inter.noise 2014
MELBOURNE AUSTRALIA
16-19 NOVEMBER

Standardization of Korean head-related transfer function based on tensor-singular value decomposition

Daehyuk SON[1]; Youngjin PARK[2]; Sei-jin JANG[3]

[1] KAIST, South Korea

[2] KAIST, South Korea

[3] NSSC, KETI, South Korea

ABSTRACT

Researches on three-dimensional multimedia has been performed actively in recent years. Virtual sound technology corresponding to virtual image should be provided to implement 3D multimedia with high quality. Head-related transfer function (HRTF) plays a key role in this research area. HRTFs measured in changing azimuth, elevation, and distance for each and every subject is necessary for ideal solution. However, it is practically impossible to measure all subjects' HRTFs, so various HRTF databases have been built by many researchers. Because HRTF displays quite different aspects from subject to subject, HRTF of dummy head has been used for generic usage. However, mannequin's HRTF showed much worse performance comparing with individual case so this solution should be improved. Therefore, this research deals with standardization of HRTF based on tensor-singular value decomposition method and verification with subjective listening test of four subjects. HRTF data used in this paper are extracted from Korean HRTF database which is built by author.

Keywords: Head-related transfer function, virtual sound technology, standardization
I-INCE Classification of Subjects Number(s): 61.7

1. INTRODUCTION

Virtual three dimensional sound technology based on headphones or a number of speakers refer to a system or technology that makes listener hear the sound at any location around head to feel like a virtual source is created and delivered to the listener. To implement and research of virtual three dimensional sound, it is known as head-related transfer function (HRTF) plays an important role. HRTF is an acoustic transfer function between sound pressure of arbitrary sound source and sound pressure measured in front of listener’s ear drum. Ideally, all listeners’ HRTFs are measured to implement ideal virtual sound but it is time-consuming and hard work. In this reason, many researchers tried to develop their own HRTF measuring system and build HRTF database to apply for practical applications (1, 2, 3). However, there is no existing HRTF database for Korean which releases anthropometry information, until now. To supplement this limitation, our laboratory built HRTF database composed of Korean subjects’ HRTFs (4).

There exist various researches to personalize HRTF with different approach. Non-individualized HRTFs measured from a dummy head microphone system are widely used in typical VAD. However, many previous reports showed that non-individualized HRTFs may cause high error rate and low performance in sound localization. In this reason, suggestion of standard HRTF is accomplished in this work. As a standardization method, tensor-singular value decomposition was adopted and HRTFs of 84 Koreans were used. By controlling subject dimension only, a method to extract standard HRTF is suggested. Standard HRTF of Korean is verified by comparing sound source localization performance with individual HRTF for four subjects in upper median plane.

[1] infinitude@kaist.ac.kr
[2] yjpark@kaist.ac.kr
[3] sjjang@keti.re.kr

2. HRTF standardization based on tensor-SVD

2.1 Korean HRTF database

Korean HRTF database which is built by author was used in this paper as a dataset (4). 84 subjects' HRTFs were measured by using our laboratory's HRTF measurement system and measuring space. To reduce occurrence of HRTF distortion caused by subjects' glasses and clothes, wearing glasses were not allowed and clothes were unified to fitted short-sleeved t-shirt. Measurement system adopted vertical polar coordinate and 48 speakers are attached along the median plane. The microphones used to measure are miniature microphones, DPA 4061 High End Binaural Microphones, and closed duct condition was formed by fixing the microphones in the subjects' ears by polymer type ear tips. Also, to reduce HRTF distortion caused by subjects' head movement during measuring, we selected radius 1.5cm circle as acceptable range of head movement and 3° as acceptable range of head rotation. Head fixing device was not used during HRTF measurement because measuring time of our system is about 10 minutes only. In addition, subjects could confirm their head position in real-time through the monitor. It made possible to sustain their concentration and to minimize head movement, simultaneously.

Figure 1 – HRTF measurement system

HRTF was measured for 84 Koreans, 41 males and 43 females, as shown in figure 1. Measurement criteria are in table 1 as follows. Spatial resolution is dense enough as 5 degree for azimuth and elevation. Log sine sweep signal with sweep duration 100ms is adopted as input signal to reduce measuring time. As the result, measuring time per one person is less than 10 minutes which may help increasing subjects' concentration.

