3-D SOUND REPRODUCTION THROUGH STEREO INNER-PHONES

Kazuhiro Iida Junichi Matsuda Kazuhiro Nakamura Motokuni Itoh Masayuki Morimoto

Matsushita Communication Ind. Co., Ltd
Multimedia Solution Laboratories
Saedo, Tsuzuki, Yokohama, 224-8539, Japan
iida.kazuhiro@jp.panasonic.com

Kobe University
Environmental Acoustics Laboratory
Rokko, Nada, Kobe, 657-8501, Japan
mrmt@kobe-u.ac.jp

ABSTRACT

A 3-D sound reproduction method through stereo inner-phones is proposed. This method uses HRTFs and a correction filter, *Hc*, which compensates the changes of the transfer functions of the ear canals caused by insertion of the stereo inside-phones. The results of the localization tests in the horizontal and the median plane indicated that the perceived direction agrees with the simulated one. Furthermore, non-individual HRTFs among subjects are selected by localization tests. The detailed observation of the selected non-individual HRTFs shows that they have emphasized characteristics in directional dependence.

1. INTRODUCTION

The previous 3-D sound reproduction systems have some problems to be settled. One of the most serious problems is that the most previous systems with loudspeakers put limitation on the listening area [1,2]. On the other hand, listeners perceive a sound image inside of their heads when they use headphones or inner-phones. Shimada and Hayashi [3] proposed the specific stereo inner-phones that include microphones in them. In this paper, a new 3-D sound reproduction method, which does not require the specific inner-phones, but uses ordinary inner-phones, is proposed. To reproduce the input signals to both eardrums correctly through ordinary stereo inner-phones, a correction filter, *Hc*, which compensates the changes of the transfer functions of the ear canals caused by insertion of the stereo inner-phones, has been introduced.

Another problem is that the accuracy of sound image localization decreases when HRTF of others is used [4]. Many studies have been made to settle this problem. Middlebrooks [5,6] has proposed the scale factors, which correct the inter-subject differences in HRTF. Møller [7] has obtained the non-individual HRTFs by selection of the typical subject whose HRTFs provide accurate localization to others. In this study, the non-individual HRTFs are obtained for each direction by choosing the subject whose HRTF of the direction provides accurate localization to others.

2. NEW 3-D SOUND REPRODUCTION METHOD THROUGH STEREO INNER-PHONES

2.1. Principle

The sound pressure at the entrance of the ear canal of a subject, *P1*, is described as Eq. (1).

$$P1 = S \times R \times HRTF(EEC) \quad (1)$$

where *S* is a source signal, *R* is a transfer function of a room, *HRTF(EEC)* is a head-related transfer function measured at the entrance of ear canal of a subject.

The sound pressure at the eardrum, *P2*, are described as:

$$P2 = S \times R \times HRTF(ED)$$

$$= S \times R \times HRTF(EEC) \times H(EC) \quad (2)$$

where *HRTF(ED)* is a head-related transfer function measured at the eardrum of a subject, and *H(EC)* is a transfer function in the ear canal (Fig.1).

Let *P3* denote the sound pressure at the eardrum of a subject when *P1* is presented to the subject through stereo inner-phones. *P3* is described as Eq. (3).

$$P3 = P1 \times SIP \times H(EC\_SIP) \quad (3)$$

where *SIP* is a transfer function of stereo inner-phones and *H(EC_SIP)* is a transfer function in the ear canal with the stereo inner-phones inserted.

Let *Hc* denote a compensation filter, which equalize *P3* to *P2*. From Eqs. (1), (2), and (3), *Hc* is expressed as follows;

$$P2 = P3 \times Hc \quad (4)$$

$$Hc = H(EC) / \{SIP \times H(EC\_SIP)\} \quad (5)$$

Therefore, the sound pressure at the eardrum of a subject in the original sound field is reproduced through stereo inner-phones, by multiplying *P1* by *Hc*.

