A method of 3-D sound image localization with externalization through headphones using a new correction filter of headphone-to-eardrum transfer functions

Kazuhiro Iida *Atsunobu Murase* *Masayuki Morimoto*

AV Core Technology Development Center
Matsushita Electric Industrial Co., Ltd
iida.kazuhiro@jp.panasonic.com

Environmental Acoustics Lab.
Fac. of Eng., Kobe University

Abstract

In this paper, a new headphone-based 3-D sound reproduction method is proposed. This method uses HRTFs and a correction filter, *Hc*, which compensates for the headphone transfer functions and for the changes in the transfer functions of the ear canals that occur when the subject wears stereo headphones. The results of tests of sound image localization using the proposed method showed that: The locations of reproduced sound images were perceived with much the same accuracy as those of real sound sources in the horizontal and the median planes. Importantly, subjects also perceived all sound images well apart from their heads.

1. Introduction

It is well known that successful 3-D sound localization can be accomplished if head-related transfer functions (HRTFs) are accurately reproduced [1]. However, many previously described 3-D sound reproduction systems [2-8] have left several problems unsolved. When headphones are used, it is important to reproduce HRTFs accurately at the subject's eardrum, not at the entrance of the ear canal. Failure to do so often results in the creation of sound images that seem to be located inside of the subject's head, and also may result in front-back confusions [9]. Wightman and Kistler [2,3] have proposed a method to measure HRTFs at the eardrum and headphone-to-eardrum transfer function for each subject, using a probe tube positioned deep in the subject's ear canal, with its tip about 1-2mm from the eardrum. Møller *et al*. [4] and Pralong and Carlile [5] have concluded that correction for the headphone response measured at the ear-canal entrance is adequate for successful binaural reproduction through headphones. However, since the variability of headphone-to-ear-canal transfer functions is substantial, especially as headphones are removed and replaced upon the subject's head, Kulkarni and Colburn [6] have claimed that making appropriate corrections is actually quite problematic. Nonetheless, it may be that such variability is unlikely to have an adverse effect on the spatial fidelity, as suggested by McAnally and Martin [7]. One solution to the problem is to include microphones in a special pair of "inner-phones," as proposed by Shimada and Hayashi [8].

In this paper, a new 3-D sound reproduction method is proposed that uses ordinary headphones, and does not require specific headphones for successful correction of the transfer functions of the ear canals. The use of a correction filter, *Hc*, is introduced here that can compensate for the changes in the transfer functions of the ear canals that occur while wearing the stereo headphones, and can reproduce the input signals at the subject's eardrums correctly without the use of special headphones.

2. New 3-D Sound Reproduction Method through Stereo Headphones

2.1. Principle

The sound pressure at the entrance of the ear canal of a subject, $P1_{l,r}$, is defined as follows:

$$P1_{l,r} = S \times R \times HRTF(EEC)_{l,r} \quad (1)$$

where S is a source signal, R is a transfer function of a room, $HRTF(EEC)_{l,r}$ is head-related transfer functions measured at the **E**ntrance of the **E**ar **C**anal (EEC) of a subject, and subscript l and r indicate the left and right ears, respectively. The sound pressure at the eardrum, $P2_{l,r}$, is defined as:

$$P2_{l,r} = S \times R \times HRTF(ED)_{l,r}$$

$$= S \times R \times HRTF(EEC)_{l,r} \times H(EC)_{l,r} \quad (2)$$

where $HRTF(ED)_{l,r}$ is head-related transfer functions measured at the subject's **E**ar**D**rum (ED), and $H(EC)_{l,r}$ is a transfer function measured in the **E**ar **C**anal (EC) of the subject (see Fig.1). Let $P3_{l,r}$ denote the sound pressure at the eardrum of a subject when $P1_{l,r}$ is presented to the subject through stereo headphones. $P3_{l,r}$ is then defined as:

