Acoust. Sci. & Tech. **29**, 6 (2008)

ACOUSTICAL LETTER

Improved method to individualize head-related transfer function using anthropometric measurements

Song Xu, Zhizhong Li* and Gavriel Salvendy

Institute of Human Factors & Ergonomics, Department of Industrial Engineering, Tsinghua University, Beijing, 100084, P. R. China

(*Received 25 March 2008, Accepted for publication 12 May 2008*)

Keywords: Individualized HRTF, 3D sound, Principal components analysis, Anthropometry
PACS number: 43.66.Pn, 43.66.Qp [doi:10.1250/ast.29.388]

1. Introduction

Humans have an inherent ability to encode the directional information of an incident sound in an acoustical transfer function. The head, external ears, and torso of a listener can transform the spectral information of a sound as it travels from a sound source to the eardrum [1]. The head-related transfer function (HRTF) is used to describe this physical transformation. However, the HRTF greatly varies from person to person. Several studies have addressed the issue of HRTF individualization [2–7]. It has been suggested that marked perceptual distortion might occur when one listens to sounds spatialized with nonindividualized HRTFs. Individualized HRTFs can increase the number of the cues necessary for spatial localization, particularly for elevation localization.

In this study we aim to improve an existing method of individualizing the HRTF based on listeners' anthropometric measurements. The estimation errors obtained using the existing and improved methods are compared.

2. Method

2.1. Existing method of individualizing HRTF

The procedure of the existing method for HRTF estimation proposed by Kistler and Wightman [8] involves the following steps: (1) standardizing the HRTFs; (2) conducting principal components analysis (PCA) on these standardized HRTFs and detecting the basis functions that can be considered to represent the basic spectral shapes (each spectrum can be approximated by a weighted sum of these basis functions); (3) estimating the HRTFs by reconstruction from these basis functions. Only the magnitude components of the HRTFs are considered in this method. There are two basic assumptions: first, the HRTFs are assumed to be minimum-phase functions; second, it is assumed that the dependence of interaural time delay (ITD) is of no perceptual relevance [8]. This means that the ITD at each source position can be represented by a constant. On this basis, it was found that the PCA scores could be predicted by the multiple regression of some anthropometric measurements [9,10].

2.2. Improved method of individualizing HRTF

Details of the existing method of individualizing HRTFs are referred to in [8–10]. The differences between the improved and existing methods lie in the choice of the basis functions. In the existing method, the basis functions are obtained from the HRTFs at all positions but not from the HRTF at each specified position [8], which consequently might reduce the amount of directional information in the HRTFs. In the improved method, the basis functions at each position of interest are obtained from the HRTFs at the specified position but not from those at all positions. The new method is expected to significantly reduce the estimation errors. The procedure of the improved method is shown in Fig. 1.

After a discrete fast Fourier transform (FFT) is performed on the raw head-related impulse responses (HRIRs) at a specified position (φ, θ), the log-magnitude HRTFs are obtained (φ = azimuth, θ = elevation). Owing to the variation of HRTFs, it is necessary to standardize them before performing PCA. The HRTFs are standardized by subtracting the mean and dividing by the standard deviation. PCA is applied to the standardized HRTFs for each position of interest. The principal component (the PCA scores) of the data can be expressed as linear weighted combinations of the raw standardized HRTFs. The weight coefficients (PCA loadings) are the basis functions mentioned above. The number of principal components can be determined according to the variance explained by the selected components.

The PCA scores, which are considered as the weights of basis functions chosen to predict the HRTF, can be predicted by the regression of some anthropometric measurements. Using the least-squares method, we can construct a multiple regression curve using the PCA scores and corresponding anthropometric measurements. Thus, an individualized standard HRTF can be regenerated from the chosen basis functions and their corresponding weights estimated by the regression.

