Acoust. Sci. & Tech. **35**, 3 (2014)

TECHNICAL REPORT

Dataset of head-related transfer functions measured with a circular loudspeaker array

Kanji Watanabe[1,*], Yukio Iwaya[2,†], Yôiti Suzuki[3,‡],
Shouichi Takane[1,§] and Sojun Sato[1,¶]

[1]*Faculty of Systems Science and Technology, Akita Prefectural University,*
84–4 Aza Ebinokuchi, Tsuchiya, Yurihonjo, 015–0055 Japan
[2]*Department of Electrical Engineering and Information Technology,*
Faculty of Engineering, Tohoku Gakuin University,
1–13–1 Chuo, Tagajo, 985–8537 Japan
[3]*Research Institute of Electrical Communications and Graduate School of Information Sciences, Tohoku University,*
2–1–1 Katahira, Aoba-ku, Sendai, 980–8577 Japan

(*Received 21 August 2013, Accepted for publication 25 November 2013*)

Abstract: In this paper, we describe a dataset of head-related transfer functions (HRTFs) measured at the Research Institute of Electrical Communications, Tohoku University. The current dataset includes HRTFs for 105 subjects at 72 azimuths × 13 elevations of spherical coordinates. Anthropometric data for 39 subjects are also included. The measurement and postprocessing methods are outlined in this paper. These data will be freely accessible for nonprofit academic purposes via the Internet. Moreover, this dataset will be included in an international joint project to gather several HRTF datasets in a unified data format.

Keywords: HRTF, Dataset, Circular loudspeaker array, Anthropometry, Postprocessing

PACS number: 43.58.–e, 43.58.Vb [doi:10.1250/ast.35.159]

1. INTRODUCTION

The characteristics of head-related transfer functions (HRTFs) are determined by scattering around the listener's external ear, head, shoulder, and torso. These characteristics are known to be perceptual sound localization cues extracted by the auditory system [1]. While the HRTF is a function of source positions, it should differ significantly among individuals owing to anthropometric variety. Therefore, a dataset consisting of HRTFs of many individuals is beneficial to studies on issues such as human localization and the realization of a three-dimensional sound system [2,3].

Several laboratories have developed HRTF datasets [4–8] that include HRTFs measured at many source directions. The number of subjects is usually more than one [5–8]. There is also a dataset that contains subjects' anthropometric data [5]. In general, the measurement of a set of HRTFs takes time and effort owing to the large number of source directions and subjects, and the measurement conditions are often limited. As a result, publicly available HRTF data are still insufficient and the publication of further HRTF datasets will be very useful.

The dataset reported in this paper contains the HRTFs measured at a number of source directions for a number of subjects. The anthropometric data of several subjects are also included. We previously reported a dataset of HRTFs measured at high spatial resolution for 59 subjects [9]. Since that report, we have been accumulating more HRTF data and have adopted postprocessing techniques to improve the quality of measured data, including those reported in [9]. Up to now, we have valid HRTF data for 105 subjects. In this paper, we describe the reconstructed dataset. Moreover, our dataset also provides three-dimensional anthropometric data on structures including the pinna, head, and shoulders for 39 of the subjects. In the rest of this paper, we will describe the methods of measurement

---

[*] e-mail: kwatanabe@akita-pu.ac.jp
[†] e-mail: iwaya@ieee.org
[‡] e-mail: yoh@riec.tohoku.ac.jp
[§] e-mail: takane@akita-pu.ac.jp
[¶] e-mail: sojun_sato@akita-pu.ac.jp

and postprocessing, the evaluation of the measured data, and their potential benefit.

2. MEASUREMENT OF HRTFS

A set of HRTFs for each subject was measured according to the definition described in [1]. By this method, an HRTF for a specified ear for a certain sound source position is obtained using two transfer functions: (1) the transfer function of the sound propagation path from the sound source at a certain direction to the entrance of the subject's ear canal, and (2) the transfer function of the sound propagation path from the same source to the position corresponding to the center of the head with no subject present. The HRTF is calculated as the ratio of transfer function (1) to (2). In this paper, both transfer functions were obtained by measuring the impulse responses and calculating their discrete Fourier transform. The obtained HRTF was stored in the form of its corresponding impulse response, generally called head-related impulse responses (HRIRs).

