Performance Evaluation of 3D Sound Field Reproduction System Using a Few Loudspeakers and Wave Field Synthesis

Munenori Naoe[12], Toshiyuki Kimura[12], Yoko Yamakata[1], and Michiaki Katsumoto[1]
[1]Universal Media Research Center,
National Institute of Information and Commnications Technology
4-2-1 Nukui-Kitamachi, Koganei, Tokyo, 184-8795 Japan
[2]Graduate School of Engineering,
Tokyo University of Agriculture and Technology
2-24-16 Naka-cho, Koganei, Tokyo, 184-8588 Japan
{nm.s512.ex, t-kimura, yamakata, katumoto}@nict.go.jp

Abstract

A conventional 3D sound field reproduction system using wave field synthesis places a lot of loudspeakers around the listener. However, since such a system is very expensive and loudspeakers come into the listener's field of vision, it is very difficult to construct an audio-visual system with it. We developed and evaluated a 3D sound field reproduction system using eight loudspeakers placed at the vertex of cube and wave field synthesis. We compared the sound localization of a loudspeaker array with that of seventeen loudspeakers placed around the listener and found that their localization capabilities of twelve directions were good.

1. Introduction

Recently, several sound field reproduction techniques have been developed for auditory virtual reality systems. By practical application of these techniques, people in different places can conduct and participate in events such as conferences (teleconferencing system) and music concerts (tele-ensemble system) at the same time. Thus, it can be stated that the use of telecommunication systems in society will increase rapidly as these systems are capable of producing more realistic environments and sensations than conventional systems (TV phone and 5.1 ch audio).

Wave field synthesis [1, 2, 3, 4, 5] is a sound field reproduction technique that synthesizes wave fronts by using Huygens' principle. The original sound is first recorded using a microphone array in a control area and then reproduced in a listening area by a loudspeaker array. The arrays are placed at the boundaries of their respective areas. The positions of the microphones and loudspeakers are the same with regard to their respective areas. This technique enables multiple listeners to move about in a listening area or to turn their heads and still hear the same sound. This type of sound field reproduction is not possible with conventional sound field reproduction techniques such as the binaural [6] and transaural [7] techniques.

In conventional sound field reproduction systems that use wave field synthesis, loudspeakers are placed in a line [1, 3] or surround the listener on a horizontal plane [2, 4, 5] in order to reproduce the sound field of a 2D space. The sound field reproduction system, in which a lot of loudspeakers are placed around the listener, is also proposed in order to reproduce the sound field of a 3D space [8]. However, since these systems are very expensive and the loudspeakers are visible in the listener's field of vision, it is very difficult to construct an audio-visual system using these systems.

The number of microphones and loudspeakers used by the system can be reduced by considering the auditory capability of the listeners, even if wave fronts are reproduced in the low-frequency range [4]. Thus, by performing a listening test and gauging the auditory capability of the listeners, a practical system can be constructed using only the minimum required number of microphones and loudspeakers.

In this study, we investigate the performance of a 3D sound field reproduction system with eight loudspeakers placed at the vertices of a cube. The use of wave field synthesis is proposed to reproduce the 3D sound field, even when the number of loudspeakers is considerably reduced to prevent the loudspeakers from appearing in the listener's field of vision. The auditory capability of the proposed system is evaluated by the localization test.

The diagram of the proposed 3D sound field reproduc-

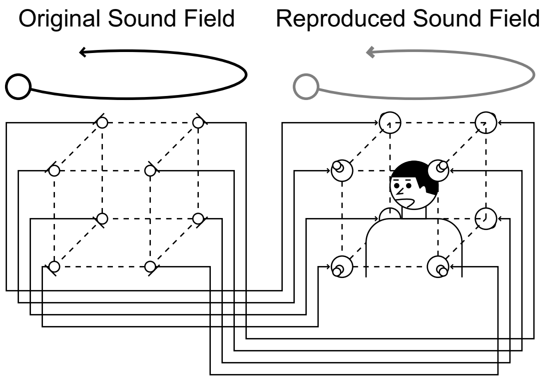

Figure 1. Diagram of proposed 3D sound field reproduction system.

Table 1. Azimuth and elevation angles of sound sources in the localization test.