A case is studied to test the hypothesis that the new method can significantly reduce the estimation errors of HRTFs at a given position (φ, θ). HRTFs are extracted from the CIPIC HRTF database [11], which includes the HRTFs of both left and right ears for 43 human subjects (27 men and 16 women) and 2 mannequins at 1,250 positions. In other words, a total 112,500 HRTFs are recorded in the form of HRIRs for 25 azimuths and 50 elevations. The database is well documented and contains subjects' anthropometric data for 37 measurements. It should be noted that some measurement data for a few subjects were not recorded in the database; the data is complete for only 35 subjects. The subjects' ear canals were blocked when measuring the HRTFs. The HRIRs in the database were compensated to remove the effect of the

*e-mail: zzli@tsinghua.edu.cn

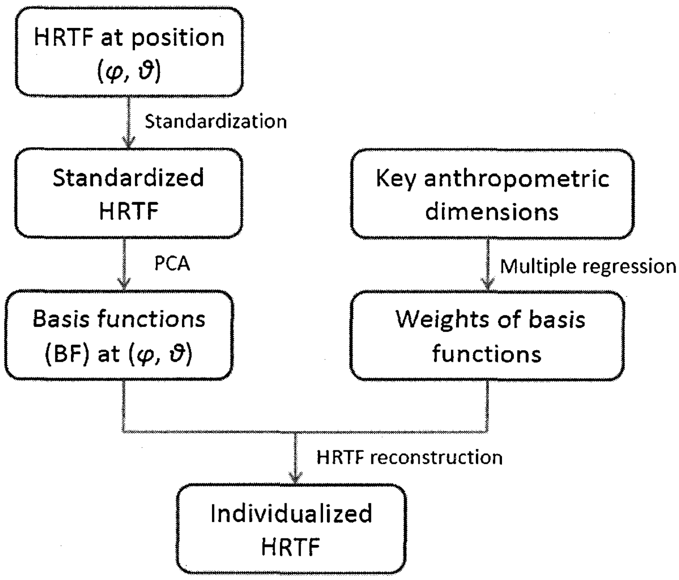

Fig. 1 Improved method of HRTF estimation based on anthropometric measurements.

loudspeakers and microphones. Details on the experiment are referred to in [11].

2.3. Evaluation of HRTF estimation method

There are two different methods for evaluating HRTF estimation. The first is subjective evaluation by sound localization experiments. Some performance indices such as error rate, correct rate, and confusion rate at specified positions are often adopted [12–14]. However, one disadvantage of this method is that it is time-consuming. The second is objective evaluation. In previous studies several objective evaluation tools have been proposed [10,15,16]. Spectral distortion (SD) has been widely used to evaluate speech quality in communication and speech recognition [17,18]. It is also used to evaluate the errors of HRTF interpolation [16] or HRTF estimation based on anthropometric measurements [10]. SD is defined as follows:

$$SD = \sqrt{\frac{1}{N}\sum_{j=1}^{N}[h(f_j) - \hat{h}(f_j)]^2} \text{ [dB]},$$

where $h(f_j)$ and $\hat{h}(f_j)$ are the measured and estimated HRTF log-magnitudes (in dB) respectively, f_j is the jth frequency, and N is the number of frequencies. A previous study on HRTF interpolation on the horizontal plane suggested that an SD of 5.7 dB can satisfy rough sound localization [16]. In this research, SD was adopted to describe the error of HRTF estimation.

3. Results and discussion

In the case study, we compared the HRTF estimation errors (SD) obtained by the two methods at 1,250 available positions. The two methods adopted the same anthropometric measurements (head height, height, neck width, neck height, seated height, shoulder width, pinna offset down, cymba concha height, and cavum concha depth) and 10 principal components were used. The average SDs using the existing and improved methods are 6.2971 and 5.1048 dB respectively.

