AN IMPROVED ANTHROPOMETRY-BASED CUSTOMIZATION METHOD OF INDIVIDUAL HEAD-RELATED TRANSFER FUNCTIONS

Xuejie Liu, Xiaoli Zhong

School of Physics and Optoelectronics, South China University of Technology

ABSTRACT

Individual head-related transfer functions (HRTFs) are necessary for rendering authentic spatial perceptions in spatial audio applications. To obtain individual HRTFs while avoiding tedious and complicated measurement and calculation, an improved customization method based on anthropometry matching is proposed. In the method, a set of HRTFs, which is the best match to the pinna shape of the listener using four pinna-related anatomical parameters, is selected as the listener's individual HRTFs from a pre-acquired HRTF baseline database. A series of subject localization experiments was conducted to verify the effectiveness of the proposed method compared with the existing method. Results show that the median-plane localization performance of the customization method proposed in the present work is prior to the existing method, though performance improvement varies with source position.

Index Terms—head-related transfer function, customization, anthropometry matching, median-plane localization

1. INTRODUCTION

The propagation of sound waves from sound source to ears can be regarded as a linear and time-invariant acoustic filtering system. The system transfer function is known as head-related transfer function (HRTF) [1, 2], which represents the interaction between incident sound waves and anatomical structures (such as the head, torso, and pinnae). Due to unique anatomical features of individual, HRTF significantly differs from person to person. One important application of HRTF is virtual sound reproduction, in which a virtual sound source at a specific spatial position can be generated by filtering a mono stimulus with a pair of corresponding HRTFs (for left and right ears) and then reproducing via headphones. Many studies suggest that individual HRTFs of the listener's own are necessary for authentic virtual sound reproduction, otherwise localization errors, such as degraded elevation localization performance and increase of front-back and up-down confusions, may occur[3]. Therefore, customization of listener's individual HRTFs is directly related to the perceptual performance of virtual sound reproduction.

Empirical measurement and computer calculation based on anatomy scanning are two straightforward customization methods of individual HRTFs [e.g. 4, 5]. However, measurement is time-consuming and requires sophisticated apparatus. In calculation, the apparatus for acquiring detailed images of anatomical geometry is expensive, not to mention the intensive computation burden for calculating the HRTFs at full audible frequency range for various source positions. Fortunately, the strong relationship between HRTFs and shapes of human anatomical structures allows for customizing HRTFs by anthropometry-based method other than measurement and calculation. That is, the similarity in anatomy leads to the similarity in HRTFs [6].

Among those HRTF-related anatomical structures, pinna is of most individuality, and determines mid- to high-frequency spectral characteristics of HRTFs [7]. On the basis of matching pinna-related anatomical parameters, Zotkin *et al*. customized mid- to high-frequency individual HRTFs for a new listener by selecting the best match from a pre-acquired HRTF baseline database [8]. This anthropometry-based customization method is simple and easy to be implemented. Experimental results showed that the method improved localization accuracy as well as subjective perception of virtual auditory scene. In Zotkin's method, seven pinna parameters measured from the reference image were used as matching index, neglecting ear rotation and flare angle. Moreover, there exists some correlation among different pinna parameters, so correlation analysis is necessary to find approximately inter-independent pinna parameters. In this way, the number of pinna parameters needed in matching can be reduced, simplifying the anthropometry-based customization method.

This paper proposes an improved anthropometry-based customization method, in which four pinna parameters used for matching were selected by spectral distortion comparison and correlation analysis. A series of subjective localization experiments was conducted to verify the effectiveness of the proposed method.

2. METHODOLOGY

In the present work, the spatial position of a sound source with respect to head center is specified by distance ($0 \le r \le$

$+\infty$), elevation ($-90º \le \phi \le 90º$) and azimuth ($0º \le \theta \le 360º$). $\phi = -90º$, 0º and 90º denote the bottom, horizontal and top direction, respectively. In the horizontal plane, azimuth θ = 0º, 90º, 180º, and 270º denote the front, right, back and left direction, respectively.

HRTFs for the two ears are usually defined as

$$H_L(r,\theta,\phi,f) = \frac{P_L(r,\theta,\phi,f)}{P_0(r,f)}, \quad H_R(r,\theta,\phi,f) = \frac{P_R(r,\theta,\phi,f)}{P_0(r,f)}, \tag{1}$$

where P_L and P_R represent sound pressures at the left and right ears respectively; P_0 represents the sound pressure at the center of the head with the head absent. In the far field with r > 1.0 m, HRTFs are approximately irrelevant to r, hence the variable r is neglected in the following.