2.2. Calculation of Compensation Filter Hc

In order to obtain *Hc*, *H(EC)* and *H*(*EC_SIP*) were measured using a dummy-head, since it is difficult to measure those of the subject. In this study, the KEMAR dummy-head [8] was adopted, because its transfer function in the ear canal and the

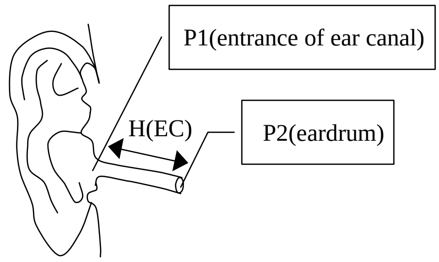

Figure 1. *Sketch of the anatomy of the external ear*

shape of the pinnae are similar to those of human being. *H(EC)* is obtained from *HRTF(ED)* and *HRTF(EEC)* by Eq. (6).

$$H(EC) = HRTF(ED) / HRTF(EEC) \tag{6}$$

HRTF(ED) was measured using the microphones, which were located at the position of the eardrums of the KEMAR dummy-head. *HRTF(EEC)* was measured at the entrance of the ear canal of the KEMAR dummy-head using ear-microphones (Fig. 2). According to the previous studies [9], *H(EC)* is supposed to be independent of the incident angle.

H(*EC_SIP*) was obtained by emitting the M-sequence signal through the stereo inner-phones, which were inserted into the ear canals of the KEMAR dummy-head, and receiving the signal with the microphones located at the position of the eardrums of the KEMAR dummy-head (Fig.3).

Figure 4 shows the obtained *Hc* of various incident angles. The difference of *Hc* due to the incident angle is small up to 15kHz. Each *Hc* has a peak around 2.5kHz and a broad dip around 5kHz. These frequency characteristics infer that the resonance frequency of the ear canal moves from 2.5 kHz to 5kHz by the insertion of the stereo inner-phones.

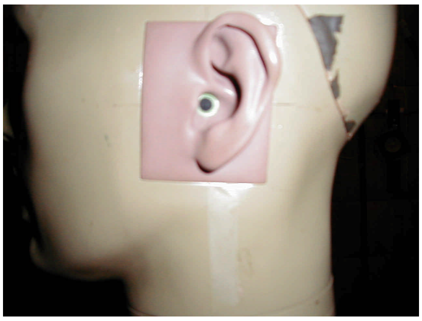

Figure 2. *Placement of microphones for HRTF(EEC) measurements.*

Figure 3. *Placement of stereo inner-phones for H(EC_SIP) measurements.*

Figure 5 shows the obtained *Hc* of various kinds of inner-phones. This figure indicates that these *Hc* have common characteristics up to 12kHz, namely, a peak around 2.5kHz and a dip around 5kHz. The peak level, however, depends on the kinds of inner-phones.