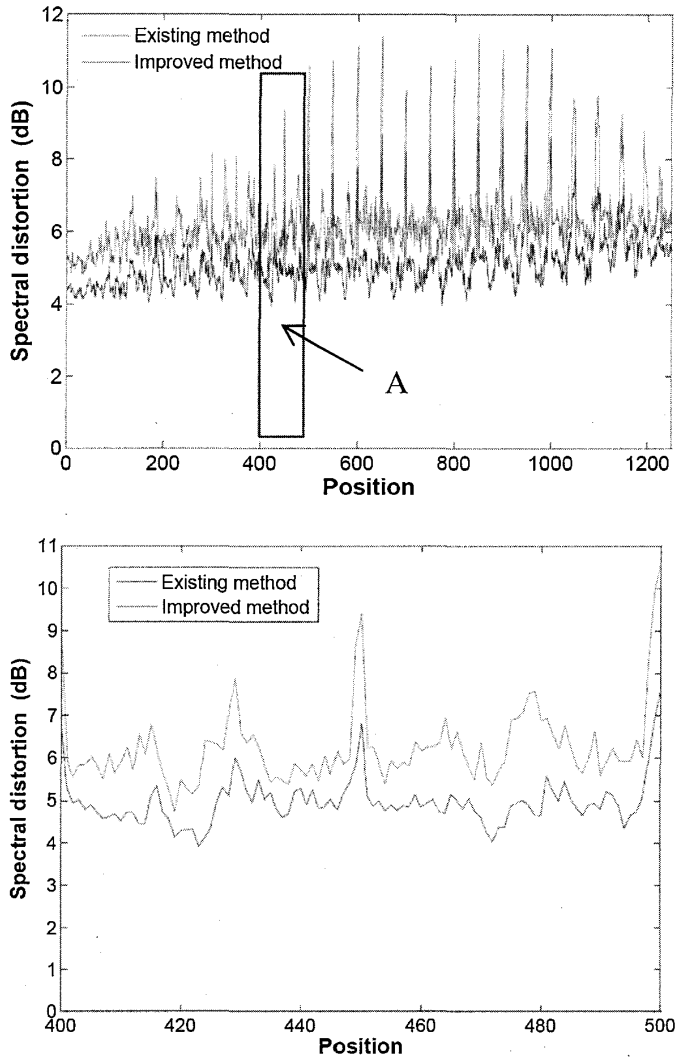

Fig. 2 Mean spectral distortion of HRTF estimation for left ear using two methods.

Using the improved method, a listener's HRTF SD is less than 5.7 dB, which is believed to be sufficiently low for rough localization. The SDs are even less than 5.0 dB at many positions. The results show that the improved method can reduce the SD of HRTF estimation by -1.2% to 39.0% depending on the position, and on average by 18.6% (see Fig. 2).

In order to verify the statistically significant differences of HRTF estimation errors between the two methods, a paired-sample t-test was performed on the SDs obtained by HRTF estimation at ten selected positions: top, front, left, right, back, bottom-front, bottom-back, top-left, top-right, and top-back (see Table 1). It was found that the improved method significantly reduced the SDs at all ten positions ($\alpha = 0.05$). The significant reduction of HRTF estimation errors can be attributed to the use of new basis functions that possess some positional information, while the functions used in the existing method contain no such information. The new method has a greater cost than the existing method because of the need to store the information of more basis functions. However, current computers have sufficient processing ability to deal with this disadvantage.

Table 1 Comparison of HRTF estimation of spectral distortion between the existing and improved methods at ten selected positions (for left ear, $\alpha = 0.05$).

