Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Effects of headphone calibration functions on subjective impressions of spatial sound

Ryouichi Nishimura, Hiroaki Kato, Parham Mokhtari and Hironori Takemoto

National Institute of Information and Communications Technology (NICT), Kyoto, Japan

PACS: 43.38.Md, 43.38.Vk, 43.58.Vb, 43.60.Dh, 43.66.Pn, 43.66.Qp

ABSTRACT

If the same sound pressure as when a listener were listening to the sound without headphones could be reproduced at the eardrum, the listener would perceive three-dimensional sound even when the sound is presented through headphones. Headphone calibration is therefore required to compensate for individual variations in the transfer function of a listener's ear canal with and without headphones. From a practical point of view, a headphone calibration function applicable to many listeners would be attractive. We measured headphone calibration functions for 245 listeners, and from these data derived the mean calibration function. Its effects on the subjective impressions on spatial features of the reproduced sound were tested through listening tests. Participants compared various virtual three-dimensional sound signals generated by applying different calibration functions: mean calibration function and ones measured using different ears such as artificial ear (B&K 4153), head and torso simulator (B&K 4128), and listener's own ear. Sound stimuli were presented in pairs, and rated by the participants in terms of diffuseness, externalization, and positional closeness between the perceived sound image and the actual loudspeaker corresponding to it. Statistical analyses of the results revealed that the mean headphone calibration function works well in terms of diffuseness and closeness.

INTRODUCTION

From an audiological point of view, it is reasonable to expect that precise reproduction of eardrum vibration at both ears can create an auditory sensation as if the listener were in the original sound scene. The eardrum vibrations usually differ from person to person even when they are listening to the same sound source because differences in their morphological characteristics cause different reflection and diffraction around the head, ears, and even within the ear canals. Head-related transfer functions (HRTFs), one definition of which is a transfer function from a sound source at a particular location in a free field to the eardrum of both ears of the listener, contain all such information. Sound source signals filtered with HRTFs would therefore be perceived as realistic spatial sound if they are presented to both the ears of the listener properly. Using a headphone is the easiest way to control sound pressure at the left and right ears separately as well as independently. It however often involves issues such as tonal changes and sound localization inside the head. It should be emphasized that signals obtained by the methods such as convolving HRTFs with a source signal or binaural recording must be reproduced not at the headphone drivers but at the eardrum. Consequently, a calibration for headphone listening for each individual listener is required.

There have been several research papers on headphone calibration. Møller analyzed this issue by use of equivalent circuits [1]. Hammershøi and Møller further investigated it and described that the sound transmission to the eardrum from the entrance of the ear canal can be considered independent of direction of sound incidence [2]. This inversely implies that the sound transmission to the eardrum from the entrance of the ear canal depends only on the ear canal and the eardrum regardless of the environmental condition outside the entrance of the ear canal. Wightman and Kistler measured HRTFs and the calibration functions using a probe microphone located very close to the eardrum [3]. However, it is difficult to maintain high accuracy and reliability with such measurements. Pralong and Carlile suggested that individual calibration of headphone listening is as important for sound localization performance as the individualization of HRTFs [4]. Referring to [1], Ozawa developed an effective method for how binaural signals recorded with a dummy head should be adjusted to individual listeners [5], but in his method the listener must go to the recording environment at least once for perfect calibration. In all of these former investigations, ear canals were assumed open when recording.

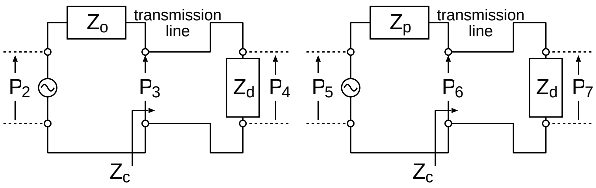
Other than measuring HRTFs, computer simulations are promising because the task of measuring HRTFs is time consuming and imposes a considerable burden on listeners, and may introduce errors caused by listener's movements. Morphological information of the listener is necessary to simulate HRTFs by means of numerical analysis such as the boundary element method (BEM) [6] or the finite difference time domain (FDTD) method [7]. It is however difficult to measure the shapes of the ear canal and eardrum with an optical scanner and magnetic resonance imaging (MRI) because the laser beam cannot reach the inside of ear canal which is not a straight tube and because the soft-tissue wall of ear canal is too thin to be imaged. Consequently, simulations are typically carried out assuming the ear canals blocked. As a result, HRTFs obtained under such conditions do not exactly match those observed at the eardrums assuming open ear canals.