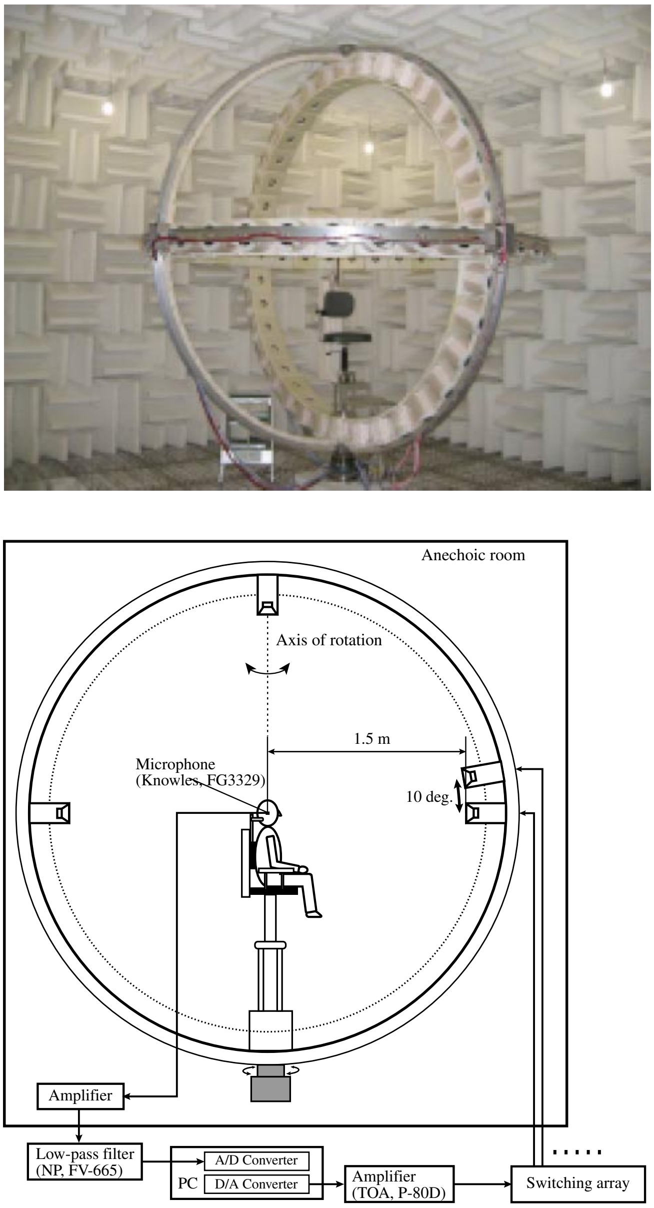

Fig. 1 HRTF measurement system. The vertical circular loudspeaker array consists of 35 loudspeakers. The loudspeakers are arranged at intervals of 10° in elevation except for the angle directly below the subject. The circular loudspeaker array can revolved automatically around the axis of rotation with a resolution of 0.1°.

The optimized Aoshima's time-stretched pulse (OATSP) [10] was used as the source signal to measure the impulse responses. Figure 1 shows the appearance and schema of the HRTF measurement system. The measurement was conducted in the anechoic room of the Research Institute of Electrical Communication, Tohoku University. Although the loudspeakers were arranged both horizontally and vertically, as shown in the upper panel of Fig. 1, a vertical circular loudspeaker array consisting of 35 loudspeakers (FE83E, Fostex) was used, as shown in the lower panel of Fig. 1. The loudspeakers were arranged at intervals of 10° of the elevation angle excluding the angle directly below the subject. The circular loudspeaker array could be automatically rotated around the vertical axis, as shown in Fig. 1, with a resolution of 0.1°. Each subject was seated so that the center of his/her interaural axis was aligned with the center of the circle. The distance between the center of the subject's head and each loudspeaker was 1.5 m, and the head was gently held in place by a small head rest. The subject's ear canals were blocked [11] and miniature microphones (FG3329, Knowles) were placed at the entrance to the blocked ear canal. The OATSP signals had a length of 8192, a resolution of 16 bits, and a sampling frequency of 48 kHz. The interval of source directions was set to 5° azimuth and 10° elevation in spherical coordinates with the origin at the center of the subject's head. It took one hour to one and a half hours to measure HRTFs for one subject, including one or two break(s) of a few minutes each. Although the room temperature was not recorded at the time of measurement, it was controlled so as to be comfortable for the subjects, using an air conditioner during the break. It should be noted that we observed effects of the measurement apparatus on the measurement. One of these was individual loudspeaker variability. Although the loudspeaker characteristics should be canceled out according to the definition of the HRTF, their influence could not be completely removed because the inverse filtering of the loudspeaker characteristics includes errors. Another influence was that of the subject's chair. When the measurement at the center of the head was performed, acoustic absorbing material (nonwoven polyester fabric) was used to reduce this influence as much as possible. However, some disturbances were observed in the