Number	θ	ϕ	Number	θ	ϕ
1	-90°	-45°	10	90°	0°
2	0°	-45°	11	135°	0°
3	90°	-45°	12	180°	0°
4	180°	-45°	13	-90°	45°
5	-135°	0°	14	0°	45°
6	-90°	0°	15	90°	45°
7	-45°	0°	16	180°	45°
8	0°	0°	17	—	90°
9	45°	0°			

tion system is shown in Figure 1. First, a sound is captured using a cubic microphone array in the original sound field, as shown on the left in Figure 1. Second, the captured sound is played by the cubic loudspeaker array in the reproduced sound field, as shown on the right in Figure 1. As a result, the 3D sound field captured by the microphone array is reproduced by the loudspeaker array. Thus, as shown in Figure 1, the listener, who is in the loudspeaker array, feels that sound is moving above their head when sound is moving above the microphone array.

2. Localization test

In order to evaluate the auditory capability of the proposed 3D sound field reproduction system, a localization test was performed.

2.1. Synthesis of multichannel signals

The multichannel signals played by the loudspeaker array were synthesized on a computer. Directional perception mainly depends on the direct sounds from a sound source. Thus, the original sound field was assumed to be a free space. The room impulse response from the sound source to the ith microphone (i = 1...8), $g_i(n)$, is denoted as follows:

$$g_i(n) = \frac{1}{d_i}\delta\left\{n - \text{round}\left(\frac{d_i F_s}{c}\right)\right\}, \quad (1)$$

where F_s (=48 kHz) is the sampling frequency, c (=340 m/s) is the sound velocity, $\delta(n)$ is Dirac's delta function, and d_i (=$|\mathbf{r}_0 - \mathbf{r}_i|$) is the distance between the sound source and the ith microphone. The values of $\mathbf{r}_0$ and $\mathbf{r}_i$ (position vectors of the sound source and the ith microphone, respectively) were set as follows:

$$\mathbf{r}_0 = (d\cos\theta\cos\phi \quad d\sin\theta\cos\phi \quad d\sin\phi)^T, \quad (2)$$

$$\mathbf{r}_i = \begin{cases} (-0.2 & -0.2 & -0.2)^T & (i=1) \\ (\ \ 0.2 & -0.2 & -0.2)^T & (i=2) \\ (\ \ 0.2 & 0.2 & -0.2)^T & (i=3) \\ (-0.2 & 0.2 & -0.2)^T & (i=4) \\ (-0.2 & -0.2 & 0.2)^T & (i=5) \\ (\ \ 0.2 & -0.2 & 0.2)^T & (i=6) \\ (\ \ 0.2 & 0.2 & 0.2)^T & (i=7) \\ (-0.2 & 0.2 & 0.2)^T & (i=8) \end{cases}, \quad (3)$$

where d (=1, 3 m) denotes the distance between the sound source and the listening position and θ and ϕ are the azimuth and elevation angles, respectively, in the listening position. The values of θ and ϕ were set as shown in Table 1.

If the source signal is represented by $s(n)$, $x_i(n)$, which represents the channel signals recorded by the ith microphone, is denoted as follows:

$$\begin{aligned} x_i(n) &= D_i\{g_i(n) * s(n)\} \\ &= \frac{D_i}{d_i}s\left\{n - \text{round}\left(\frac{d_i F_s}{c}\right)\right\}, \end{aligned} \quad (4)$$

where * is the convolution. Previous studies have indicated that the sound is only recorded from outside the control area according to D_i (the directivity of the ith microphone) [5]. In this study, D_i was set to shotgun directivity as follows:

$$D_i = \begin{cases} \cos\theta_i & (|\theta_i| \le 90^\circ) \\ 0 & (|\theta_i| > 90^\circ) \end{cases}, \quad (5)$$

where θ_i (incident angle of the sound source in the ith microphone) is defined as follows:

$$\theta_i = \cos^{-1}\left\{\frac{\mathbf{r}_i\cdot(\mathbf{r}_0 - \mathbf{r}_i)}{|\mathbf{r}_i||\mathbf{r}_0 - \mathbf{r}_i|}\right\}, \quad (6)$$

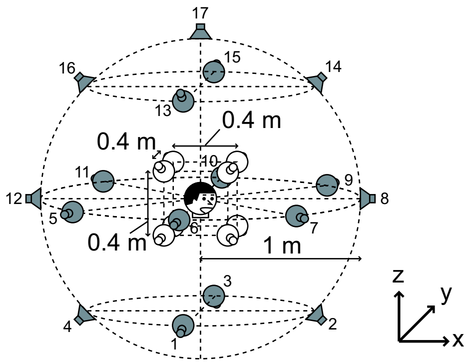

Figure 2. Position of a listener and the loudspeakers in the localization test.