It is apparent from these research papers that a calibration function plays an important role for reproducing the desired auditory sensation. With the aid of electrical equivalent circuit models, this paper begins with theoretical backgrounds on the necessity of headphone calibration to reproduce sound pressures at the eardrum. Variations in headphone calibration functions among different human ears are then discussed. After that, experimental setups for the listening tests, which aimed to reveal effects of calibration functions on the perception of reproduced sound space, and the results are presented.

HEADPHONE CALIBRATION

Equivalent circuit

Møller investigated relationships of sound pressure at the entrance to the ear canal and at the eardrum by means of equivalent circuits depicted in Fig. 1 [1]. Figure 1(a) corresponds to



(a) Free-air listening condition (b) Headphone listening condition

Figure 1: Equivalent circuits considered in [1]

a free-air listening condition, where the listener is listening to the sound without headphones. Figure 1(b), on the other hand, corresponds to a headphone listening condition, where the listener is listening to the sound with headphones. The meanings of the impedance variables that appear in Fig. 1 are as follows:

- Z_o : Equivalent impedance outside of the ear canal without headphones (radiation impedance),
- Z_p : Equivalent impedance outside of the ear canal with headphones,
- Z_c : Equivalent impedance inside of the ear canal (ear-canal impedance),
- Z_d : Impedance of the eardrum.

All variables are functions of frequency and therefore dependent on angular frequency ω. The notation of (ω) is however omitted hereafter for brevity, as long as no special attention is necessary on the argument. Conforming to the former literature [1], variables relevant to sound pressure are defined as follows:

- P_1 : Sound pressure at the center of the head while the listener is not present (defined in [1] but not used in this paper),
- P_2 : Sound pressure at the entrance to the blocked ear canal without headphones,
- P_3 : Sound pressure at the entrance to the open ear canal without headphones,
- P_4 : Sound pressure in front of the eardrum without headphones,
- P_5 : Sound pressure at the entrance to the blocked ear canal with headphones,
- P_6 : Sound pressure at the entrance to the open ear canal with headphones,
- P_7 : Sound pressure in front of the eardrum with headphones.

The constant voltage source corresponds to the sound pressure observed at the entrance to the blocked ear canal. The sound pressure observed at the entrance to the open ear canal can be obtained by dividing the voltage of the constant voltage source according to the impedance outside and inside of the ear canal as

$$P_3 = \frac{Z_c}{Z_o + Z_c} \cdot P_2, \tag{1}$$

$$P_6 = \frac{Z_c}{Z_p + Z_c} \cdot P_5. \tag{2}$$

The ratio of the sound pressure at the entrance to the open ear canal with and without headphones is then obtained by dividing Eq. (1) by Eq. (2) as

$$\frac{P_3}{P_6} = \frac{Z_p + Z_c}{Z_o + Z_c} \cdot \frac{P_2}{P_5}. \tag{3}$$

Because the acoustical characteristics of the listener's ear canal and eardrum can be considered consistent under both the headphone and the free-air listening conditions, if the sound pressure at the entrance to the ear canal is the same under both the conditions, the equality is expected to hold also at the eardrum. In other words, if P_6 equals to P_3, P_7 can be expected completely equal to P_4, resulting in the headphone calibration for perfect reproduction of sound pressure at the eardrums. It is therefore easy to calibrate headphones if P_3 and P_6 can be obtained directly.

Equation (3) suggests that if $(Z_p + Z_c)/(Z_o + Z_c) \simeq 1$, sound pressure in front of the eardrum under the free-air listening condition P_4 can be reproduced by making the sound pressure at the entrance to the blocked ear canal under the headphone listening condition P_5 equal to that under the open listening condition P_2 because P_6 becomes identical to P_3 as $P_3/P_6 \simeq P_2/P_5$ is satisfied in this case. The coefficient $(Z_p + Z_c)/(Z_o + Z_c)$ in Eq. (3) is called the pressure division ratio (PDR), and headphones that realize PDR $\simeq 1$ have so-called free-air equivalent coupling (FEC) characteristics [8]. The use of headphones having FEC characteristics, i.e., PDR $\simeq 1$, is attractive especially when the HRTFs are calculated by computer simulations because only P_2 instead of P_3 is generally obtained in computer simulations.