$$P3_{l,r} = P1_{l,r} \times HP_{l,r} \times H(EC\_HP)_{l,r} \quad (3)$$

where $HP_{l,r}$ is a transfer function of stereo headphones and $H(EC\_HP)_{l,r}$ is a transfer function in the **E**ar **C**anal with the stereo **H**ead**P**hones (EC_HP) inserted. Let $Hc_{l,r}$ denote a correction filter, which equalize $P3_{l,r}$ to $P2_{l,r}$. From Eqs. (1), (2), and (3), $Hc_{l,r}$ is expressed as follows;

$$P2_{l,r} = P3_{l,r} \times Hc_{l,r} \quad (4)$$

$$Hc_{l,r} = H(EC)_{l,r} / \{HP_{l,r} \times H(EC\_HP)_{l,r}\} \quad (5)$$

Therefore, the sound pressure at the eardrum of a subject in the original sound field is reproduced through stereo headphones, by multiplying $P1_{l,r}$ by $Hc_{l,r}$.

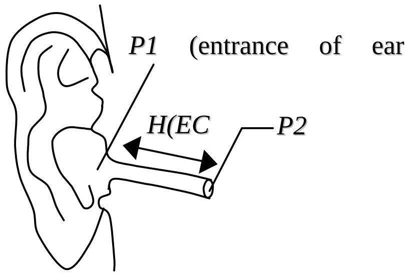


Figure 1: Sketch of the anatomy of the external ear.

2.2. Calculation of compensation filter *Hc*

In order to obtain $Hc_{l,r}$, the transfer functions $H(EC)_{l,r}$ and $H(EC\_HP)_{l,r}$ were measured using a dummy-head, since it is hard to measure those for a human subject. In this study, the KEMAR dummy-head was adopted, because its ear-canal transfer functions are reported practically the same as those of human subjects [10]. $H(EC)_{l,r}$ is obtained from $HRTF(ED)_{l,r}$ and $HRTF(EEC)_{l,r}$ by Eq. (6).

$$H(EC)_{l,r} = HRTF(ED)_{l,r} / HRTF(EEC)_{l,r} \quad (6)$$

$HRTF(ED)_{l,r}$ was measured using the microphones, which were located at the position of the eardrums of the KEMAR dummy-head. $HRTF(EEC)_{l,r}$ was measured at the entrance of the ear canal of the KEMAR dummy-head. According to the results of previous studies (e.g., [11]), $H(EC)_{l,r}$ is expected to be independent of the sound incident angle.

$H(EC\_HP)_{l,r}$ was obtained by emitting an M-sequence signal through the stereo headphones, which were worn on the KEMAR dummy-head, and receiving the signal with the microphones located at the position of the eardrums of the KEMAR dummy-head.

Figure 2 shows the obtained $Hc_{l,r}$ of various incident angles. As was expected, the difference of $Hc_{l,r}$ due to the incident angle is small up to 15kHz. Each $Hc_{l,r}$ has a peak around 2.5kHz and a broad dip around 5kHz. These frequency characteristics show that the resonance frequency of the ear canal moves from 2.5 kHz up to 5kHz when the subject wears the headphones.

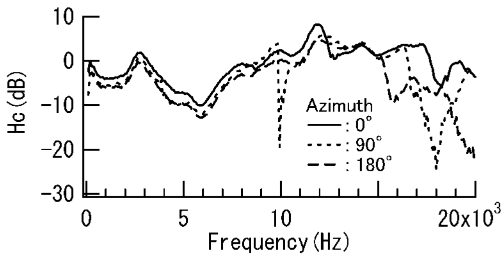


Figure 2: Examples of obtained Hc at various incident angles.

3. Localization Test 1: Examination of the 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 image localization by the proposed method.

3.1. Procedure of localization tests

The HRTFs of the subjects in the horizontal plane and the upper median plane were measured in an anechoic chamber at 30 degree steps. The distance from the sound sources to the center of the subject's head was 1.5m. The ear-microphones shown in Fig.3 were used in the HRTF measurements. The ear-microphones were made using the following procedure. Molds of the ear canals of each subject were made. The miniature electret condenser microphones (diameter: 5mm) and silicon resin were put into the molds. In the measurements, the ear-microphones were put into the ear canals of the subject to satisfy the condition of "blocked entrances," recommended by Hammershøi and Møller [12] for HRTF measurements.