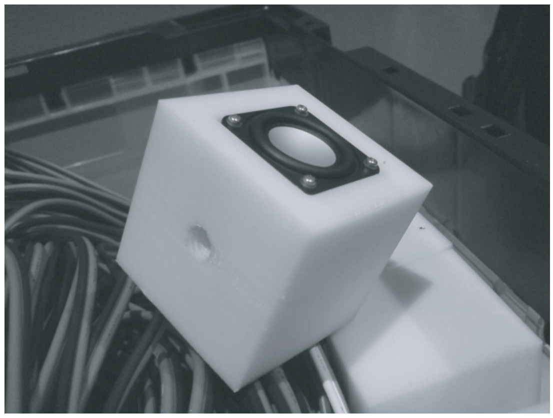

Figure 3. Image of manufactured loudspeakers.

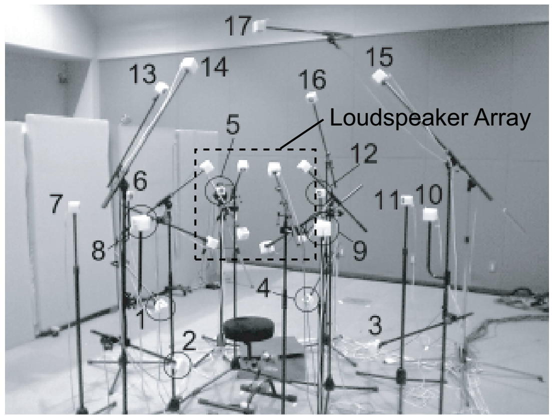

Figure 4. Setup of the loudspeaker array and loudspeakers for the control condition.

2.2. Experimental environment

The localization test was performed in a room with a reverberation time of 180 ms. Twenty-five loudspeakers were placed in the positions shown in Figure 2. The listening position was placed on the center of a sphere. The white loudspeakers indicate eight loudspeakers placed at the vertex of a cube having sides measuring 0.4 m. The gray loudspeakers indicate seventeen loudspeakers placed on a sphere with a radius of 1 m; these loudspeakers were used for the control condition. The values of the azimuth and elevation angles of seventeen loudspeakers in the listening position are the same as those shown in Table 1. Loudspeakers were manufactured by mounting a loudspeaker unit (Aurasound: modified NSW1-205-8A) on a loudspeaker box as shown in Figure 3. The setup of the loudspeaker array and loudspeakers for the control condition is shown in Figure 4. The white boxes in Figure 4 denote the manufactured loudspeakers. A background noise level of A-weighted level of 23 dB and the sound pressure level in the listening position was set to A-weighted level of 60 dB.

The three conditions in the localization test are shown in Figure 5. In the control condition (a), the sound source itself was presented to subjects by playing a sound source from one loudspeaker selected from a group of seventeen. In conditions (b) and (c), eight channel signals calculated by Eq. (4) were played from eight loudspeakers. Subjects reported feeling that there are synthetic sound images in the positions occupied by the gray circles, as shown in Figure 5.

2.3. Experimental procedure

Six males and one female participated as listeners in this test. The flowchart of the localization test is shown in Figure 6. In the test, two sound sources (white noise and speech) were used. White noise was synthesized using MATLAB and the speech was an excerpt from SQAM-CD [9]. The test was divided into two sessions for each sound source. The order of presentation of the sound sources was randomized for each listener. One hundred fifty-three main trials were performed following thirty-four practice trials were conducted. During the main trials, rest periods were allowed after every set of 51 trials. The orders of the trials were randomised for each listener. The details of the practice and main trials are shown in Table 2.