Fig. 2 Example of three-dimensional shading applied to the mesh data. The original data were measured using a three-dimensional scanner (VIVID910, Konica Minolta) and were saved in stereolithography (STL) format consisting of numerous meshes composed of triangulated surfaces.

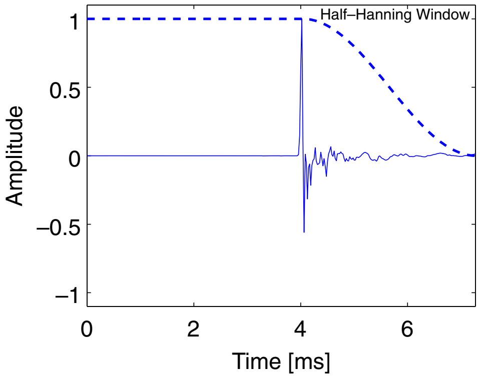

Fig. 3 Example of a measured head-related impulse response (solid line) and the window function applied (dashed line) to eliminate possible reflections.

frequency characteristics of HRTFs, especially for sound sources at elevations lower than −40°.

3. MEASUREMENT OF ANTHROPOMETRY

The anthropometric dimensions of and around the ears, including the pinna, head, shoulders, and torso, are naturally important parameters in the determination of HRTFs, and the frequency characteristics of HRTFs are known to be strongly correlated with the subject's anthropometric dimensions [12,13]. Therefore, three-dimensional images of subjects above and including their shoulders were measured for several subjects whose HRTFs were measured from May 2003 to April 2008. A three-dimensional scanner (VIVID910, Konica Minolta) was used for this measurement. All measured anthropometric dimensions are included in this dataset, in the data format of stereolithography (STL). Figure 2 is an example of the measured three-dimensional image for an artificial head and torso (SAMRAI, Koken) used as the subject.

4. POSTPROCESSING

4.1. Windowing of Measured Impulse Responses

Measured impulse responses contained possible reflections from the circular loudspeaker array, which was made of aluminium covered by sound-absorbing material (glass wool). Therefore, a window function was applied to each impulse response. Since window shape and length substantially affect both the waveform and the frequency characteristics of the windowed signal, proper selection of the window function is important. It is desirable that the window does not affect the main response of the original HRIR and HRTF, and that the window function converges smoothly to zero at its edge. Well-known window functions include the rectangular window, Hamming window, Hanning window, and Blackman window. The rectangular window does not affect the characteristics of the main response in the time domain within the window range. However, if a discontinuity exists at the edge of the window, the frequency characteristics are affected. The other windows are designed so as to reduce the influence of the main response on the frequency characteristics at the expense of waveform distortion. Since the end of the Hamming window does not converge to zero in the time domain, it may affect the frequency characteristics. Considering the balance between the distortion of the main response and the effectiveness of the elimination of the reflection, we selected the Hanning window. In this study, half of the Hanning window was applied from the peak of the impulse response, as shown in Fig. 3. The original length of the Hanning window was set to 350 points. This value was arbitrarily determined for use in checking the measured impulse responses for all source directions. After processing was complete, the length of the impulse response was set at 512 points with zero padding. To compare the influence of windowing on the frequency characteristics, the magnitude characteristics of an HRTF (azimuth, 0°; elevation, 0°; right ear, subject 35) when the Hanning window and rectangular window were applied are shown in Fig. 4. As shown in this figure, ripples in the characteristics seen with the rectangular window become inconspicuous with the Hanning window.