Table 1 – Measurement criteria

Features	Specifications
Spatial Resolution	Azimuth 5°, Elevation 5° (Azimuth : -180~175°, Elevation : -30~85°)
Measuring place	Experimental room

Input signal	Log sine sweep signal, sweep duration 100ms
Source distance	1.1581m
Sampling frequency	44100Hz
Measuring time per one person	Less than 10 minutes
Number of measured subjects	84 (41 males and 43 females)

2.2 Tensor-singular value decomposition (SVD)

Tensor-singular value decomposition, so called tensor-SVD method is an extension method of ordinary matrix SVD. A tensor means a higher-order array, a matrix is a second-order tensor and a vector is a first-order tensor, for example. Tensor-SVD is usually used in image processing area because it can handle high-dimensional data like high-quality image (5). Following figure can be helpful to understand tensor-SVD easily which was introduced by Lathauwer et al (6). Figure 2 expresses a decomposition of third-order tensor *A*.

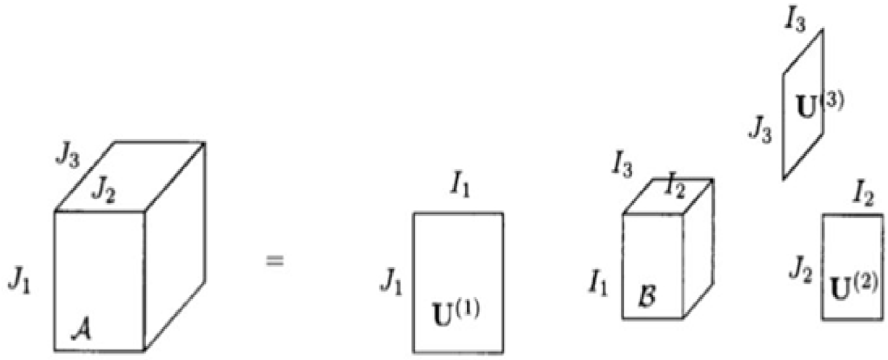

Figure 2 – Mimetic diagram of tensor-SVD of third-order tensor *A*

Tensor *A* can be decomposed with core tensor *B* and factors for each axis $U^{(1)}$, $U^{(2)}$, and $U^{(3)}$. This decomposition process can be represented as an equation (1). Multiplication of *B* with $U^{(1)}$ means that every column of *B* has to be multiplied from the left with $U^{(1)}$. Almost all properties of tensor-SVD can be regarded as extension form of matrix SVD. Specific explanations of tensor-SVD is omitted in this paper.

$$A = B \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \quad (1)$$

In case of Korean HRTF data, subjects' HRIRs which means HRTFs in time domain can be represented as fourth-order tensor having dimension of 670 (time)*72 (azimuth)*24 (elevation)*84 (subjects). The idea of standardization method is reducing dimension of subject axis to one and not controlling other axes.

2.3 Extracted Korean standard HRTF

In this chapter, a method of extracting Korean standard HRTF using tensor-SVD is introduced. As I mentioned earlier, reducing dimension of subject axis to one is a main idea of standardization. Tensor decomposition of Korean HRTFs can be represented like equation (2). In equation, *C* means core tensor which plays a role of standard HRTF. $U^{(t)}$, $U^{(a)}$, $U^{(e)}$, and $U^{(s)}$ represent factors of time, azimuth, elevation, and subject axis respectively. In this paper, $U^{(s)}$ becomes vector form and other factors become one.

$$HRIRs = C \times_t U^{(t)} \times_a U^{(a)} \times_e U^{(e)} \times_s U^{(s)} \quad (2)$$

Before analysis, initial time delay of whole position is removed to reduce unnecessary error caused by different time delay of each point. Magnitude of HRIRs for each subject was normalized to prevent distortion of standard HRTF caused by different magnitude from one subject to others. Core tensor earned by tensor-SVD can be considered as Korean standard HRTF because subject axis is the only tuning target. Analysis is accomplished to male and female HRTFs separately. This is because physical information of male and female subjects are quite different and features of HRTFs showed difference too. Also, magnitude of factor represents how much the subject contributes for making standard HRTF. Standard HRTF for male and female subjects are as figure 3 and 4. 41 Korean males' and 43 Korean females' HRTFs were used as analyzing data. Left side figure represents magnitude spectrum of left ear standard HRTF in median plane from elevation -30 degree to 85 degree. Right side figure represents magnitude of factor for each subject.

