Subjective Effect of Synthesis Conditions in 3D Sound Field Reproduction System Using a Few Transducers and Wave Field Synthesis

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

ABSTRACT

In a conventional 3D sound field reproduction system using wave field synthesis, numerous loudspeakers are placed around the listener. However, since such a system is very expensive and loudspeakers are in the listener's field of vision, it is very difficult to construct an audio-visual virtual reality system. We have proposed a 3D sound field reproduction system using wave field synthesis and eight transducers, which are placed at the vertex of a cube. In this study, the effect of synthesis conditions on the localized perception was evaluated when the synthesis conditions, the directivity of microphones, and the size of cubic arrays, were varied. As a result, the performance of the localized perception was good when shotgun microphones were used and the size of arrays was that of a cube, measuring 0.4 m on each side.

Categories and Subject Descriptors

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—audio input/output; H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing—signal analysis, synthesis, and processing

General Terms

Experimentation, Human Factors, Performance.

Keywords

Sound field reproduction, Wave field synthesis, Microphone directivity

1. INTRODUCTION

Recently, several sound field reproduction techniques have been developed for auditory virtual reality systems. By the 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 societies will increase rapidly as these systems are capable of creating more realistic environments 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 using 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 if conventional sound field reproduction techniques such as binaural [6] and transaural [7] techniques are used.

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. In order to reproduce the sound field of a 3D space, the sound field reproduction system, in which numerous loudspeakers are placed around the listener and inverse filters are applied, is also proposed [8]. However, since these systems are very expensive to develop and the loudspeakers are visible in the listener's field of vision, it is very difficult to construct an audio-visual virtual reality 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 the 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.

We have proposed the 3D sound field reproduction system using wave field synthesis and eight transducers, which are

*He is now working in a FUJISOFT INCORPORATED.

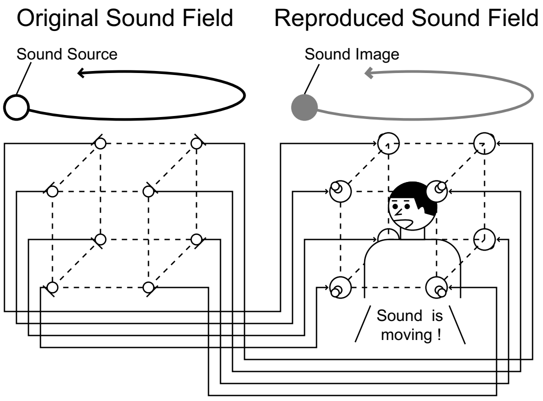

Figure 1: Diagram of proposed 3D sound field reproduction system.

placed at the vertices of a cube [9]. The proposed system reproduces 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 diagram of the proposed system is shown in Figure 1. First, a sound is captured using a cubic microphone array in the original sound field, as shown in the left-hand side of Figure 1. Second, the captured sound is replayed by the cubic loudspeaker array in the reproduced sound field, as shown in the right-hand side of Figure 1. As a result, the 3D sound field captured by the microphone array is reproduced by the loudspeaker array. Thus, as shown in the right-hand side of Figure 1, the listener, who is in the loudspeaker array, feels that sound is moving above his/her head when the sound is moving above the microphone array.

In a previous study [9], the localized capability of the proposed system was evaluated by the localization test. As a result, although it was indicated that the localized performance of 12 directions was good in the evaluated 17 directions, the localized performance of the remaining 5 directions was not good. However, the synthesis conditions to reproduce 3D sound fields, the directivity of microphones, and the size of cubic arrays, were fixed in the previous localization test. If these synthesis conditions were varied, the localized performance of the remaining 5 directions may be improved.

In this study, the effect of synthesis conditions on the localized perception is evaluated by the localization test and the synthesis condition in which the localized performance is improved is considered.