In this study, P_3 is assumed obtainable. Hence, a transfer function from a driving signal of the headphones to the sound pressure P_6 is required to reproduce the sound pressure at the eardrum. This transfer function is therefore regarded as a calibration function.

Calibration functions

Table 1 briefly lists the five calibration functions considered in this study. Procedures to obtain each calibration function are as follows:

'Phone' The inverse of the headphone characteristics measured using an artificial ear (B&K 4153) was regarded as a calibration function excluding potential effects by the pinna. To avoid effects brought about by using a microphone different from one used in the measurements of other calibration functions, sound pressure was measured with a microphone (Knowles FG-23629) placed in close proximity of the microphone installed in the artificial ear. An optimized Aoshima's time-stretched pulse of 32768 points was presented 11 times consecutively, at a sampling rate of 48 kHz, and observations of the latter 10 trials were adopted for the calculation of the transfer function using the discrete Fourier transform (DFT). The obtained transfer function, which was represented in the frequency domain, was inverted in complex amplitude to yield a calibration function.

'Own' A microphone (Knowles FG-23629) was placed in the listener's open ear canal, with its face aligned with the entrance of the ear canal. A listener wore circum-aural closed-type headphones (Senheiser HDA-200) over it. The rest of the measurement procedure was the same as that for obtaining the calibration function of 'Phone.'

'HATS' A head and torso simulator (B&K 4128) was used instead of a human listener. The rest of the measurement procedure was the same as others.

Table 1: Five calibration functions compared in this study

Acronym	Compensation for	Method to obtain
None	Nothing	–
Phone	Characteristics of the headphone	Measurement using an artificial ear (B&K 4153)
HATS	Characteristics of the headphone coupled with a typical external ear	Measurement using a head and torso simulator (B&K 4128)
Own	Characteristics of the headphone coupled with the listener's own ear	Measurement for each listener
Mean	Averaged transfer function	Calculation from the data for 245 listeners

'Mean' Calibration functions were measured for 245 individuals (177 male and 68 female), who visited the open laboratory last year and experienced an audio demonstration. The measurements were carried out in a small meeting room where people were chatting in the opposite side of the room. In order to obtain the mean calibration function, the amplitude spectrum of the obtained calibration functions were first converted into logarithmic amplitude scale and then averaged over all the individuals, separately for the left and right ears. After that, its minimum phase was derived from the amplitude spectrum. The obtained minimum phase and the amplitude spectrum were combined to compose a complex frequency spectrum.
The inverse Fourier transform of the spectrum was regarded as the calibration function. Therefore, the calibration function $g(t)$ was obtained as

$$g(t) = \mathrm{IDFT}[|G(\omega)|\exp(-j\arg[G(\omega)])], \qquad (4)$$

where the amplitude spectrum $|G(\omega)|$ was calculated by

$$|G(\omega)| = 10^{-\overline{|H(\omega)|}} = 1/10^{\overline{|H(\omega)|}}, \qquad (5)$$

$$\overline{|H(\omega)|} = \frac{1}{N}\sum_{n=1}^{N} \log_{10}|H_n(\omega)|, \qquad (6)$$

$$N = 245$$

The amplitude spectra of the four calibration functions are depicted in Fig. 2 , where the 'Own' calibration function is a mere

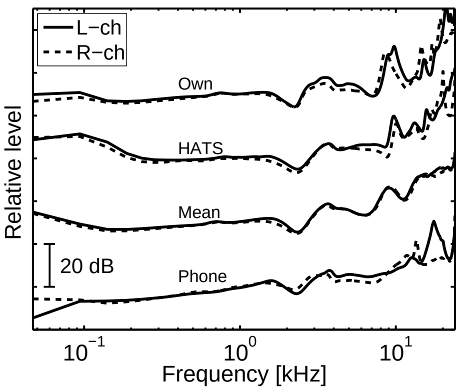


Figure 2: Amplitude spectra of the calibration functions, displayed with equal vertical spacing at 1 kHz