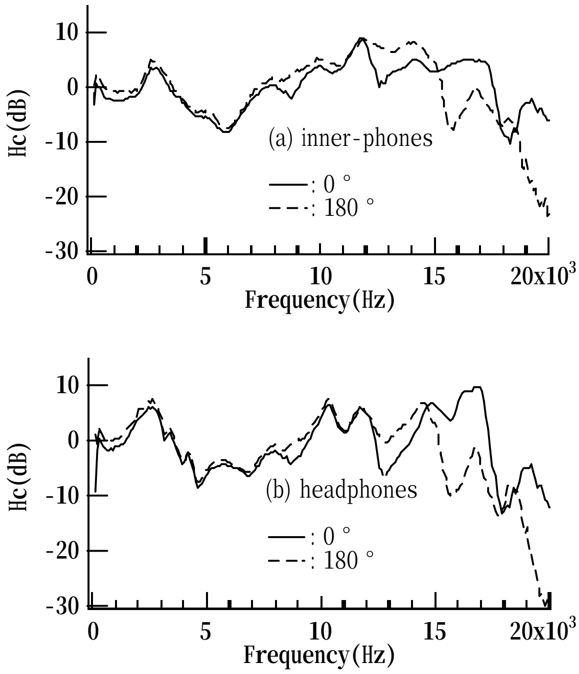

Figure 4. *Examples of the obtained Hc of various incident angles.*

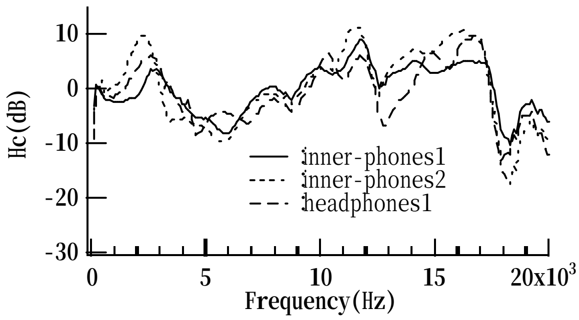

Figure 5. *Examples of the obtained Hc of various kinds of inner-phones.*

2.3. Accuracy of 3-D Localization

Localization tests in the horizontal and the median plane were carried out to confirm the accuracy of 3-D sound reproduction by the proposed method.

2.3.1. Procedure of localization tests

The HRTFs of the subjects in the horizontal plane and the upper median plane were measured at every 30°, using ear-microphones (Fig.6). The ear-microphones were made by the following procedure [10]:

1. The molds of the ear canals of each subject were made.
2. The miniature electret condenser microphones (diameter: 5mm) and silicon resin were put into the molds.
3. The ear-microphones were put into the ear canals of the subject, then the surfaces of the microphones located at the entrances of ear canals (blocked entrances).

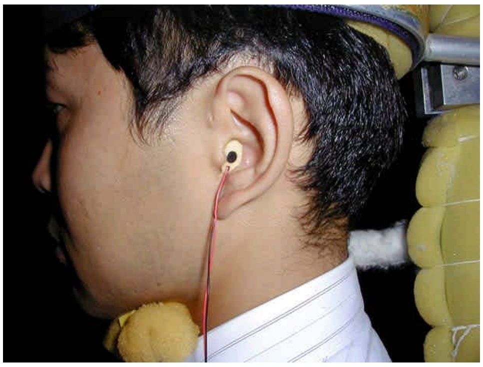

Figure 6. *Placement of the ear-microphones for HRTF measurement.*

The source signal is the wide-band white noise (20Hz - 20kHz). Stimuli were made by convolving the noise with the measured subject's own HRTF, and with *Hc*. In total, nineteen kinds of stimuli (12 azimuths + 7 elevations) were prepared.

Stimuli were divided into two groups. One group consisted of the horizontal plane stimuli, and the other the median plane stimuli. The stimuli were presented through stereo inner-phones (Panasonic, RP-HJ535). Each stimulus was presented 10 times in random order at 70dBA, measured with a coupler in accordance with the IEC318. The tests were carried out in a quiet listening room. The duration of the stimuli was 3s and the interval between two stimuli was 7s. The task of the subjects was to mark down the perceived azimuth and elevation of the sound image on the recording sheet. Subjects were four males with normal hearing sensitivity.

2.3.2. *Results and discussions*

Figure 7 shows one subject's responses for the stimuli in the horizontal plane and those in the median plane.

For the stimuli in the horizontal plane, the perceived azimuth almost agrees with the target azimuth. In case the target azimuth of 0 and 30 degrees, however, a small number of front-back confusions are observed. The perceived elevations were concentrated around 0 degree (in the horizontal plane).

For the stimuli in the median plane, almost of all the responses distribute around a diagonal line. In case the target elevation is 0 degrees, however, the subject sometimes perceived a sound image at the rear. The perceived azimuths were concentrated around 0 degrees (in the median plane).

These results are common in the other subjects. Therefore, it seems proper to conclude that the subjects localize a sound image accurately by the proposed 3-D sound reproduction method through stereo inner-phones.