4.2. Additional Calibration of Gains and Frequency Characteristics Due to Loudspeaker Differences

Multiple loudspeakers were used in the measurement, as shown in Fig. 1. As described in Sect. 2, differences among them would be automatically calibrated because the measured impulse responses for a subject were normalized based on the impulse responses of each loudspeaker measured at the position corresponding to the center of the head with no subject present. However, as neither the inverse filtering nor the actual measurement was ideal, we

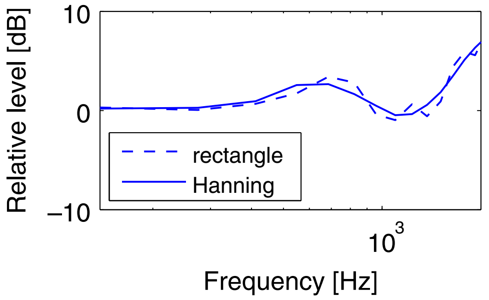

Fig. 4 Effects of window functions. The frequency characteristics of an HRTF in the frequency region below 2 kHz (azimuth, 0°; elevation, 0°; right ear).

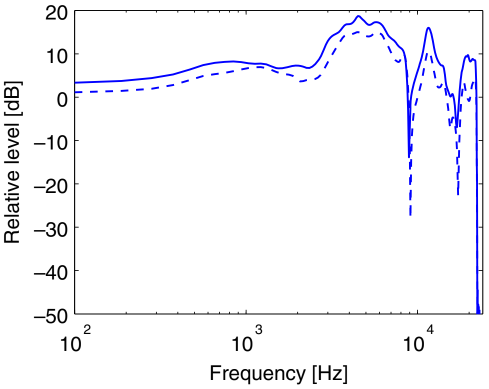

Fig. 5 Example of the effect of the loudspeakers used in the measurements. The two HRTFs shown are of adjacent azimuths (right ear, azimuths, 90° and 95°; elevation, 0°), each measured with a different loudspeaker. The HRTF at 90° was measured by the loudspeaker in front of the circular loudspeaker array while that at 95° was measured by the loudspeaker behind the array.

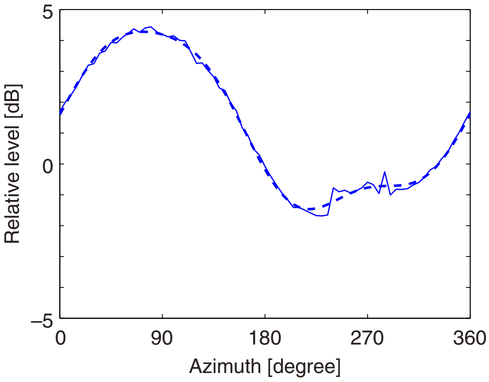

Fig. 6 Example of level change as a function of azimuth at 200 Hz (solid line) and its approximation (dashed line) by superposition of three sinusoidal functions.

observed some systematic coloration depending on the loudspeaker frequency characteristics. That is, to measure HRTFs for a specific elevation angle, two loudspeakers were located at the same elevation in the vertical loudspeaker array (Fig. 1). Both were used as sound sources, from the frontal and rear hemicycle areas, respectively. We found that a small but systematic spatial discontinuity still remained after the specified signal processing described in Sect. 2. This discontinuity seems to be due to the characteristics of each loudspeaker as well as the anterior-posterior asymmetry of the measurement system. Figure 5 shows an example of this artifact. This figure exhibits the frequency characteristics of HRTFs for adjacent azimuths (90° and 95°; elevation, 0°; subject 35), measured using the two loudspeakers located at the same elevation angle but at opposite positions in the circular loudspeaker array. There is an overall gain difference of a few dB between the two HRTFs. Therefore, to calibrate the difference, the overall gains of the rear HRTFs, which were measured using the rear loudspeakers, were adjusted to the same level as the front HRTFs, so that HRTFs varied smoothly as a function of the azimuth for all directions.

4.3. Low-Frequency Compensation

Below several hundred Hz, in general, HRTFs are expected to converge to 0 dB as frequency decreases because the dimensions of the human head and torso are sufficiently smaller than the corresponding wavelengths in this frequency range. However, measured HRTFs sometimes show a displacement from 0 dB in the low-frequency region. Moreover, direct current (DC) offsets are often observed in measured HRTFs [14]. This phenomenon arises because the outputs from the loudspeakers and microphones decay toward DC and lack a DC component. This artifact can be compensated by using the *a priori* knowledge mentioned above, i.e., that HRTFs converge to 0 dB toward 0 Hz.