2.1. HRTF database and preprocessing

CIPIC HRTF database constructed by Algazi *et al*. in 2001 is adopted here [4]. The database consists of HRTFs (44.1kHz sampling frequency and 200-point length) at 1250 spatial source positions for each of 45 subjects. Moreover, the database also includes measurements of 27 anthropometric parameters — 17 for the head/torso and 10 for the pinna. In the present work, 35 out of 45 subjects with complete 27 anthropometric measurements are used, in which 5 subjects named T_1~T_5 consist of the test group, while the other 30 subjects named B_1~B_{30} consist of the baseline. As to anthropometric parameters, the whole 10 pinna parameters are used as alternatives in the selection procedure to determine which pinna-related parameters are most appropriate in the matching.

It's generally accepted that the pinna determines mid- to high-frequency spectral characteristics of HRTFs which are crucial for mid-plane localization as well as the discrimination of left-right or up-down directions. Therefore, the present work uses the HRTFs in the frequency band of 5kHz ~ 12kHz. Moreover, the HRTF magnitudes are smoothed by a moving frequency window whose bandwidth equals one equivalent-rectangular-bandwidth (ERB) accounting for the frequency resolution of the inner ear [9].

2.2. Parameter selection

In the anthropometry-based customization method of individual HRTFs, selecting appropriate pinna parameters for the matching procedure is crucial. In the present work, the hypothesis is that the similarity of HRTFs is consistent with the similarity of relevant anatomical parameters.

To evaluate the similarity between HRTFs from subjects i and j quantitatively, spectral distortion in the frequency range of [f_l, f_h] is defined as

$$SD\ (\theta,\phi) = \sqrt{\frac{1}{f_h - f_l}\sum_{f_l}^{f_h}\left(20\log_{10}\frac{|H_i(\theta,\phi,f)|}{|H_j(\theta,\phi,f)|}\right)^2}\quad (dB). \tag{2}$$

Moreover, to evaluate the anthropometric similarity of parameter d between subjects i and j, anthropometric deviation E is defined as

$$E = \frac{(d_i - d_j)^2}{\sigma^2}\quad , \tag{3}$$

where σ^2 represents the variance of the anthropometric parameter d among all subjects in the baseline.

The steps of selecting appropriate pinna parameters for the matching procedure are followed as

Step 1, according to Eq. (2), calculate SD between each test subject T_i (i=1, 2,…, 5) and each subject B_j (j=1, 2,…, 30) in the baseline. Then, for each test subject T_i, select the subject with minimal value of SD from the baseline as the most similar subject B_j.

Step 2, according to Eq. (3), calculate E for each of 10 pinna parameters between the test subject T_i and the most similar subject B_j. Then, sort the 10 pinna parameters in ascending order, and those parameters with $E \le 0.1$ are selected as the most relevant pinna parameters for HRTFs.

Step 3, above steps are applied for the left and right ears, respectively. Considering the selected pinna parameters for each ear of each test subject, we choose six pinna parameters which are most relevant to HRTFs, namely, Cavum concha height, Cavum concha width, Pinna height, Pinna width, Pinna rotation angle, Pinna flare angle [4].

Step 4, to find approximately inter-independent pinna parameters and thus reduce the number of pinna parameters in the matching procedure, correlation analysis is applied among the six pinna parameters. Finally, four pinna parameters (namely, Cavum concha height, Pinna height, Pinna rotation angle, Pinna flare angle) with correlation coefficient less than 0.5 are chosen as the matching indexes for HRTF customization.

So far, the improved anthropometry-based customization method of individual HRTFs can be carried out

(1)Measure the four pinna parameters (namely Cavum concha height, Pinna height, Pinna rotation angle, Pinna flare angle) of a new listener out of the baseline;

(2)Calculate the anthropometric deviation E in Eq. (3) for each pinna parameter, then sum up to obtain the total anthropometric deviation E_{total}.

(3)The subject in the baseline with the minimal value of E_{total} is selected as the most similar subject to the new listener. Then, the individual HRTFs of the subject are regarded as the customized HRTFs for the new listener.