The listeners were instructed to report the perceived direction of sound by listing the number of the direction in an answer sheet. The relation between the perceived directions and direction numbers which described in an answer sheet is shown in Figure 7. The listeners were allowed to turn their heads freely while listening to the sounds.

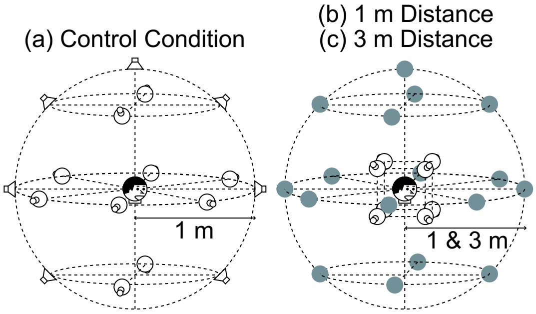

Figure 5. Three experimental conditions in the localization test.

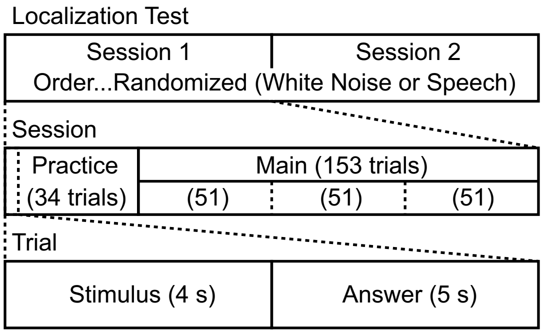

Figure 6. Flow chart of the localization test.

2.4. Results and discussions

The auditory capability of the proposed system was evaluated by calculating the accuracy rates defined as follows:

$$\text{Accuracy rate} = \frac{\text{The number of correct answers}}{\text{The number of presentations}}. \quad (7)$$

The accuracy rates and the results of the chi-square test for each direction are shown in Tables 3 and 4. * and ** in Tables denote that there are the significant differences of 5% and 1% levels between the control condition and the conditions of the proposed system by the chi-square test. It was observed that in five directions (6, 8, 10, 12 and 17), the accuracy rates of the proposed system were lower than those of the control condition because there were the significant differences of the 1% level in almost all cases. However, in other directions, the accuracy rates of the proposed system were almost the same as those of the control condition since there were no significant differences in almost all cases. Thus, it is considered that the performance of the proposed system is good in all the directions, except in the five directions stated above.

In order to evaluate the perceived directions in the five directions described above, the answer rates were calculated. The answer rates are defined as follows:

$$\text{Answer rate} = \frac{\text{The number of answers}}{\text{The number of presentations}}. \quad (8)$$

The results of the answer rates for the five presented directions (6, 8, 10, 12, and 17) are shown in Figure 8. The sound is presented from the back of the heads placed at the center of two concentric circles. The numbers placed on two concentric circles denote the direction numbers used in the localization test. The white circles on the numbers denote the answer rates of the perceived directions. It is indicated that the answer rates are high if the white circles are large.

It was observed that in some cases when the direction number was actually 6, an erroneous answer of 7 or 13 was provided. Thus, it is considered that when a sound image is produced from the direction on the right hand side of the listeners, they localize the sound image towards the forward and upper directions.

On the other hand, when the direction number was 8, the most common erroneous answer was 14. Thus, it can be inferred that when a sound image is produced from the region in front of the listeners, the listeners localize the sound image towards the upper direction. It seems that the answer rates of the direction number 14 in the white noise is more

Table 2. Details of the practice and main trials in the localization test.

	Element	Note
Practice (34)	= 17 directions × 2 conditions	(a) and (b) of Figure 5
Main (153)	= 17 directions × 3 conditions × 3 repetitions	(a)–(c) of Figure 5

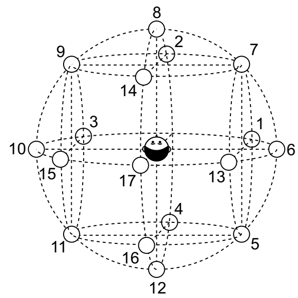

Figure 7. Relation between perceived directions and direction numbers.

Table 3. Accuracy rates and results of the chi-square test for a white noise in the localization test.