Figure 6 shows an example of the directional dependence of HRTFs in the low-frequency region (200 Hz) on the horizontal plane. The displacement value from 0 dB systematically changes with source direction and the curve is nearly sinusoidal. Therefore, this directional dependence was approximated by the superposition of three (1st–3rd-order) sinusoidal functions whose basic cycle was 360°. The approximated curve is also shown in Fig. 6 by a dashed line. According to the *a priori* knowledge mentioned above, the magnitude characteristics in the low-frequency region should converge to 0 dB. In this study, the level differences among azimuths were adjusted to be 0 dB by subtracting the gain obtained from the approximated curve. The magnitude characteristics of the microphone used in the measurement declined in frequency regions below around 200 Hz. In high-frequency regions, directional dependence is not sinusoidal on the contralateral side owing to a superposition of two waves with different phases, each of which propagates along two paths: one

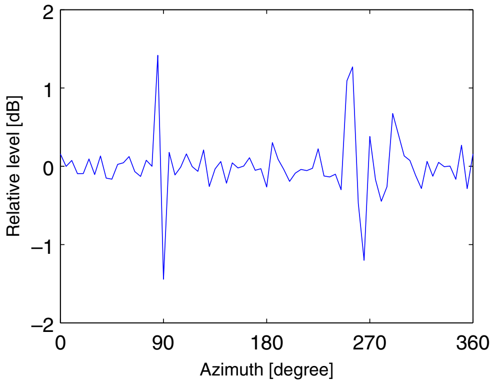

Fig. 7 Difference in overall level between two HRTFs at adjacent azimuths as a function of azimuth (elevation, −40°; left ear, subject 35).

through the front of the head, the other through the back of the head [13]. For these reasons, the value of the approximated curve for 200 Hz was used as the representative value to adjust the level differences.

4.4. Treatment of Measured Data

Figure 7 shows an example of the difference in overall levels between two HRTFs of adjacent azimuths as a function of the source direction azimuth at an elevation of −40°. A notable discontinuity is observed at around 90°. Discontinuities among source directions were corrected with the overall directional dependence obtained from the measured HRTFs. However, large pulsive differences such as that at 90° cannot be fully removed by such a method. Moreover, observation of the measured HRTFs for all subjects revealed that the discontinuity in HRTFs became marked below elevations of −30° for most subjects; this may be because of the influence of the chair. Therefore, we decided to include all data of −30° or above with a cautionary statement.

5. EVALUATION OF HRTFS

As an example of HRTF evaluation, Fig. 8 schematically shows the measured HRIRs of one subject after the postprocessing described in Sect. 4. The abscissa is time, the ordinate is the source azimuth, and the brightness indicates the amplitude of impulse responses. The peak of each HRIR can be seen as a white line, and it is clear that the change as a function of azimuth is sinusoidal. This is because the arrival time for source azimuths depends on the distance from the source to the ear. Figure 9 illustrates the HRTFs for the same subject as in the case of Fig. 8. The abscissa is frequency, the ordinate is the source azimuth, and the brightness indicates the relative level of the magnitude spectrum. In the low-frequency region, the magnitude characteristics seem almost flat and their values converge to 0 dB. This means that the postprocessing compensation we applied for the low-frequency region was very effective for obtaining HRTF data.

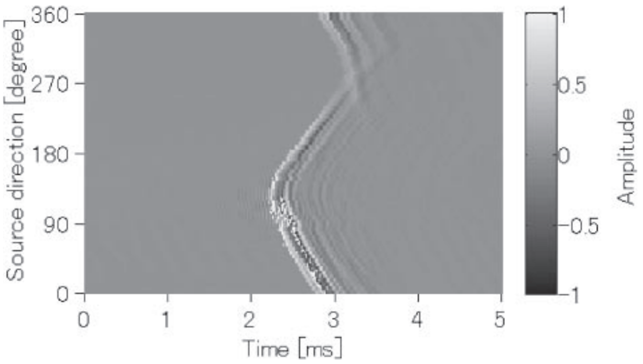

Fig. 8 HRIRs on horizontal plane (left ear).