The source signal was wide-band white noise extending from 20Hz to 20kHz. The stimuli were created by convolution of the noise with the subject's own HRTFs, and with *Hc*. In total, stimuli for 19 locations were prepared. The stimuli were divided into two groups. One group consisted of sources located at 12 azimuth angles on the horizontal plane, and the other group consisted of sources located at 7 elevation angles on the median plane. The stimuli were presented through stereo inner-type headphones (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 each stimulus 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 a response-recording sheet. Subjects were four males with normal hearing sensitivity.

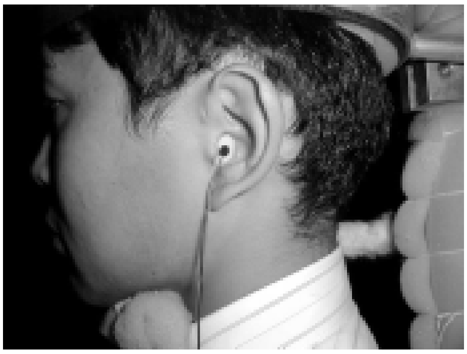


Figure 3: Placement of ear-microphones for HRTF measurement.

3.2. Results and discussions

Figure 4 shows one subject's responses for the stimuli in the horizontal plane and those in the median plane. The responses of the other subjects were almost the same as those of this subject.

For the stimuli in the horizontal plane, the reported azimuths nearly agree with the target azimuths (Fig.4 (a)). 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, Fig. 4 (b) shows that almost all of the responses were distributed around a diagonal line. In the case of 0 degree target elevation, however, the subject sometimes reported the sound image to be located to the rear. The perceived azimuths were concentrated around 0 degrees (in the median plane).

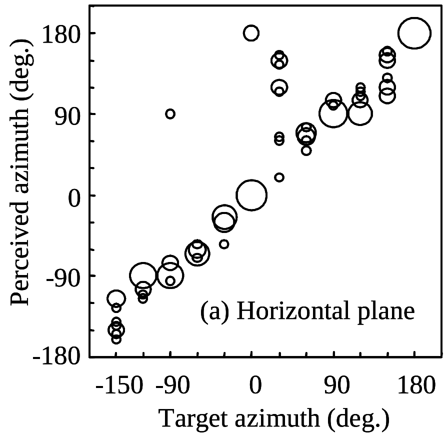
Thus, it seems proper to conclude that the subjects localize sound images accurately using the proposed headphone-based 3-D sound reproduction method.

4. Localization Test 2: Effects of *Hc* on Sound Image Externalization

Localization tests were carried out to examine the effects of *Hc* on the externalization of the sound image, examining whether it was located inside or outside of the subject's head.

4.1. Procedure of localization tests

The HRTFs of the subjects in the horizontal plane (-30, 0, 30, 120, and –120 degrees) were measured in a listening room using the ear-microphones. The ear-microphones were put into the ear canals of the subjects to satisfy the condition of "blocked entrances." The distance from the sound sources to the center of the subject's head was 3.0m. The length of the HRTF was



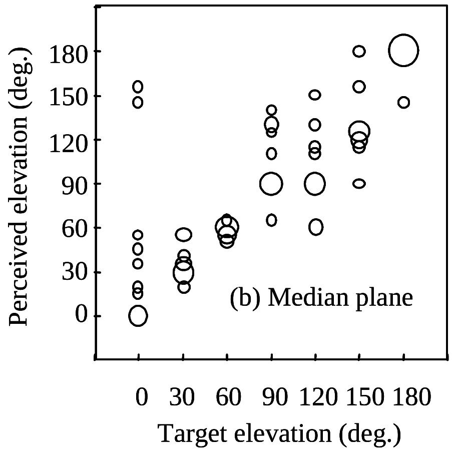


Figure 4: Examples of the responses to the stimuli; (a) perceived azimuth for the stimuli in the horizontal plane, (b) in the median plane.