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°			

2. LOCALIZATION TEST

2.1 Synthesis of multichannel signals

The multichannel signals replayed by the loudspeaker array were synthesized on a computer. Since directional perception mainly depends on the direct sounds originating from a sound source, 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:

$$g_i(n) = \frac{1}{d_i}\delta\Big\{n - \mathrm{round}\Big(\frac{d_i F_s}{c}\Big)\Big\}, \tag{1}$$

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:

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

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

where d (=1, 3 m) denotes the distance between the sound source and the listening position, θ and ϕ are the azimuth and elevation angles, respectively, in the listening position, and A (=0.4, 0.5 m) denotes the size of the cubic microphone array. The values of θ and ϕ were set as shown in Table 1. These values are the same as those in the previous localization test [9].

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

$$\begin{aligned} x_i(n) &= D_i\{g_i(n) * s(n)\} \\ &= \frac{D_i}{d_i}s\Big\{n - \mathrm{round}\Big(\frac{d_i F_s}{c}\Big)\Big\}, \end{aligned} \tag{4}$$

where $*$ is the convolution. Previous studies have indicated that the sound is only recorded from outside the control area according to D_i (directivity of the ith microphone) [5]. Although D_i was set to one type (shotgun directivity) in the previous localization test [9], in the current test, D_i was set to two types, unidirectional and shotgun directivities as shown in Figure 2, as follows:

$$\text{(Unidirectional)} \qquad D_i = \frac{1+\cos\theta_i}{2}, \tag{5}$$

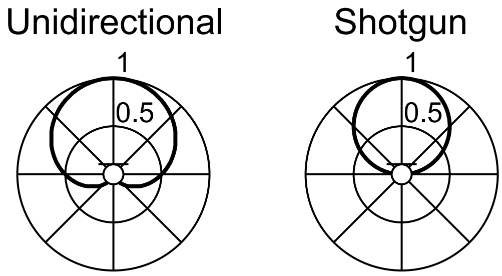

Figure 2: Directivity patterns of microphones in the localization test.

Table 2: Synthesis conditions used in the localization test.

Number	Microphone directivity D_i	Array size A
(i)	Unidirectional	0.4 m
(ii)	Unidirectional	0.5 m
(iii)	Shotgun	0.4 m
(iv)	Shotgun	0.5 m

$$\text{(Shotgun)} \qquad D_i = \begin{cases} \cos\theta_i & (|\theta_i| \leq 90^\circ) \\ 0 & (|\theta_i| > 90^\circ) \end{cases}, \tag{6}$$

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

$$\theta_i = \cos^{-1}\left\{ \frac{\mathbf{r}_i \cdot (\mathbf{r}_0 - \mathbf{r}_i)}{|\mathbf{r}_i||\mathbf{r}_0 - \mathbf{r}_i|} \right\}. \tag{7}$$

Four synthesis conditions used in the localization test are shown in Table 2. The synthesis condition (iii) is the same as that of the previous localization test [9].

2.2 Experimental environment

The localization test was performed in a room at a reverberation time of 115 ms. Twenty-five loudspeakers were placed in the positions as shown in Figure 3. The listening position was placed at the center of a sphere. The white loudspeakers indicate eight loudspeakers placed at the vertex of a cube with sides measuring 0.4 m or 0.5 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 as described below. 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: NSW1-205-8A suitable) on a loudspeaker box as shown in Figure 4. The setup of the loudspeaker array and loudspeakers for the control condition is shown in Figure 5. The white boxes in Figure 5 denote the manufactured loudspeakers. A background noise level was A-weighted level of 20 dB and the sound pressure level in the listening position was set to A-weighted level of 60 dB.

The five experimental conditions in the localization test are shown in Figure 6. In the control condition (a), the sound source signal $s(n)$ was replayed from one loudspeaker selected from a group of seventeen loudspeakers. As a result, listeners feel that there are sound sources in the positions of the loudspeakers, as shown in Figure 6(a). In other four conditions(b), (c), (d), and (e), eight channel signals $x_i(n)$ calculated by Eq. (4) were replayed from eight loudspeakers. It is to be noted that the gray lines of the microphones in the left-hand side of Figure 6(b)-(e) denote the directivity of microphones. As a result, listeners feel as if there are synthetic sound images in the positions occupied by the gray circles, as shown in the right-hand side of Figure 6(b)-(e). It is to be noted that two synthesis conditions, (i) and (ii), as shown in Table 2, are included in the two experimental conditions, (b) and (c), further, two synthesis conditions, (iii) and (iv), as shown in Table 2, are included in the two experimental conditions, (d) and (e).