Position*(φ, θ)	Improved method		Existing method		Improved by %	t-test (paired)	
	Mean(dB)	Std.(dB)	Mean(dB)	Std.(dB)		t	p
Top-right (80°, 45°)	5.4513	1.5402	6.1593	1.2666	11.5%	−2.753	0.009
Top-left (−80°, 45°)	4.1187	1.3998	4.8134	1.4163	14.4%	−3.779	0.001
Bottom-back (0°, 225°)	6.9757	3.6321	11.8230	6.3002	41.0%	−4.860	<0.001
Bottom-front (0°, −45°)	5.3469	1.9194	5.9673	1.6146	10.4%	−2.405	0.021
Right (80°, 0°)	5.1847	1.1601	5.8289	1.0396	11.1%	−6.133	<0.001
Left (−80°, 0°)	4.2626	1.3465	4.9123	1.3233	13.2%	−3.745	0.001
Top (0°, 90°)	4.1037	1.2777	6.0062	2.6868	31.7%	−4.417	<0.001
Back (0°, 180°)	4.8039	1.3676	5.5476	1.6004	13.4%	−3.583	0.001
Top-back (0°, 135°)	4.6026	1.6870	6.1376	2.0583	25.0%	−4.153	<0.001
Front (0°, 0°)	5.2226	1.5591	6.1376	2.0583	14.9%	−2.098	0.043

*The coordinate system was defined in reference [11].

4. Conclusions

A new improved method for HRTF estimation based on listeners' anthropometric measurements was developed. The existing and improved methods were compared by a paired-sample t-test. The results suggested that the SDs using the new method were significantly reduced at all the selected positions.

References

[1] J. C. Middlebrooks, "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.*, **106**, 1493–1510 (1999).

[2] E. M. Wenzel, F. L. Wightman, D. J. Kistler and S. H. Foster, "Acoustic origins of individual differences in sound localization behavior," *J. Acoust. Soc. Am.*, **84**, S79 (1988).

[3] E. M. Wenzel, M. Arruda, D. J. Kistler and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, **94**, 111–123 (1993).

[4] H. Møller, M. F. Sørensen, D. Hammershøi and C. B. Jensen, "Head related transfer function of human subjects," *J. Audio Eng. Soc.*, **43**, 300–321 (1995).

[5] H. Møller, C. B. Jensen, D. Hammershøi and M. F. Sørensen, "Using a typical human subject for binaural recording," *Proc. 100th AES Convention* (1996).

[6] H. Møller, M. F. Sørensen and C. B. Jensen, "Binaural technique: do we need individual recordings?" *J. Audio Eng. Soc.*, **44**, 451–469 (1996).

[7] S. Xu, Z. Z. Li and G. Salvendy, "Individualization of head-related transfer function for three-dimensional virtual auditory display: a review," in *LNCS: Virtual Reality* (Springer, Berlin/Heidelberg, 2007), pp. 397–407.

[8] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, **91**, 1637–1647 (1992).

[9] N. Inoue, T. Kimura, T. Nishino, K. Itou and K. Takeda, "Evaluation of HRTFs estimated using physical features," *Acoust. Sci. & Tech.*, **26**, 453–455 (2005).

[10] T. Nishino, N. Inoue, K. Takeda and F. Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Appl. Acoust.*, **68**, 897–908 (2007).

[11] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, "The CIPIC HRTF database," *Proc. IEEE WASPAA 2001*, pp. 99–102 (2001).

[12] F. L. Wightman and D. Kistler, "Multidimensional scaling analysis of head-related transfer functions," *Proc. IEEE WASPAA 1993*, pp. 98–101 (1993).

[13] C. Jin, P. Leong, J. Leung, A. Corderoy and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements," *Proc. 1st IEEE Pacific-Rim Conf. on Multimedia*, pp. 235–238 (2000).

[14] B. U. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," *Proc. ICAD 2003*, pp. 259–262 (2003).

[15] J. Chen, B. D. Van Veen and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Am.*, **97**, 439–452 (1995).

[16] N. Takanori, K. Shoji, T. Kazuya and I. Fumitada, "Interpolation of the head related transfer function on the horizontal plane," *J. Acoust. Soc. Jpn. (J)*, **55**, 91–99 (1999).

[17] N. Nocerino, F. K. Soong, L. R. Rabiner and D. H. Klatt, "Comparative study of several distortion measures for speech recognition," *Speech Commun.*, **4**, 317–331 (1985).

[18] P. A. Laurent, "Expression of spectral distortion using line spectrum frequencies," *IEEE Trans. Acoust. Speech Signal Process.*, **5**, 481–484 (1997).