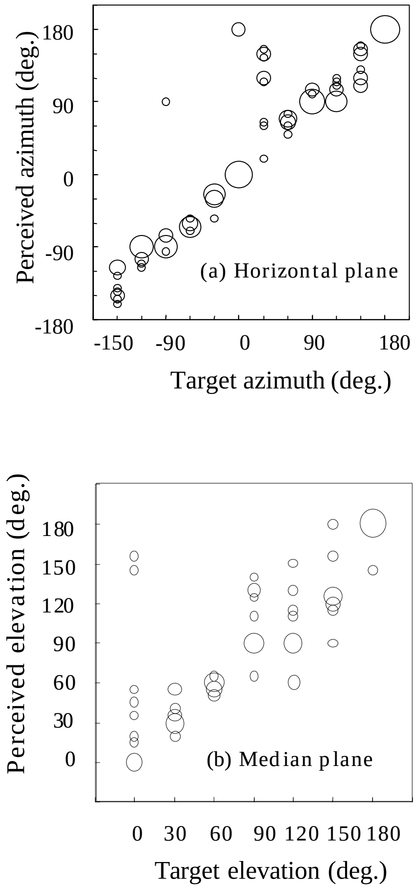

Figure 7. *Examples of the responses to the stimuli; (a) perceived azimuth for the stimuli in the horizontal plane, (b) perceived elevation for the stimuli in the median plane.*

3. NONINDIVIDUALIZATION OF HRTF

The non-individual HRTFs are obtained for each direction by choosing the subject whose HRTF of the direction provides the accurate localization to others.

3.1. Localization Tests to Select Non-individual HRTFs

3.1.1. *Procedure of localization tests*

The HRTFs in the upper median plane were measured at every 30° from the front to the rear by the ear-microphones for nine male. The source signal was a female voice (2s). Stimuli were prepared by convolving the source signal with the median plane HRTF of nine male. Sixty-three kinds of stimuli (7 directions x 9 persons) were prepared. The presentation system of the stimuli consists of DSPs and near-ear loudspeakers (AKG K1000). The DSPs were used for convolution of the source signal and the HRTF. Besides it, the DSPs compensated the transfer functions from the near-ear loudspeakers to the entrances of ear canals of the subjects, and the cross-talks between two ears were negligible small. Each stimulus was presented 5 times in random order at 60dB. The tests were carried out in a quiet listening room. The task of the subjects was to mark down the perceived elevation of a sound image on the recording sheet. Subjects were five males among the nine males, who participated in the HRTF measurement.

3.1.2. Results and discussions

Figure 8 shows one subject's responses to the stimuli, in case that his own HRTFs were used. The responses distribute along a diagonal line. These results are common in other subjects. Therefore, the subjects localized a sound image accurately with the experimental system.

On the other hand, Fig.9 shows the responses of five subjects to all the stimuli, in case that the nine males HRTFs were used. The responses scattered over the figure, and a number of front-back confusions are observed. Theses results show that the accuracy of median plane localization is low when the HRTF are measured from other subjects, as shown in previous studies.

Then, in order to extract the non-individual HRTFs, the mean localization error of each subject for each HRTF was calculated by Eq. (7).

$$e\,(HRTF(ID,S)) = \overline{|\,S_{j,k} - R_{j,k}\,|} \qquad (7)$$

where ID is the participants of HRTF measurement, S is the target elevation, R is the perceived elevation, j is the subjects, and k is the number of trials.

For each direction, the HRTF, whose mean localization error is the minimum is selected among nine participants. Figure 10 shows the responses of the five subjects to the stimuli, in case that the selected non-individual HRTFs. Since these responses distribute along a diagonal line, the HRTFs has robustness against the inter-subject differences for the five subjects.

3.2. A set of nonindividualized HRTF

Figure 11 shows all HRTFs used in the experiments. The bold line denotes the selected non-individual HRTFs and the shaded area denotes the distribution of other HRTFs. These figures show that the peaks and dips of the selected non-individual HRTFs are more sharp than other HRTFs for front (0, 30°) and back (150, 180°). For the upper direction (60, 90, 120°), the behavior of the selected non-individual HRTFs is gentler than others. These results infer that the selected non-individual HRTFs have emphasized characteristics in the directional dependence.