In contrast to the Zotkin's work, the present work determines the pinna parameter for matching using the combination of spectral distortion comparison and correlation analysis. Moreover, four pinna parameters are used for matching in the present work, rather than seven pinna parameters in the Zotkin's work [8].

3. EXPERIMENTS AND RESULTS

To validate that the proposed HRTF customization method can effectively improve the virtual localization performance, a series of subject localization experiments was carried out.

Six listeners aging from 20 to 30 with normal hearing participated in the experiment. To measure pinna parameters accurately, we have proposed a measuring method based on extracting anthropometric parameters from scanned 3D head models [10]. First, the 3D head models of the six listeners were scanned by a laser scanner (UNIscan, Creaform Company), then two dedicated softwares (Pro Engineering and Solidworks), were used to extract relevant anthropometric parameters from the scanned head models. Details can be found in Ref. [10].

To compare the localization performance of our proposed method with that of the Zotkin's method, both anthropometry matching methods were implemented on the basis of the pinna parameters extracted from individual 3D head models. Moreover, KEMAR artificial model was constructed using the average anthropometric measurements across certain human populations and HRTFs of KEMAR are widely used in the applications of virtual sound reproduction. So HRTFs of KEMAR were also included in the experiment. Therefore, three types of HRTFs were compared here.

(1) Customized HRTFs using the proposed method in the present work, HRTF_{our};
(2) Customized HRTFs using the Zotkin's method, HRTF_{ref};
(3) Measured HRTFs for KEMAR artificial model, HRTF_{KEMAR}.

Due to the fact that the necessity of individual HRTFs is more urgent in sound source localization in the median plane than other spatial positions, six target virtual source positions in the median plane (θ = 0º, ϕ = −45°, 0°, 22.5°, 45°, 67.5°, and 90°) were tested. Input stimulus was a 10 s white noise. First, the stimulus was filtered with a pair of HRTFs at the target position, synthesizing corresponding binaural signals; Then binaural signals were reproduced through a pair of Sennheiser HD-250 headphone with headphone equalization at a sound pressure level equivalent to a free-field presentation of about 75 dB.

Listeners were asked to judge the perceived virtual source position. At each target position, each listener repeatedly judged six times. The binaural signals for various HRTF types, target positions, and repetitions were rendered in a random order. For each listener, there were 3 HRTF types × 6 target positions × 6 repetitions = 108 judgments. At each experimental condition (HRTF type and target position), there were 6 listeners × 6 repetition = 36 judgments.

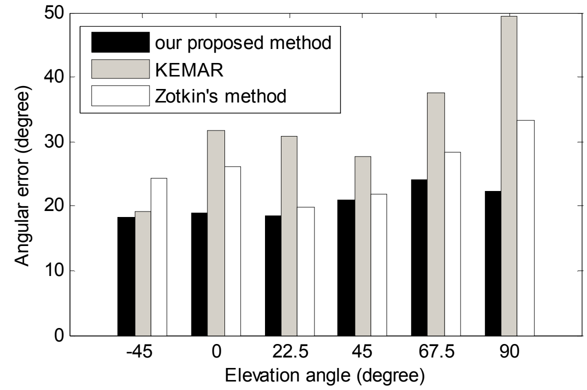

Fig. 1 Elevation angular error of the subjective localization experiment. The dark, grey, and white bars represent the results of our proposed HRTF_{our}, HRTF_{KEMAR}, and Zotkin's HRTF_{ref}, respectively.

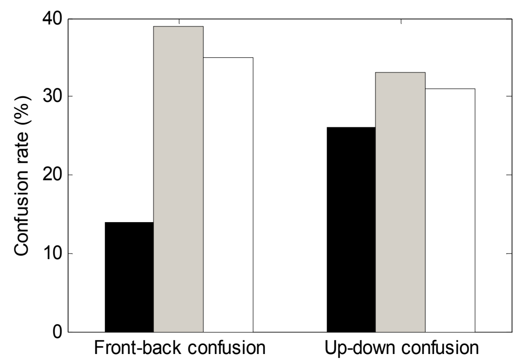

Fig. 2 Front-back and up-down confusion rates of the subjective localization experiment. As in Fig. 1, the dark, grey, and white bars represent the results of our proposed HRTF_{our}, HRTF_{KEMAR}, and Zotkin's HRTF_{ref}, respectively.