Number	Control condition	1 m distance	3 m distance
1	1.00	0.81*	0.76*
2	0.95	1.00	0.95
3	1.00	0.81*	0.86
4	0.95	0.76	0.90
5	1.00	1.00	1.00
6	1.00	0.43**	0.52**
7	1.00	1.00	1.00
8	1.00	0.33**	0.43**
9	1.00	1.00	1.00
10	1.00	0.52**	0.43**
11	1.00	1.00	0.95
12	1.00	0.48**	0.33**
13	1.00	0.90	0.90
14	0.95	1.00	1.00
15	1.00	0.76*	0.86
16	0.86	0.71	0.71
17	1.00	0.38**	0.33**

Table 4. Accuracy rates and results of the chi-square test for a speech in the localization test.

Number	Control condition	1 m distance	3 m distance
1	1.00	0.86	0.81*
2	0.86	0.76	0.86
3	1.00	1.00	1.00
4	0.90	0.81	0.71
5	0.95	0.95	0.95
6	1.00	0.43**	0.57**
7	1.00	1.00	0.90
8	1.00	0.67**	0.62**
9	1.00	0.76*	0.90
10	1.00	0.57**	0.52**
11	1.00	0.90	0.90
12	0.95	0.76	0.33**
13	1.00	0.76*	0.81*
14	1.00	0.95	0.95
15	0.95	0.71*	0.71*
16	0.86	0.67	0.52*
17	0.90	0.48**	0.52**

than those of the direction number in the speech. Thus, it is considered that the listeners localize the sound image towards the upper direction when the white noise rather than the speech is used.

For the direction number 10, the erroneous answer was 3. This implies that the listeners localize a sound image towards the downward direction when the sound image is produced from the direction on the left hand side of the listener.

In the case when the direction number was 12, the erroneous answers were 4, 16, and 17. Thus, it can be inferred that when a sound is produced behind the listeners, they localize the sound image blurred towards the vertical direction.

When the direction number was 17, the most common erroneous answer was 14. Thus, it can be considered that the listeners localize a sound image towards the forward direction when the sound image is produced from above the listener. It seems that the answer rates of the direction number 14 in the white noise is more than those of the direction number in the speech. Thus, it is considered that the listeners localize the sound image towards the forward direction when the white noise rather than the speech is used.

It should be noted that identical signals were played from four loudspeakers in all the five directions. Thus, it is considered that the blur and bias in sound images occurred due to phantom sources in the five directions.

3. Conclusions

In this study, we have proposed the 3D sound field reproduction system with eight loudspeakers to reproduce sound. Wave field synthesis allows us to reproduce a 3D sound field even when the number of loudspeakers is made very small in order to prevent the loudspeakers from appearing in the listener's field of vision. The auditory capability of the proposed system was evaluated by the localization test. It was found that a good performance was observed for twelve of the seventeen directions that were used in the test. Moreover, in future studies, we plan to develop a method to improve the localized accuracy of the remaining five directions and test its performance using the localization test.

References

[1] H. Fletcher, "Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: Basic requirement," *Bell Sys. Tech. J.*, vol. 13, no. 2, pp. 239–244, April 1934.