instance because it exists for each of the 245 participants. A dip appears at approximately 2 kHz in all the calibration functions, and could be considered as a characteristic of the headphones used in the tests. Ear canal resonance could be another possible cause of the dip. Two dips appear at approximately 6 kHz and 10 kHz in all the calibration functions except 'Phone' so that they could be regarded as effects originating from the pinna. In the calibration functions of 'Own' and 'HATS', multiple dips and peaks appear above 10 kHz, while the calibration function of 'Mean' shows a flat response in this frequency region. This difference implies that these dips and peaks could originate from the reflection and diffraction at small convex and concave parts of the listeners' pinnae. Accordingly, these peaks and dips can be regarded as individual differences. Nevertheless, there is another possibility that wearing headphones in different ways might cause the difference. The actual reason has not been determined yet.

Individual differences

For further investigation on individual differences in the calibration functions, they were analyzed with an auto-regressive model. The order of the model was determined experimentally to obtain a convergent pole distribution, and the eighth order was adopted. Figure 3 is a polar plot of the pole distribution

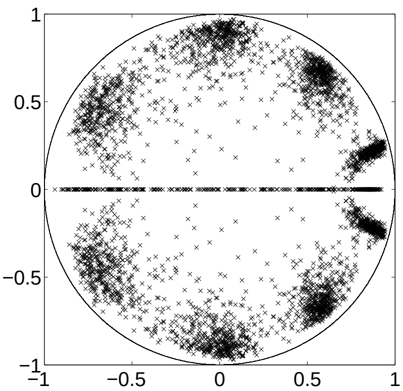


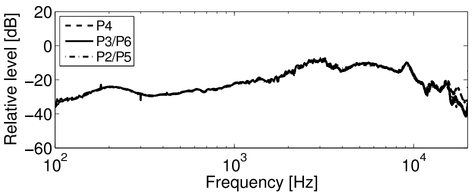
Figure 3: Pole distribution of eighth-order auto-regressive models of the calibration functions measured for 245 participants

of the 245 transfer functions from the headphone driving function to the sound pressure at the entrance to the open ear canal. Because the sampling frequency was set to 48 kHz, the right-most and left-most positions of the circle correspond to 0 Hz and 24 kHz, respectively.

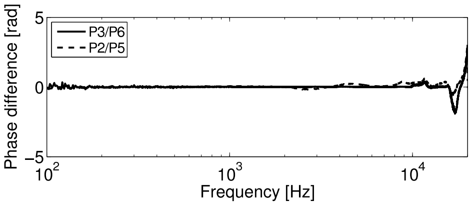
Commonly to all the listeners, a pole appears in proximity to 2 kHz but its depth varies considerably. This pole therefore could be considered corresponding to the dip which has been seen at approximately 2 kHz in all the calibration functions in Fig. 2, and could be caused by ear canal resonance or the headphone characteristic. Two poles at approximately 6 kHz and 10 kHz are broad in frequency and shallow in depth, or apart from the unit circle. Yet these poles are certainly observed also in the 'Mean' calibration function as depicted in Fig. 2, where the poles appear as dips because of the inverse relationship. Fourth poles have a similar distribution to the second and the third in Fig. 3 but disappear in the 'Mean' calibration function probably due to a distribution broader in frequency and shallower in depth.

Validation of calibration

A measurement was conducted to validate that a calibration function calibrates the sound pressure correctly at the eardrum so as to be the same as when a listener were listening to the sound without headphones. In the measurement, the sound pressure captured at the microphone installed in the ear of a head and torso simulator (B&K 4128) was regarded as the sound pressure at the eardrum. Figure 4 shows the results where the



(a) Power spectrum at the eardrum



(b) Phase spectrum difference at the ear drum

Figure 4: Sound pressures at the eardrum calibrated with (solid line) and without (dashed line) the calibration function. The sound source was located in the listener's facing direction.

solid and dashed lines correspond to those measured with and without the calibration function, respectively. Because the calibration function was measured with the same combination of the headphones and the head and torso simulator as in the reproduction, it is regarded as the 'Own' calibration function. The upper and lower panels are the power and phase spectrum, respectively. In the upper panel, P_4 is the sound pressure measured in front of the eardrum without headphones, and therefore is regarded as the desired characteristics. Because the lower panel is represented in the phase differences from the desired characteristics, phase difference of zero is desirable.