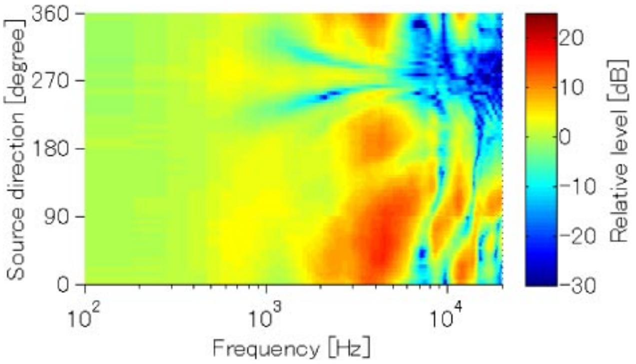

Fig. 9 HRTFs on horizontal plane (left ear).

HRTFs vary according to source direction in a characteristic manner, producing reliable cues that human beings use for localizing sound images [1,2,15–17]. If HRTFs measured at a number of positions are available, directional dependences can be effectively analyzed. As illustrated in Figs. 8 and 9, the variations of the impulse responses and frequency spectra as a function of source direction can be clearly observed. Such visualizations led to the discovery of certain characteristics related to localization cues as well as clarification of the mechanisms underlying the human localization system. In addition, the use of a large number of data points makes it possible to develop various studies for modeling the cues included in HRTFs. For example, Watanabe *et al.* focused their attention on the sinusoidal variation of HRIRs, and modeled the interaural time difference (ITD) as the superposition of various sinusoidal functions [18]. Iida *et al.* investigated HRTF modeling using spectral notches [19].

The availability of a large number of HRTFs can greatly contribute to, for example, studies on the human auditory system and the development of applications utilizing HRTFs. In addition, datasets in which HRTFs are derived from many subjects are particularly valuable because HRTFs vary significantly among individuals. We hope that the dataset described in this paper can contribute substantively in these ways. The dataset is also beneficial because it includes anthropometric data of subjects (39 subjects).

6. SUMMARY

We described a dataset of HRTFs derived from 105 subjects (210 ears). The number of source directions per subject ear was 865; HRTFs were measured at 5° intervals along the azimuth and 10° intervals of elevation from -30 to 90° in spherical coordinates. A subset containing the anthropometric data of 39 subjects was also available. These data will be freely accessible for non-profit academic purposes via the Internet (http://www.riec.tohoku.ac.jp/pub/hrtf/). Moreover, we are participating in an international joint project to gather several HRTF datasets and unify the data format [20]; the present dataset will be included in this international package.

ACKNOWLEDGEMENT

This work was partially supported by JSPS KAKENHI Grant Numbers 24700120, 24240016, and 23300043, and the Consortium R & D Project for Regional Revitalization by METI.

REFERENCES

[1] J. Blauert, *Spatial Hearing* (The MIT Press, Cambridge, Mass., 1983).

[2] M. Morimoto and K. Nomachi, “Binaural disparity cues in median-plane localization,” *J. Acoust. Soc. Jpn.* (*E*), **3**, 99–103 (1982).

[3] F. L. Wightman and D. J. Kistler, “Headphone simulation of free-field listening. II: Psychophysical validation,” *J. Acoust. Soc. Am.*, **85**, 868–878 (1989).

[4] B. Gardner and K. Martin, “HRTF measurements of a KEMAR dummy-head microphone,” http://sound.media.mit.edu/resources/KEMAR.html(ref. 2012.12.26).

[5] 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*, pp. 99–102 (2001).

[6] Human Behavior Signal Processing, Speech, Acoustics, Language, and Music at Nagoya University, “Head Related Transfer Functions,” http://www.sp.m.is.nagoya-u.ac.jp/HRTF/index.html(ref. 2012.12.26).

[7] Yano Laboratory at Nagaoka University of Technology, “HRTF Database,” http://www.nagaoka-ct.ac.jp/ee/lab_syano/index_e.html(ref. 2012.12.26).

[8] S. Takane, D. Arai, T. Miyajima, K. Watanabe, Y. Suzuki and T. Sone, “A database of head-related transfer functions in whole directions on upper hemisphere,” *Acoust. Sci. & Tech.*, **23**, 160–162 (2002).

[9] K. Watanabe, Y. Iwaya, J. Gyoba and Y. Suzuki, “Head-related transfer function corpus for virtual auditory environment,” *Inf. Technol. Lett.*, **4**, 237–240 (2005) (in Japanese).