A statistical analysis was applied to the localization results of all six subjects. First, results with in-head-localization were excluded from the raw data. Then, if reversal errors (front-back and up-down confusions) appeared, reversal was resolved through spatial reflection prior to analyzing the perceived angular error. Figures 1 and 2 show the mean results of the subjective localization experiment across the six listeners, including elevation angular error, front-back confusion rate, and up-down confusion rate. In the Fig.1, the elevation angular error of HRTF_{KEMAR} is larger than that of either HRTF_{our} or HRTF_{ref} except at elevation −45°, suggesting the localization performance using customized HRTF_{our} or HRTF_{ref} is better than using generic HRTF_{KEMAR}. This is consistent with existing studies [e.g. 3]. Moreover, the localization performance using customized HRTF_{our} is improved in comparison with HRTF_{ref}, though the improvement is less obvious at elevations 22.5° and 45°. As to the two other measures of localization performance (front-back confusion rate and up-down confusion rate), $HRTF_{KEMAR}$ performs worst, while $HRTF_{our}$ is prior to $HRTF_{ref}$. On the whole, the subject localization experiment verifies that (1) virtual sound localization performance can be obviously improved using customized rather than generic HRTFs; (2) the anthropometry-based customization method proposed in the present work is prior to that proposed by Zotkin *et al* in terms of localization accuracy.

4. DISCUSSION AND CONCLUSION

HRTFs vary across individuals, and individual HRTFs of the listener's own is needed to obtain desired localization performance in virtual sound reproduction. The close relationship between HRTFs and listener's anatomical structures (particularly pinnae) makes customize individual HRTFs approximately through anthropometry-based matching possible. In reference [8], Zotkin *et al*. customized individual HRTFs by matching seven pinna parameters of a new listener to a baseline database, based on which we propose an improved customization method by only matching four pinna parameters. Compared with previous relevant studies, the contributions of the present work are

(1) This work selects the pinna parameters used in matching by the combination of spectral distortion analysis and correlation analysis;

(2) In this work, the number of pinna parameters used in matching is only four (Cavum concha height, Pinna height, Pinna rotation angle, Pinna flare angle), rather than seven in existing studies [8].

(3) The four pinna parameters selected in this work include not only pinna size (Cavum concha height, Pinna height) but also pinna position relative to the head surface (Pinna rotation angle, Pinna flare angle). While, the existing studies only consider pinna size.

The results of the present work are applicable to binaural analysis and virtual sound reproduction. From the point of practical uses, procedures for quickly searching for best-matched HRTFs from a baseline are promising and urgently needed. Moreover, increasing the individual samples in the baseline database will improve the statistical results. In addition, the psychoacoustic experiments in the present work are limited in static virtual sound reproduction, in which the dynamic localization cues caused by head motion are omitted. A further work is to incorporate the results of the present work into dynamic virtual sound reproduction.

5. ACKNOWLEDGE

This work was supported by the National Natural Science Foundation of China (11474103) and National Natural Science Foundation of Guangdong Province (2014A030313267).

6. REFERENCES

[1] J. Blauert, *Spatial Hearing* (Revised edition), MIT Press, Cambridge, MA, England, 1997.

[2] B. S. Xie, *Head related transfer function and virtual auditory*. National Defense Industry Press , Beijing, 2008.

[3] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using non-individualized head-related transfer functions," J. Acoust. Soc. Am., 94(1), pp. 111-123, 1993.

[4] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp.99–102, 2001.

[5] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, D. N. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation", J. Acoust. Soc. Am. 127, pp. 370-386, 2010.

[6] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency". J. Acous. Soc. Am., 106(3), pp.1480–1492, 1999.

[7] X. L. Zhong, B. S. Xie. "Spatial Characteristics of Head-Related Transfer Function"，Chinese Physics Letters, 22(5),pp. 1166-1169, 2005.

[8] D. N. Zotkin, J. Hwang, R. Duraiswami, et al. "HRTF personalization using anthropometric measurements", Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 157-160, 2003.

[9] J. Huopaniemi, N. Zacharov. "Objective and subjective evaluation of head-related transfer function filter design," J. Audio Eng. Soc. 47(4), pp. 218–239, 1999.

[10] X. J. Liu, X. L. Zhong. "Extracting anthropometric parameters from a scanned 3D-head-model". In proceeding of 2nd International Conference on Information Science and Control Engineering, ICISCE 2015, Shanghai, China , pp. 225-228, 2015.