4. CONCLUSIONS

A method of 3-D sound reproduction through stereo inner-phones is proposed. The results of the localization tests in the horizontal and the median plane indicated the validity of the proposed method. Furthermore, non-individual HRTFs among subjects are selected by localization tests. The detailed observation of the selected non-individual HRTFs shows that they have emphasized characteristics in directional dependence.

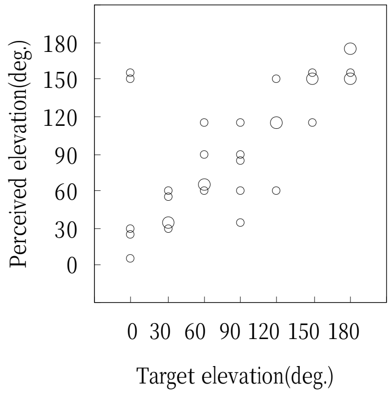

Figure 8. *Examples of responses to the stimuli generated from the subject's own HRTF.*

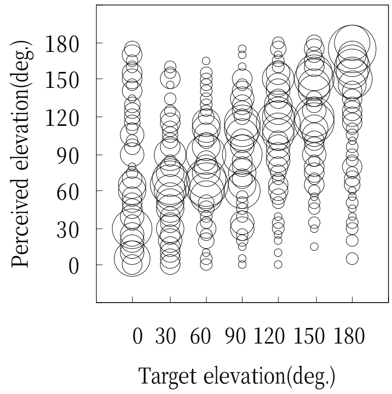

Figure 9. *Responses to the stimuli generated from the subject's own and the other subject's HRTF.*

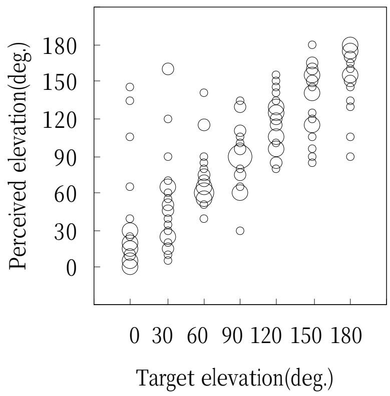

Figure 10. *Examples of responses to the stimuli generated from the selected non-individual HRTF.*