8192 samples at a sampling rate of 48kHz. The source signal again was wide-band white noise from 20Hz to 20kHz. Two kinds of stimuli were prepared for each target azimuth. One was created by convolution of the noise with the subject's own HRTFs, and no *Hc* was included. The other was created by convolution of the noise with both the subject's own HRTFs and with *Hc*. In total, 10 kinds of stimuli (5 azimuths × presence/absence of *Hc*) were prepared. The stimuli were always presented in pairs, with and without *Hc* for each azimuth (presented via stereo inner-type headphones, Panasonic, RP-HJ535). Each stimulus pair was presented 10 times in random order. Within each pair, the order of presence/absence of *Hc* was also randomized. The tests were carried out in a quiet listening room. The task of the subjects was to indicate which stimulus sound image in each pair seemed to be located further from the head, and which was nearer to the horizontal plane. Subjects were seven males with normal hearing sensitivity.

4.2. Results and discussions

Table 1 shows the results of statistical tests for the responses of all subjects to the stimuli by pair test. This test examined whether the stimuli with *Hc* were localized further from the subject's head than those without *Hc*, and whether they were perceived nearer to the horizontal plane or not.

With regard to distance perception, the sound images of the stimuli processed using *Hc* were localized significantly further from the subject's head than those without *Hc* in the case of the target azimuth of 0, 120, and –120 degrees. In case of the target azimuth of 30 and –30 degrees, no significant difference was observed.

With regard to elevation perception, or rising angle, subjects localized sound images of the *Hc*-processed stimuli significantly nearer to the horizontal plane than those without *H*c for all target azimuth angles.

Table 1: Results of t-tests for the responses of all subjects to the stimuli by pair test. The asterisks show the stimuli with *Hc* were localized further from the subject's head, and nearer to the horizontal plane than those without *Hc*.

Target azimuth (deg.)	Distance	Rising angle
-30	-	**
0	**	**
30	-	***
120	***	***
-120	***	**

***: p< 0.001, **: p<0.01

5. Conclusions

A new 3-D sound reproduction method was proposed that can compensate for the changes in the transfer functions of the ear canals caused by wearing stereo headphones. The method includes a correction filter, *Hc*, that compensates for these changes in order to reproduce HRTFs accurately at the subject's eardrum. The results of localization tests using the proposed method showed the following: The reproduced sound images could be perceived with much the same accuracy as those of real sound sources located in the horizontal and the median plane. Importantly, subjects perceived the *Hc*-processed stimuli significantly further outside of the head than stimuli processed without *Hc*. From these results, it can be concluded that the proposed method using a novel correction filter, *Hc*, provides a successful solution to common problems associated with headphone-based 3-D sound reproduction.

6. Acknowledgements

The authors would like to thank Mr. Matsuda for his help in the localization tests, and Prof. W. L. Martens (Mcgill University) for his help in proofreading and correcting the English version of this manuscript.

7. References

[1] M. Morimoto and Y. Ando, "On the simulation of sound localization," J. Acoust. Soc. Jpn. (E) 1, 167-174 (1980).

[2] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am. 85, 858-867 (1989).

[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] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sorensen, "Transfer characteristics of headphones measured on human ears," J. Audio Eng. Soc. 43, 203-217 (1995).

[5] D. Pralong and S. Carlile, "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," J. Acoust. Soc. Am. 100, 3783-3797 (1996).

[6] A. Kulkarni and H. S. Colburn, "Variability in the characterization of the headphone transfer function, " J. Acoust. Soc. Am. 107, 1071-1074 (2000).

[7] K. I. McAnally and R. L. Martin, "Variability in the headphone-to-ear-canal transfer function," " J. Audio Eng. Soc. 50, 263-266 (2002).

[8] S. Shimada and S. Hayashi, "Stereophonic sound image localization system using inner earphones," Acustica, 81, 157-160 (1992).

[9] J. Blauert, Spatial Hearing (revised edition) (MIT press, Cambridge, MA, 1997).

[10] M. Burkhard and R. Sachs,"Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. 58, 214-222 (1975).

[11] F. M. Wiener and D. A. Ross, "The pressure distribution in the auditory canal in a progressive sound field," J. Acoust. Soc. Am. 18, 401-408 (1946).

[12] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. 100, 408-427 (1996).