155	Sub124	Sub027	0.0348	Sub059	Sub162	0.0781
Sub156	Sub018	Sub010	0.0481	Sub050	Sub050	0.0492
Sub162	Sub020	Sub058	0.0523	Sub018	Sub018	0.0502
Sub163	Sub027	Sub058	0.0808	Sub126	Sub018	0.1029

To consider the relationship of ITDs and localization results, we have calculated the ITDs of all the listeners based on their personalized HRTFs. The mean value of ITDs and that of subject 1 are compared in Fig. 16. For clarity, we also inserted the variance error bar to the mean plot. It can be found that the trend of ITD with the change of angles for different listeners is similar and is consistent with that of the results in Fig. 15.

In Table 7 the best/worst matching results of 35 subjects in the CIPIC database are listed. It can be found that in most cases for both the best match and the worst match, the matched HRTFs of left and right ears belong to different subjects in the database. This is because our algorithm matches the HRTF of the left and right ears, respectively. It can also be found that in most cases different subjects chose different parts of HRTFs. Only 3 subjects chose the same subject (ID 048) at the condition of best match, while 9 subjects chose the same subject (ID 018) at the condition of worst match. From the mean error of ITD, it can be easily found that the mean errors at the best match are usually smaller than those of the worst match and all the mean errors at the best match are less than 0.1 ms, which indicate that the localization errors are smaller than 15°.

4. Conclusions

The similarity of human anthropometric structures makes it possible to individualize HRTF through anthropometric parameters. In this paper we have presented a hybrid database matching method, which combines the method of principal component analysis, correlation analysis and multiple linear regressions. By the use of CIPIC database, we select 13 parameters from the provided 27 parameters and use them as the reference parameters for database matching. The performance of the method has been tested by subjective localization experiments with 9 listeners. Besides comparing with other published results, we also compare the results at the best match condition with those at the worst match condition. The results have shown that in most cases the sound localization accuracy can be enhanced and the front–back confusion rates can be reduced by using the hybrid method.

Like most other prediction methods, our method also needs to measure the subject's anthropometric parameters, but since a number of the parameters have been selected and reduced, the troublesome measurement process can be reduced and can save some time. We will continue to study to what extent the number of parameters can be reduced in order to enhance the practicality of the algorithm.

One of the limitations of the current research is that only the horizontal HRTF data have been used for the individualization and listening tests, and in future work, the individualization method and localization experiments will be applied into the median plane.

Acknowledgements

This project was supported by Natural Science Fund of Shaanxi Province of PRC and Program for New Century Excellent Talents in University. Thanks must go to those who have participated in our measurement and localization experiments. The authors also appreciate the reviewers' constructive suggestions.

References

[1] J.P. Blauert, *Spatial Hearing*, revised edition, MIT, Cambridge, MA, 1997.
[2] J.C. Middlebrooks, Individual differences in external-ear transfer functions reduced by scaling in frequency, *Journal of the Acoustical Society of America* 106 (1999) 1480–1492.
[3] B.F.G. Katz, Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation, *Journal of the Acoustical Society of America* 110 (2001) 2440–2448.
[4] D.N. Zotkin, J. Hwang, R. Duraiswami, L.S. Davis, HRTF personalization using anthropometric measurements, *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, 2003, pp. 157–160.
[5] C. Jin, P. Leong, J. Leung, A. Corderoy, S. Carlile, 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)*, Sydney, Australia, 2000, pp. 235–238.
[6] V.R. Algazi, R.O. Duda, R.P. Morrison, D.M. Thompson, Structural composition and decomposition of HRTFs, *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, 2001, pp. 103–106.
[7] A. Kulkarni, H.S. Colburn, Infinite-impulse-response models of the head-related transfer function, *Journal of the Acoustical Society of America* 115 (2004) 1714–1728.
[8] B.R. Algazi, R.O. Duda, R. Duraiswami, N.A. Gumerov, Z. Tang, Approximating the head-related transfer function using simple geometric models of the head and torso, *Journal of the Acoustical Society of America* 112 (2002) 2053–2064.
[9] S. Fontana, A. Farina, Y. Grenier, A system for rapid measurement and direct customization of head related impulse responses, *Proceedings of the Audio Engineering Society 120th Convention*, Paris, France, 2006, pp. 6851–6870.
[10] X.Y. Zeng, Customization methods of head-related transfer function, *Audio Engineering* 8 (2007) 41–46.
[11] W.L. Martens, Principal components analysis and resynthesis of spectral cues to perceived direction, *Proceedings of the 1987 International Computer Music Conference*, San Francisco, CA, 1987, pp. 274–281.
[12] D.J. Kistler, F.L. Wightman, A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction, *Journal of the Acoustical Society of America* 91 (1992) 1637–1647.
[13] J.C. Middlebrooks, D.M. Green, Observations on a principal components analysis of head-related transfer functions, *Journal of the Acoustical Society of America* 92 (1992) 597–599.
[14] 〈http://interface.cipic.ucdavis.edu/〉 (accessed 5 July 2007).
[15] R.D. Cook, S. Weisberg, *Residual and Influence in Regression*, Chapman & Hall, New York, 1982.
[16] D.N. Zotkin, R. Duraiswami, L.S. Davis, Customizable auditory displays, *Proceedings of the 2002 International Conference on Auditory Displays*, Kyoto, 2002, pp. 1–10.
[17] V.R. Algazi, R.O. Duda, D.M. Thompson, C. Avendano, The CIPIC HRTF database, *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, 2001, pp. 99–102.
[18] E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman, Localization using nonindividualized head-related transfer functions, *Journal of the Acoustical Society of America* 94 (1993) 111–123.
[19] F.L. Wightman, D.J. Kistler, Headphone simulation of free-field listening. II: Psychophysical validation, *Journal of the Acoustical Society of America* 85 (1989) 868–878.