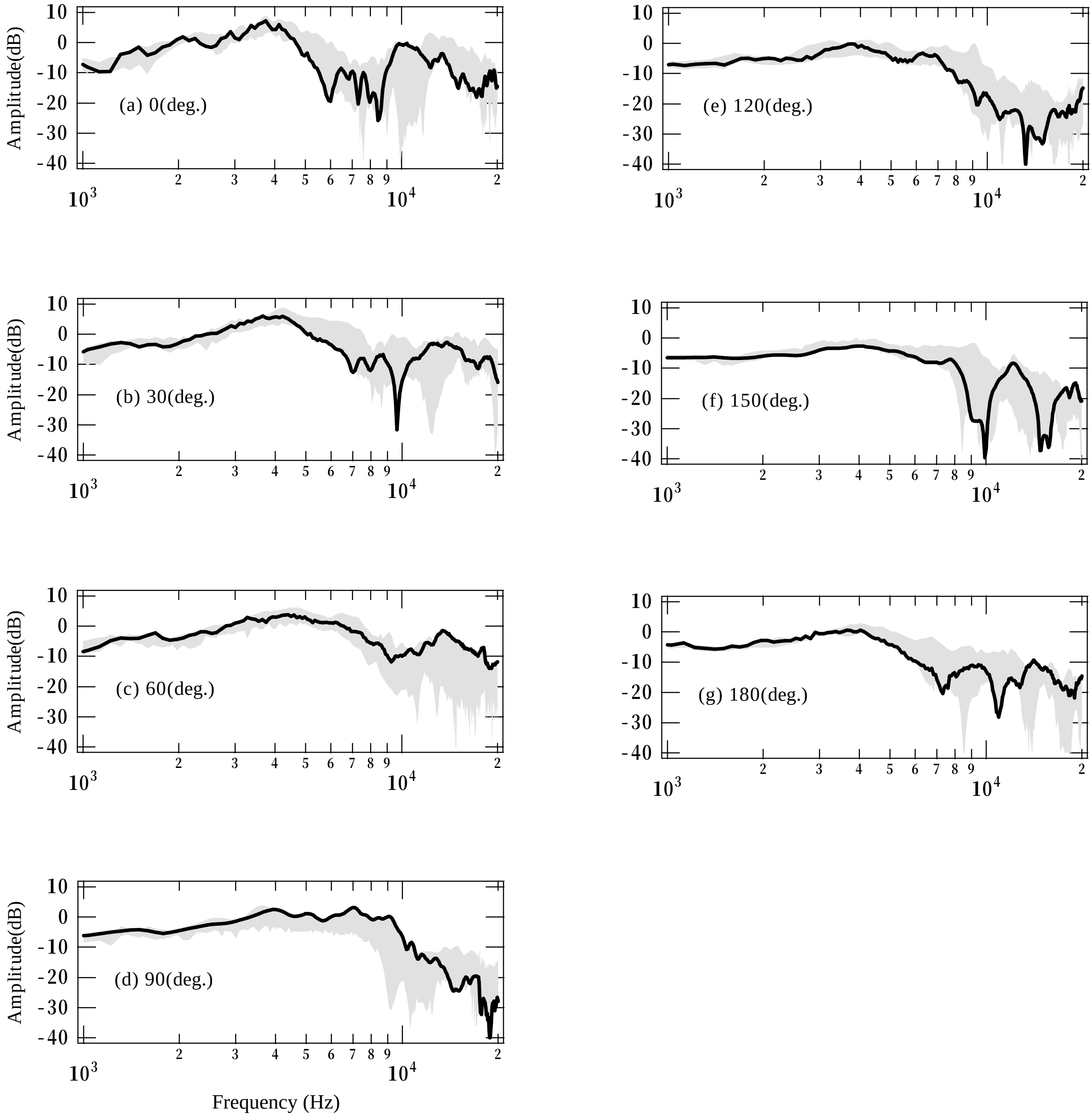

Figure 11. *Non-individual HRTF (left ear) selected by the localization tests. Shaded area denotes the range of HRTF of all the participants*

5. ACKNOWLEDGEMENT

A part of this study is supported by Telecommunications Advancement Organization of Japan, Grant No. JGN-11420.

6. REFERENCES

[1] P. Damaske,"Head-related two-channel stereophony with loudspeker reproduction," *J. Acoust. Soc. Am*., vol. 50, pp. 1109-1115, 1971.

[2] O. Kirkeby, P. A. Nelson, and H. Hamada,"Local sound field reproduction using two closely spaced loudspeakers," *J. Acoust. Soc. Am*., vol. 104, no. 4, pp. 1973-1981, Oct. 1998.

[3] S. Shimada and S. Hayashi," Stereophonic Sound Image Localization System Using Inner Earphones," *FASE*, pp. 157-160, 1992.

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

[5] J. C. Middlebrooks,"Individual difference in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am*., vol. 106, no. 3, pp. 1480-1492, Sep. 1999.

[6] J. C. Middlebrooks,"Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am*., vol. 106, no. 3, pp. 1493-1510, Sep. 1999.

[7] H. Møller , C. B. Jensen, D. Hammershøi, and F. Sørensen,"Selection of a typical human subject for binaural recording," *ACUSTICA*, vol. 82, S215, 1996.

[8] M. Burkhard and R. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.*, vol. 58, pp. 214-222, 1975.

[9] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts, 1983.

[10] K. Iida, E. Rin, M. Itoh, and M. Morimoto,"Role of spectra of input signals to both ears in median plane localization," in Proc. Mtg. Acoust. Soc. Jpn., pp. 295-296, Sep. 2000.