Considering the wave length of the sound, equivalent circuits are valid only for the frequency region lower than 10 kHz. Nevertheless, good reproduction of sound pressure was realized up to approximately 14 kHz in both the power and phase spectrum. Moreover, it is apparent from this figure that with an appropriate calibration function, reproduction of sound pressure at the eardrum can be achieved over the frequency range from approximately 100 Hz to 14 kHz. With the calibration function, log spectral distortion of the reproduced sound pressure in this frequency region was 0.9 dB on average over the six sound source directions tested while it was 1.5 dB, without the calibration function.

LISTENING TEST

In this section, effects of various calibration functions on the perception of sound space are investigated through listening tests.

Procedure

The listening tests consisted of two stages. In the first stage, a participant, with a miniature microphone placed at the entrance of each ear canal, sat on a chair surrounded by six loudspeakers (Eclipse TD508II) with different height in a listening room, as depicted in Fig. 5. The gain of each loudspeaker was cali-

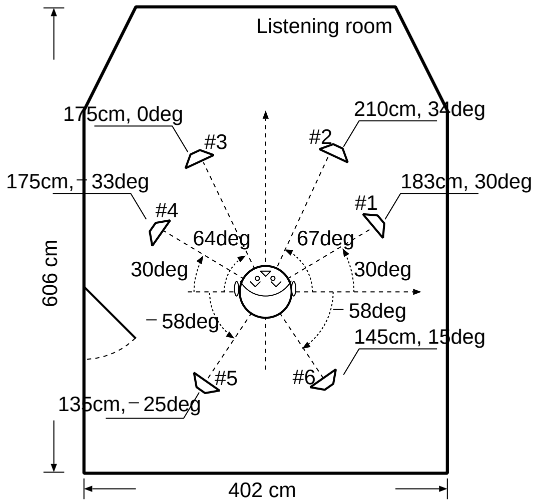


Figure 5: Loudspeaker arrangement (including distance and elevation of each speaker from the listener's head)

brated so as to make the sound pressure of each loudspeaker identical at the center of the listener's head. The sound stimulus consisted of three consecutive sound clips: the first two were a sound of hitting a stainless steel bowl recorded in an anechoic room, and the last one was a cat's voice excerpted from a sound effect CD, resulting in a sequential sound like "gong-gong-meow." The sound stimulus was presented once from each loudspeaker and one by one in an anti-clockwise order. Accordingly, each participant listened to the same sound stimulus six times but from different loudspeakers. The initial loudspeaker was randomized for each participant. While the listener was listening to a sequence of sound stimuli, sound pressure was measured with the miniature microphones placed at the entrance of the ear canal, and regarded as P_3.

In the second stage, the participant was in the same condition as in the first stage but wore headphones. Prior to the listening tests, the participant's 'Own' calibration function was measured. In addition to that, sound stimuli for the tests were generated by applying various headphone calibration functions explained in a former section to the measured signal of P_3. The amplitude of sound stimuli for each calibration function was adjusted so as to have an identical mean squared value. Then through the headphones, the participant listened to two sequences of sound stimuli, each of which was generated using different headphone calibration functions. One of the paired sound stimuli was presented first and then followed by the other. The time pattern of the presentation is illustrated in Fig. 6. After listening to both sound stimuli, the listener judged them

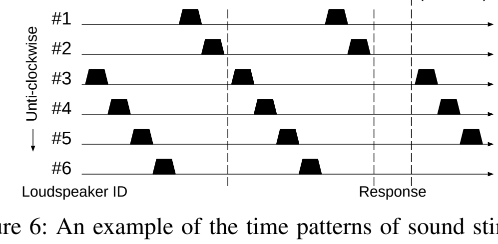


Figure 6: An example of the time patterns of sound stimulus presentation

in terms of 'diffuseness', 'externalization', and 'localization.' Judgments may vary among the loudspeakers from which sound stimulus radiates so that the participants were asked to make a judgment based on their overall impression. Every participant judged all combinations of calibration functions as well as presentation orders.