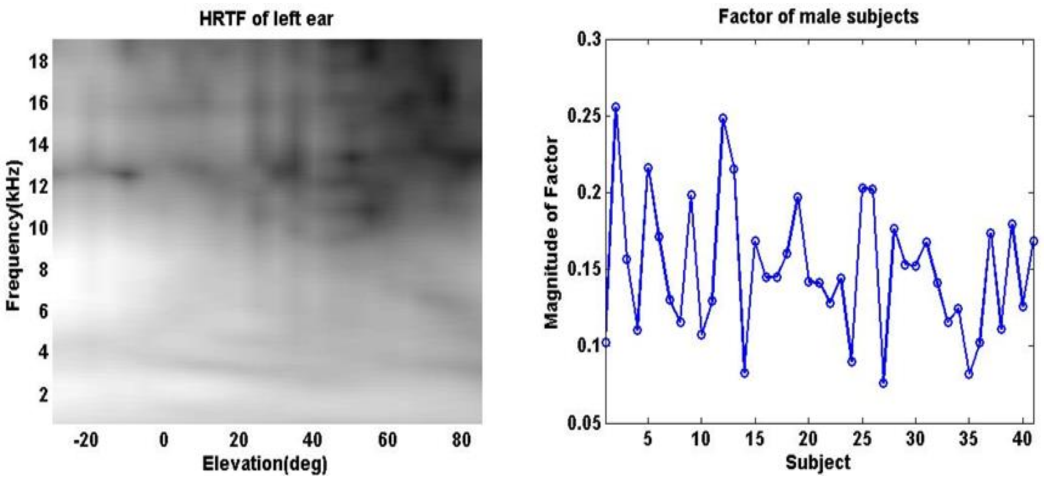

Figure 3 – Left ear standard HRTF of male in median plane (left) and magnitude of factor for subject axis (right)

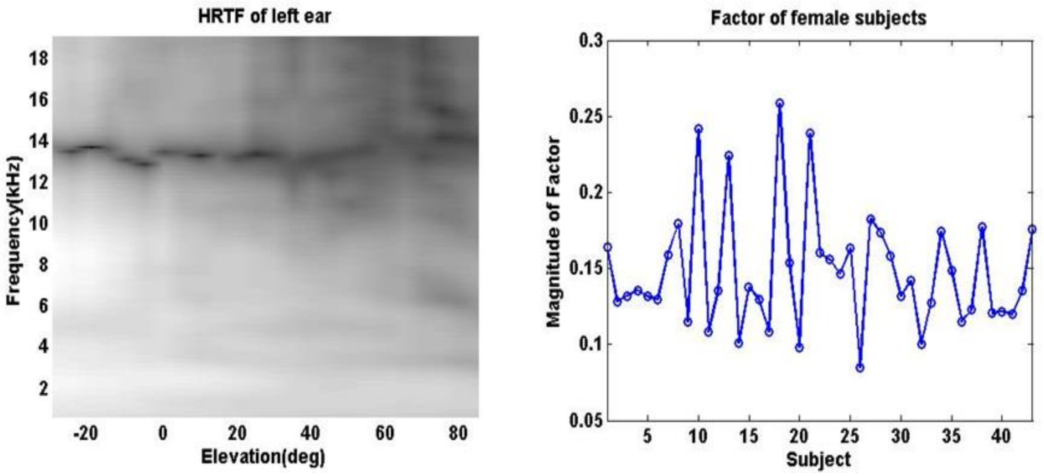

Figure 4 – Left ear standard HRTF of female in median plane (left) and magnitude of factor for subject axis (right)

From figure 3 and 4, first pinnae notch which is known as represented from 6~8 kHz is indistinct. This is because overall smoothing of data has been occurred due to reducing many subjects' HRTFs as one standard HRTF. Also, a deep notch in 12.5 kHz can be observed in Korean male standard HRTF and a deep notch in 13.8 kHz can be observed in Korean female standard HRTF. How this result has an effect on vertical perception will be discussed based on subjective listening test in further chapter. In case of magnitude of factor, each subject has different contribution to forming standard HRTF as predictable result.

3. Experiment

3.1 Subjective listening test

To confirm sound localization performance of fore-earned Korean standard HRTF, subjective listening test was accomplished. Four male subjects were involved in listening test and target angles were elevation 0, 20, 40, 60, 80 degree in upper median plane. Subjects were instructed to localize perceived angle from 0 to 180 degree elevation angle. Five repetition per each target angle was carried out and playing order was randomly decided. Four subjects were named as NH, JH, MJ, and JS. Playing signal was white noise that is inverse-filtered with HRTF. For comparison, virtual sound generated with standard HRTF and individual HRTF were used in listening test. Also, B&K open-ear type headphone was used and speaker dynamics were cancelled before sound generation. The result can be observed below figures 5~8. Left side figures represent individual HRTF case and right side figures represent standard HRTF case. X axis means target angle and y axis expresses perceived angle reported by the subject. Diagonal line represents zero error line where target angle and perceived angle is equal.