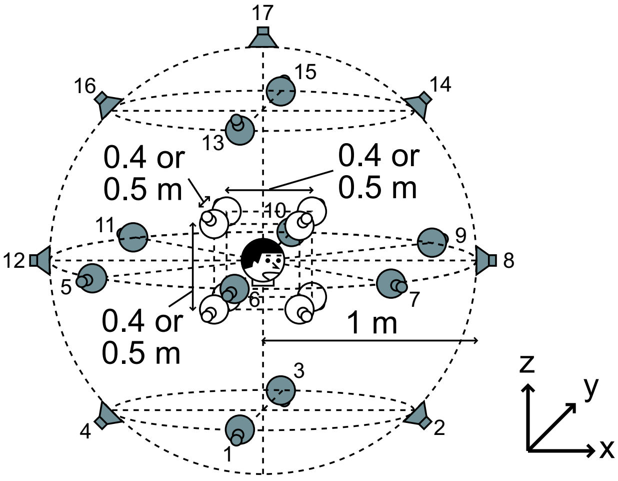

Figure 3: Position of a listener and the loudspeakers in the localization test.

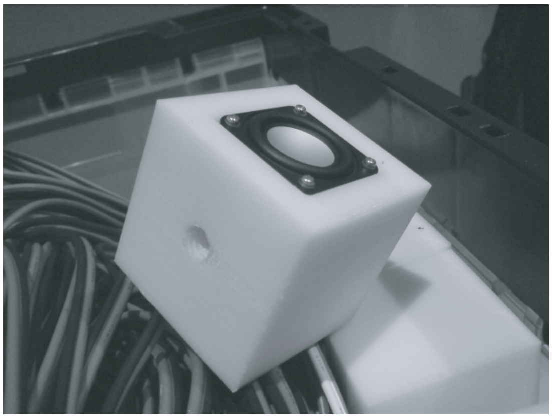

Figure 4: Image of manufactured loudspeakers.

2.3 Experimental procedure

Seven males and three females participated as listeners in this test. The flowchart of the localization test is shown in Figure 7. Because there was no difference between sound sources in the previous localization test [9] in which two sound sources (white noise and speech) were used, only one

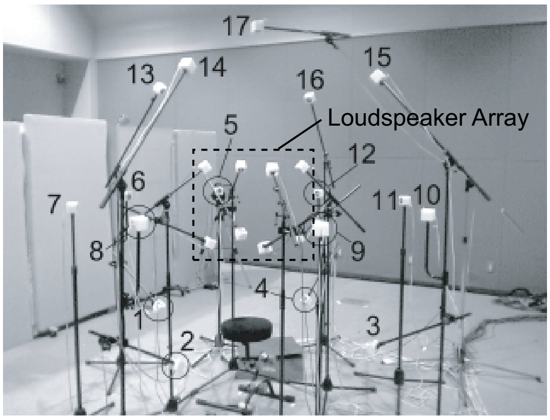

Figure 5: Setup of the loudspeaker array and loudspeakers for the control condition.

Table 3: Details of the practice and main trials in the localization test.

	Element	Note
Practice (34)	= 17 directions × 2 conditions	(a) and (b) of Figure 6
Main (170)	= 17 directions × 5 conditions × 2 repetitions	(a)–(e) of Figure 6

sound source (white noise) was used in this test. White noise was synthesized using MATLAB. The test was divided into two sessions for each array size shown in Table 2. The order of the presentation of the array sizes was randomized for each listener. Thirty-four practice trials and one hundred and seventy main trials were performed. During the main trials, rest periods were allowed after every set of 42 or 43 trials. The orders of the trials were randomized for each listener. The details of the practice and main trials are shown in Table 3.