Participants judged their impressions to the paired sound stimuli and answered using the response form shown in Fig. 7. ‘Diffuseness’ was judged on the basis of an alternative forced

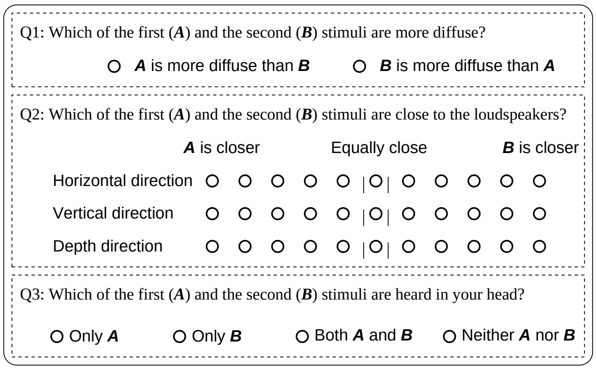


Figure 7: Response form

choice between the paired sound stimuli. ‘Localization’ was tested based on the Scheffe’s paired comparison modified by Ura [9]. ‘Externalization’ was judged for each of the paired sound stimuli. Nine male and eight female paid participants aged from 20 to 35 took part in the listening tests. They were allowed to replay each pair of sound stimuli as many times as they liked before making their judgments.

Diffuseness

Table 2 shows the results regarding ‘diffuseness,’ which is presented in terms of just noticeable difference (JND); a value greater than one indicates that a sound stimulus generated with the calibration function specified in the left column was perceived with noticeably smaller diffuseness than that by the calibration function specified in the top row. Those having values greater than one are highlighted with underlines.

Table 2: d' in terms of ‘diffuseness’

	None	Phone	HATS	Mean
Phone	0.74			
HATS	<u>1.88</u>	0.74		
Mean	<u>2.15</u>	<u>1.07</u>	0.33	
Own	<u>1.65</u>	0.68	−0.05	−0.22

These results suggest that in general the mean calibration function has an ability to make the sound image clearer or more solid, and that the listener’s own calibration function and that measured with a head and torso simulator created clearer sound images than when no calibration function was applied. Even including combinations that showed no significant differences, the mean calibration function always yielded the least diffuseness in the reproduced sound images tested.

Externalization

Figure 8 shows results with respect to ‘externalization,’ where externalization ratios for the tested calibration functions are represented by a box and whisker plot. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers show the rest of the data, and the asterisks denote outliers.

While the mean values of externalization rate for ‘None’ and

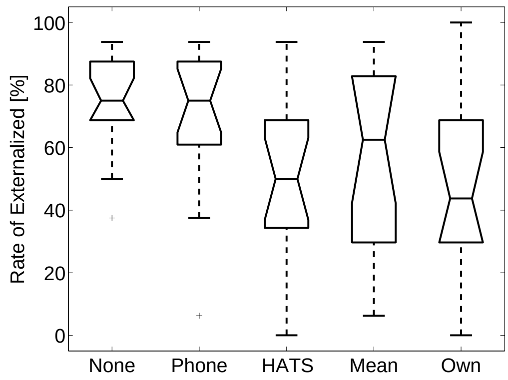


Figure 8: Box-whisker plot of the rate of externalization for each calibration function

‘Phone’ fall within a range from 70% to 80%, those for ‘HATS’, ‘Mean’, and ‘Own’ fall roughly within a range from 40% to 60%. Accordingly, no calibration and the calibration function for headphone characteristics showed higher externalization ratios than other calibration functions. However, there were considerable individual differences and no significant difference was observed. Nevertheless, ‘Mean’ showed the highest mean value among those three (‘HATS’, ‘Mean’, ‘Own’) calibration functions that showed worse performance.

Localization

Because every participant judged all the pairs, obtained responses were analyzed by means of Scheffe’s paired comparison modified by Ura. Table 3 shows the results with respect to three different directions perpendicular to each other: vertical direction (top table), horizontal direction (middle table), and depth direction (bottom table). Asterisks denote pairs for which a sound image reproduced with the calibration function specified in the left column was perceived significantly closer to the actual loudspeaker corresponding to it than that reproduced with the calibration function specified in the top row.