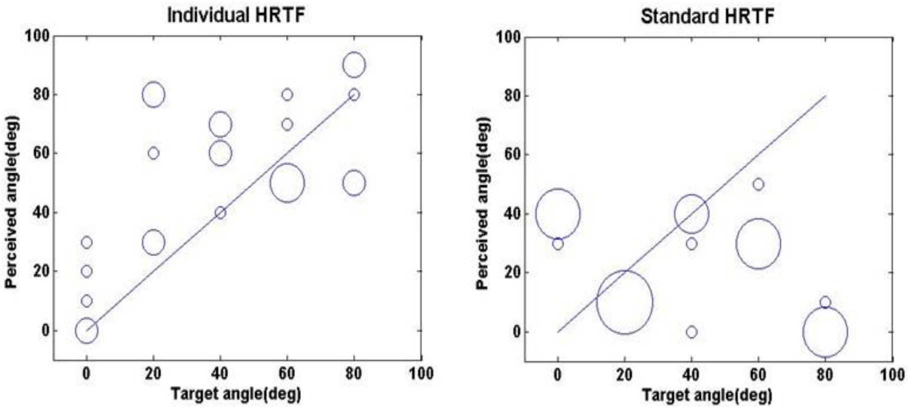

Figure 5 – Listening test result of NH: individual HRTF (left) and standard HRTF (right)

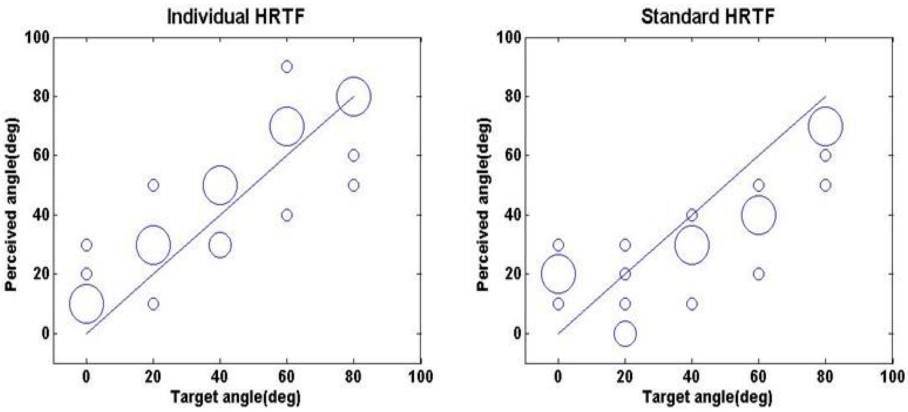

Figure 6 – Listening test result of JH: individual HRTF (left) and standard HRTF (right)

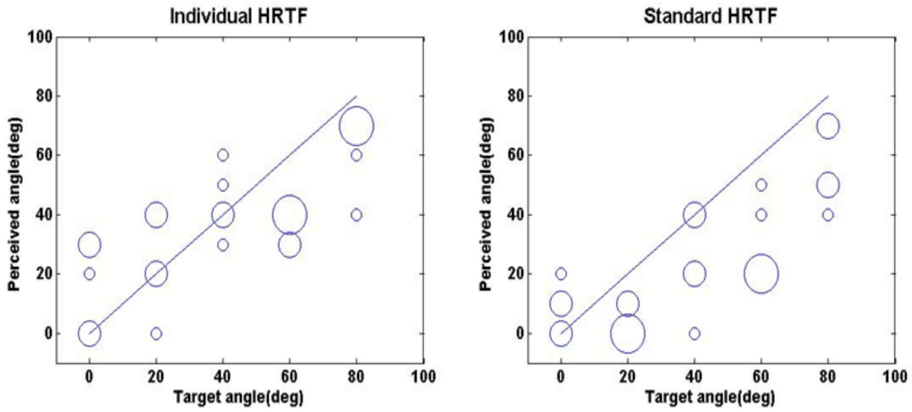

Figure 7 – Listening test result of MJ: individual HRTF (left) and standard HRTF (right)