The listeners were instructed to report the perceived direction of sound by listing the number of directions in an answer sheet. The relation between the perceived directions and the direction numbers is shown in Figure 8. The listeners were allowed to turn their heads freely while listening to the sounds.

2.4 Results and discussions

The effect of the synthesis conditions was evaluated by calculating the accuracy rates defined as follows:

$$\text{Accuracy rate} = \frac{\text{The number of correct answers}}{\text{The number of presentations}}. \qquad (8)$$

In order to evaluate the global effect of the synthesis conditions, the accuracy rates of all presented directions were calculated in each synthesis condition shown in Table 2. The accuracy rates of each synthesis condition are shown in Figure 9. The error bars in Figure 9 denote the 95% confidence intervals. From the result of comparison between the synthesis conditions in Figure 9, it is indicated that the performance of the synthesis condition (iii), in which the size of cubic arrays is 0.4 m on each side and shotgun microphones are applied, is good in the proposed system. However, because the accuracy rate of the synthesis condition (iii) is lower than that of the control condition, it is considered that the performance of the proposed system is not improved even if the synthesis condition is varied.

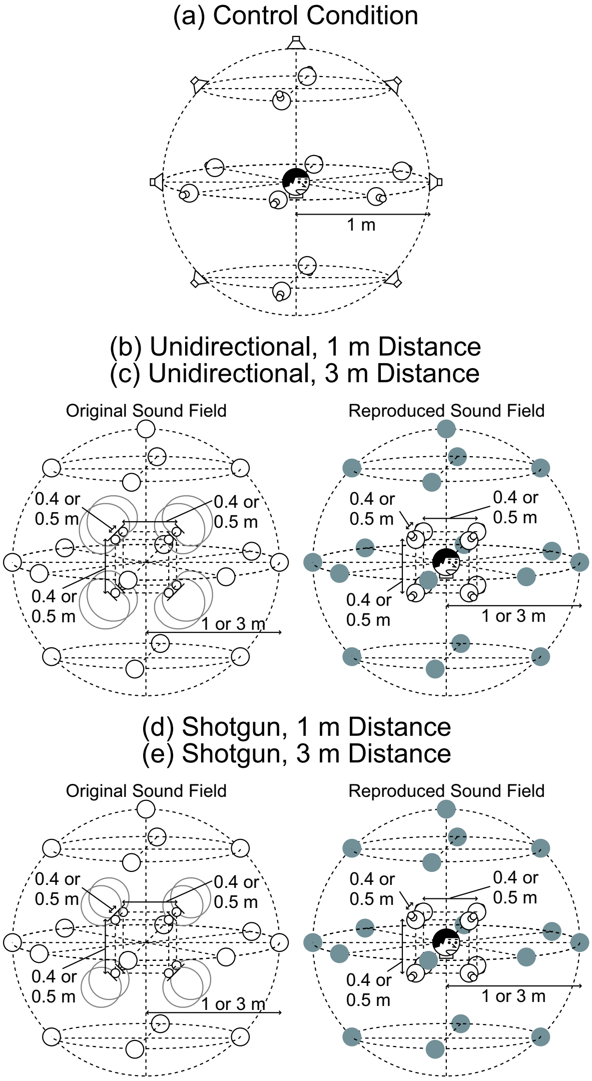

Figure 6: Five experimental conditions in the localization test.

In order to evaluate the presented direction in which the performance of the proposed system is not good enough, the accuracy rates of each presented directions were calculated and the chi-square test was performed. The accuracy rates and the results of the chi-square test for each presented direction are shown in Table 4. * and ** in Table 4 denote that there are significant differences of 5% and 1% levels between the control condition and the synthesis conditions by the chi-square test. It was observed that in six directions (4, 6, 8, 10, 12, and 17), the accuracy rates of all synthesis conditions were lower than those of the control condition

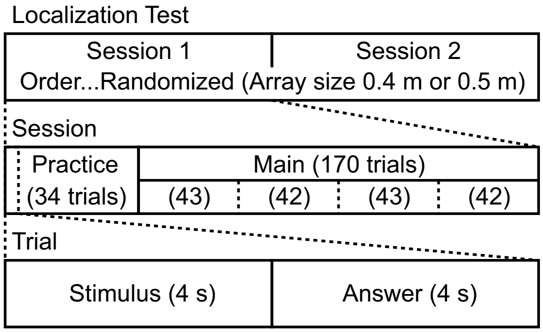

Figure 7: Flowchart of the localization test.