Table 3: Proximity of the perceived sound location to the actual loudspeaker

(a) Vertical direction

	None	Phone	HATS	Mean	Own
None	-				
Phone		-			
HATS			-		*
Mean				-	**
Own					-

(b) Horizontal direction

	None	Phone	HATS	Mean	Own
None	-				
Phone		-			
HATS			-		
Mean				-	
Own					-

(c) Depth direction

	None	Phone	HATS	Mean	Own
None	-				
Phone		-			
HATS	**	**	-		
Mean	**	**		-	
Own					-

** $p < .01$ * $p < .05$

With respect to the vertical direction, the calibration functions of 'HATS' and 'Mean' significantly outperformed that of 'Own.' With respect to the horizontal direction, on the other hand, no significant difference was seen for any pair of conditions. Meanwhile, the calibration functions of 'HATS' and 'Mean' showed better performance compared with those of 'None' and 'Phone' with respect to the depth direction.

DISCUSSION AND INFORMAL TEST

In the localization test, no significant difference was observed for the horizontal direction. While the perceived position of a sound in the horizontal direction can be determined primarily based on the interaural level and time differences, estimation of vertical and depth directions of the sound strongly relies on differences in the spectral shapes of the perceived sound. This would be a reason why effects of the calibration functions were observed only in the vertical and depth directions because the applied calibration functions modify neither level nor time differences between the two ears.

An additional informal listening test was conducted with five listeners who were engaged in the research on realistic communications and had experienced listening tests on spatial audio several times. The test results showed tendencies clearly different from those obtained with naive listeners. In the diffuseness test, the 'Own' calibration function performed significantly better (less diffuse) than all the others. The 'Mean' calibration function also performed significantly better than all the others except the 'Own' calibration function. In the externalization test, regardless of the calibration function applied, high externalization rates were obtained and due to that, no significant difference was observed. In the localization test, both of the 'Mean' and the 'Own' calibration functions performed well compared with other calibration functions. These differences between the trained and naive listeners might arise from the fact that apart from these tests, the naive listeners had probably listened to sound stimuli captured with any technique other than ordinary stereophonic or monophonic recordings. Because source signals had been captured at the entrance of the listeners' own ears, even when no calibration function was applied, the sound was heard with sufficient spatial information compared with what out naive participants usually listen to with headphones. As a result, the naive listeners might have perceived less differences among the tested sound signals, and responded with considerable variations.

CONCLUSION

Effects of headphone calibration functions on the reproduction of sound space were investigated both objectively and subjectively. In addition, to prevent the inconvenience of measuring headphone calibration functions for each individual, the mean headphone calibration function was derived and evaluated through listening tests. The test results showed that the mean calibration function performed well in terms of diffuseness and localization to some extent, compared with the other tested calibration functions. Meanwhile, some problems in conducting subjective listening tests on spatial audio with naive listeners were discovered. For more reliable conclusions, additional listening tests would be required in which instructions to participants as well as a set of test signals should be reconsidered in order to obtain responses as reliable as the trained listeners'.

REFERENCES

[1] Henrik Møller. "Fundamentals of Binaural Technology". *Applied Acoustics* 36 (1992), pp. 171–218.

[2] Dorte Hammershøi and Henrik Møller. "Sound transmission to and within the human ear canal". *J. Acoust. Soc. Am.* 100.1 (1996), pp. 408–427.

[3] Frederic L. Wightman and Doris J. Kistler. "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis". *J. Acoust. Soc. Am.* 85.2 (1989), pp. 858–867.

[4] Danièle Pralong and Simon Carlile. "The Role of Individualized Headphone Calibraiton for the Generation of High Fidelity Virtual Auditory Space". *J. Acoust. Soc. Am.* 100.6 (1996), pp. 3785–3793.

[5] Kenji Ozawa. "Expansion of an individual equalization method for binaural signals". *Acoust. Sci. & Tech.* 26.3 (2005), pp. 309–311.

[6] Brian F. G. Katz. "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation". *J. Acoust. Soc. Am.* 110.5 (2001), pp. 2440–2448.

[7] Parham Mokhtari et al. "Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data". *the 123rd AES Convention*. 7240. 2007.

[8] Henrik Møller et al. "Transfer Characteristics of Headphones Measured on Human Ears". *J. Audio Eng. Soc.* 43.4 (1995), pp. 203–217.

[9] Shin Sato. *Statistical Sensory Evaluation (in Japanese)*. Nikkagiren, 1985.