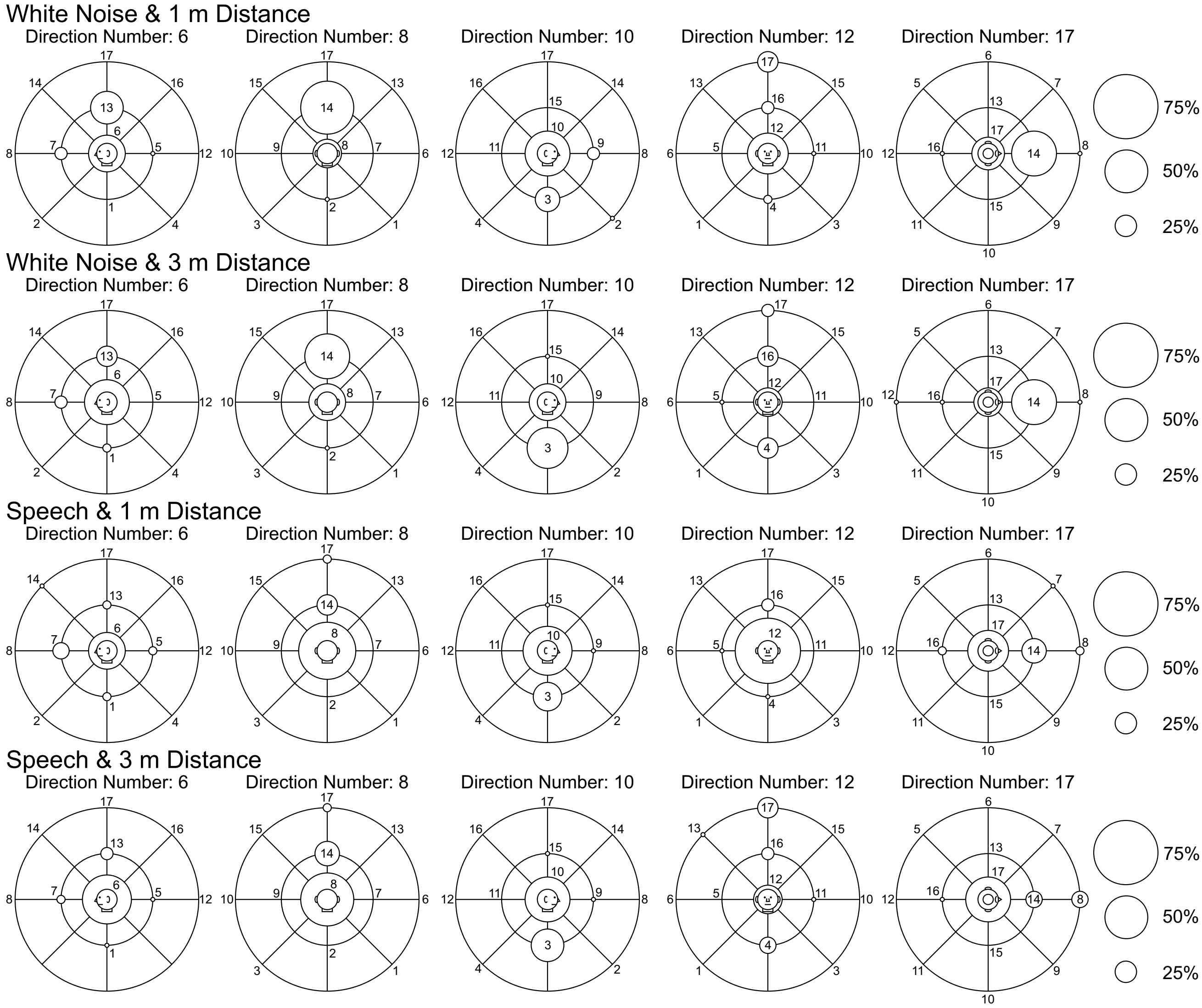

Figure 8. Answer rates of five presented directions in the localization test.

[2] M. Camras, "Approach to recreating a sound field," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1425–1431, June 1968.

[3] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[4] T. Kimura, K. Kakehi, K. Takeda, and F. Itakura, "Subjective assessments for the effect of the number of channel signals on the sound field reproduction used in wavefield synthesis," in *Proc. Int. Cong. Acoust*, Kyoto, Japan, April 2004, number Th.P1.18 in IV, pp. 3159–3162.

[5] T. Kimura and K. Kakehi, "Effects of directivity of microphones and loudspeakers in sound field reproduction based on wave field synthesis," in *Proc. Int. Cong. Acoust*, Madrid, Spain, September 2007, number RBA-15-011, pp. 1–8.

[6] J. Blauert, *Spatial Hearing*, pp. 372–392, MIT Press, Cambridge, Mass, revised edition, 1997.

[7] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.*, vol. 44, no. 9, pp. 683–705, September 1996.

[8] S. Ise, M. Toyoda, S. Enomoto, and S. Nakamura, "The development of the sound field sharing system based on the boundary surface control principle," in *Proc. Int. Cong. Acoust.*, September 2007, number ELE-04-003, pp. 1–7.

[9] Technical Centre of the European Broadcasting Union (EBU), *Sound Quality Assessment Material - Recordings for Subjective Tests*.