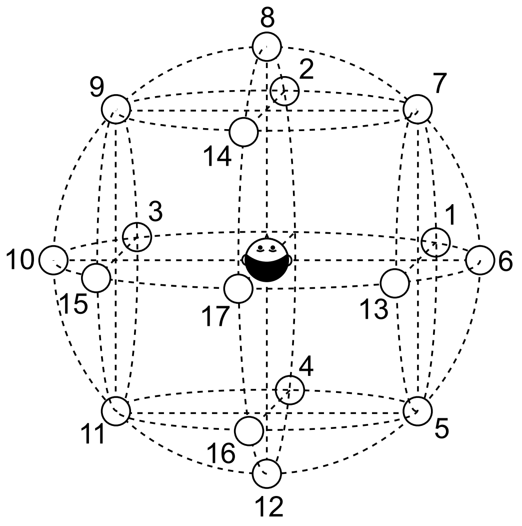

Figure 8: Relation between perceived directions and direction numbers.

because there were significant differences of the 1% level in all cases. Thus, it is considered that the performance of the proposed system is not good enough because the performance in six directions stated above is not good. On the other hand, in the other directions of the synthesis condition (iii), the accuracy rates were almost the same as those of the control condition since there were no significant differences of the 1% level in all cases. Thus, it is considered that the performance of the proposed system is good in all the directions, except in the six directions stated above if the size of cubic arrays is 0.4 m on each side and if shotgun microphones are applied in the proposed system.

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

$$\text{Answer rate} = \frac{\text{The number of answers}}{\text{The number of presentations}}. \qquad (9)$$

The results of the answer rates for the six presented directions (4, 6, 8, 10, 12, and 17) in the synthesis condition (iii) are shown in Figure 10. The sound is presented from the backward (the downward and backward in the direction

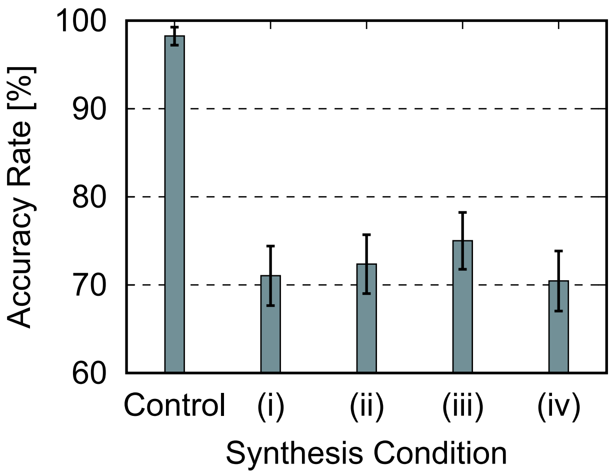

Figure 9: Accuracy rates of each synthesis condition.

number 4) direction of the heads placed at the center of two concentric circles. The numbers on the 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.

When the direction number was 4, the most common erroneous answer was 12. Thus, it can be inferred that when a sound image is produced from the region of the downward and backward directions of the listeners, the listeners localize the sound image toward the upper direction.

On the other hand, in the case of the four directions (6, 8, 10, and 12), the erroneous answers were the upper and downward directions of the presented directions. Thus, it can be inferred that when a sound is produced from the four directions (left, frontal, right, and backward) of the listeners, the listeners localize the sound image blurred toward the vertical direction. In these cases, it should be noted that identical signals were replayed from four loudspeakers. Thus, it is considered that the blurring of sound images occurred due to phantom sources synthesized by four loudspeakers in these cases.

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 toward the forward direction when the sound image is produced from above the listener.