[10] Y. Suzuki, F. Asano, H. Y. Kim and T. Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” *J. Acoust. Soc. Am.*, **97**, 1119–1123 (1995).

[11] H. Møller, “Fundamentals of binaural technology,” *Appl. Acoust.*, **36**, 172–218 (1991).

[12] E. A. G. Shaw, “Acoustical features of the human external ear,” in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. (Lawrence and Erlbaum Associates, Mahwah, N.J., 1997), pp. 25–47.

[13] K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki and K. Aso, “Estimation of interaural level difference based on anthropometry and its effect on sound localization,” *J. Acoust. Soc. Am.*, **122**, 2832–2841 (2007).

[14] D. Toledo, “Technical and perceptual issues on head-related transfer functions sets for use in binaural synthesis,” Ph.D. thesis, Acoustics, Department of Electronic Systems, Aalborg University (2011).

[15] J. Blauert, “Sound localization in the median plane,” *Acustica*, **22**, 205–213 (1970).

[16] J. Hebrank and D. Wright, “Spectral cues used in the localization of sound sources on the median plane,” *J. Acoust. Soc. Am.*, **56**, 1829–1834 (1974).

[17] F. Asano, Y. Suzuki and T. Sone, “Role of spectral cues in median plane localization,” *J. Acoust. Soc. Am.*, **88**, 159–168 (1990).

[18] K. Watanabe, Y. Iwaya, J. Gyoba, Y. Suzuki and S. Takane, “An investigation on the estimation of interaural time difference based on anthropometric parameters,” *Trans. Virtual Reality Soc. Jpn.*, **10**, 609–617 (2005).

[19] K. Iida, M. Itoh, A. Itagaki and M. Morimoto, “Median plane localization using parametric model of the head-related transfer function based on spectral cues,” *Appl. Acoust.*, **68**, 835–850 (2007).

[20] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger and M. Noisternig, “Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions,” *Proc. 2013 Int. AES Convention*, P12-3 (11 pp.) (2013).

Kanji Watanabe was graduated from Tohoku University in 1999, and received his Ph. D. degree from Tohoku University in 2005. He is currently an assistant professor at the Faculty of Systems Science and Technology, Akita Prefectural University. His research interests include spatial hearing, digital signal processing of acoustic signals. He received the Funai Best Paper Award in 2005. He is a member of the Acoustical Society of Japan, IEICE, VRSJ.

Yukio Iwaya is a professor of faculty of engineering in the Tohoku Gakuin University. He received M. A. in 1993 and Ph. D. 1999 from Tohoku University. His research interest is high-realistic spatial sound field recording and reproduction, perception of spatial hearing. He is a member of the Acoustical Society of Japan, IEICE, ASA, VRSJ, IEEE.

Yôiti Suzuki graduated from Tohoku University in 1976 and received his Ph. D. degree in electrical and communication engineering in 1981. He is currently a professor at Tohoku University. His research interests include psychoacoustics, multimodal perception, and digital signal processing of acoustic signals. He received the Awaya Kiyoshi Award in 1986 and Sato Prize in 1992 and 1994 from the Acoustical Society of Japan. He is a fellow of the Acoustical Society of America.

Shouichi Takane was born in Tokyo, Japan in 1967. He received B. E., M. E., and Ph. D. degrees from Tohoku University in 1989, 1991 and 1998, respectively. He is currently an associate professor at the Faculty of Systems Science and Technology, Akita Prefectural University. His research interests include virtual acoustic display, sound field control, numerical analysis of sound field and acoustic signal processing. He is a member of the Acoustical Society of Japan, the Virtual Reality Society of Japan, the Audio Engineering Society, the Acoustical Society of America, and other societies.

Sojun Sato was graduated from Tokyo Institute of Technology in 1970, and received his Dr. Eng. degree from Tokyo Institute of Technology in 1975. From 1977 to 2001, he joined the Electrotechnical Laboratory, where he researched precise acoustical measurements. From 2001 to 2006, he was the Head of the Acoustics and Vibration Metrology Division, National Metrology Institute of Japan, National Institute of Advanced Industrial Science and Technology. He is now a professor of the Faculty of Systems Science and Technology, Akita Prefectural University. He is a member of ASJ, INCE/J, SICE, IEEE and some other academic societies.