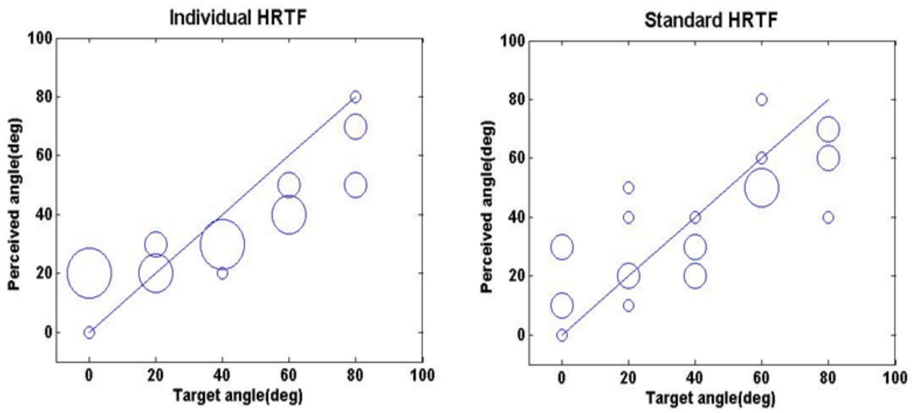

Figure 8 – Listening test result of JS: individual HRTF (left) and standard HRTF (right)

To analyze subjective listening test results quantitatively, absolute mean error was calculated for each subject. Absolute mean error was obtained by averaging difference of absolute value between target angle and perceived angle per each test set. Absolute mean error of each subject with individual HRTF case and standard HRTF case is represented below table 2. The value in bracket means standard deviation. Last column named ‘accuracy of standard HRTF’ means localization accuracy of standard HRTF compared with individual HRTF of that subject and represented by percentage scale.

Table 2 – Absolute means error of subjective listening test

Subjects	Individual HRTFs	Standard HRTFs	Accuracy of standard HRTF
NH	19.2°(16.6)	33.2°(26.6)	57.8%
JH	13.2°(9.00)	16.4°(9.50)	80.5%
MJ	15.2°(11.6)	18.8°(13.6)	80.9%
JS	12.8°(8.90)	14.0°(10.8)	91.4%

Subject NH reported maximum localization error among all subjects in both individual and standard HRTF case. This result means that subject NH has worst vertical perception ability comparing with other three subjects. Accuracy of standard HRTF compared with individual HRTF is 57.8% in case of NH. Subject JS reported minimum localization error among all subjects. Absolute mean error of subject JS in case of individual and standard HRTFs showed below 15 degree error. Also, subject JH and MJ reported quite small localization error in both individual and standard HRTF case. Overall, all subjects except subject NH reported above 80% accuracy of standard HRTF. It means that standard HRTF earned by tensor-SVD of Korean HRTF database showed quite good performance in the view of absolute mean error.

On the other hand, result of all subjects showed worse localization performance when using standard HRTF than individual HRTF. This is because magnitude spectra of standard HRTF represented dim notch patterns compared with individual HRTF. This is predictable result because many subjects' HRTFs were approximated as one basis tensor by tensor-SVD analysis. Although shape of magnitude spectra of individual and standard HRTF were quite different, but sound localization performance of standard HRTF was not much hindered according to subjective listening test result.

4. CONCLUSIONS

This paper proposes a method to extract Korean standard HRTF using tensor-SVD. We have extracted male and female Korean standard HRTF from built HRTF database. After that, subjective listening test for four subjects was implemented to verify sound localization performance of calculated standard HRTF. Test result showed that standard HRTF earned by tensor-SVD of Korean HRTF database displayed quite good performance compared with individual HRTF except one subject. Although magnitude spectra represented different shape, but vertical perception of the subject was not much hindered when we use standard HRTF earned by proposing process.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Knowledge Economy grant funded by the Korea government (No. 10037244) and BK21 Plus project.

REFERENCES

1. V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano. The CIPIC HRTF Database. Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics; 21-24 October 2001; NY, USA 2001. p. 99-102
2. http://recherche.ircam.fr/equipes/salles/listen/
3. P. Majdak, M. J. Goupell and B. Laback. 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. Atten Percept Psychophys. 2010;72(2):454-469
4. D. Son, Y. Park, Y. Park and S. Jang. Building Korean Head-related Transfer Function Database, Transactions of the KSNVE. 2014;24(4):282-288.
5. Tamara G. Kolda and Brett W. Bader. Tensor Decompositions and Applications. SIAM REIVIEW. 2009;51(3):455-500.
6. LIEVEN DE LATHAUWER, BART DE MOOR, AND JOOS VANDEWALLE. A MULTILINEAR SINGUALR VALUE DECOMPOSITION. SIAM J. MATRIX ANAL. APPL. 2000;21(4):1253-1278.