The aim of this localization test was to evaluate the synthesis condition in which the localized capability of the proposed system is improved. However, the synthesis condition, in which the localized performance is the best, was the synthesis condition (iii), which is the same as that used in the previous localization test [9], and the auditory performance of the proposed systems was not improved.

3. CONCLUSIONS

In this study, in order to improve the auditory performance of the proposed system, the 3D sound field reproduction

Table 4: Accuracy rates and results of the chi-square test for synthesis conditions in the localization test.

No.	Control	(1)	(2)	(3)	(4)
1	100%	70%**	65%**	95%	88%*
2	98%	73%**	88%	85%*	88%
3	100%	75%**	55%**	88%*	65%**
4	98%	75%**	63%**	70%**	70%**
5	100%	98%	100%	100%	98%
6	100%	50%**	33%**	40%**	30%**
7	100%	100%	100%	100%	100%
8	100%	40%**	43%**	33%**	28%**
9	100%	100%	98%	98%	100%
10	100%	35%**	28%**	38%**	18%**
11	100%	95%	98%	95%	100%
12	100%	45%**	35%**	53%**	40%**
13	100%	80%**	88%*	90%*	83%**
14	98%	93%	90%	95%	95%
15	95%	78%*	88%	85%	90%
16	88%	50%**	55%**	65%*	78%
17	95%	53%**	65%**	48%**	73%**

system using eight transducers and wave field synthesis, the effect of synthesis conditions, the size of cubic array and the directivity of microphones, on the localized perception was evaluated. The localization test was performed under the four synthesis conditions. As a result, it was indicated that the localized performance of the proposed system was best when the size of the cubic arrays was 0.4 m on each side and when shotgun microphones were applied. However, the auditory performance of the proposed system was not improved.

In future works, we plan to develop a method to improve the auditory capability of the proposed system and evaluate its performance using the localization test.

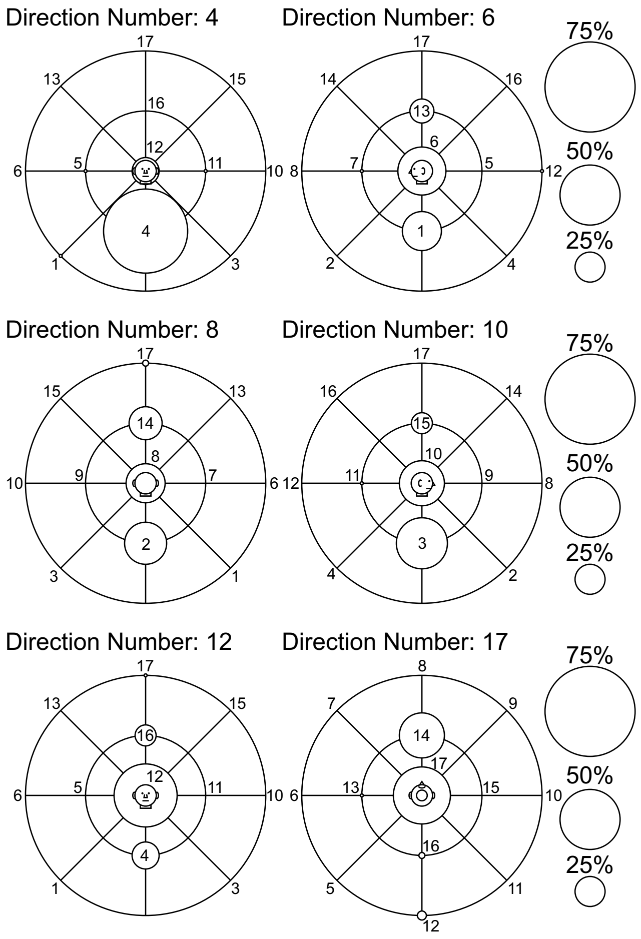

Figure 10: Answer rates of six presented directions for the synthesis condition (iii) in the localization test.

4. REFERENCES

[1] H. Fletcher. Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: Basic requirement. Bell Sys. Tech. J., 13(2):239–244, April 1934.

[2] M. Camras. Approach to recreating a sound field. J. Acoust. Soc. Am., 43(6):1425–1431, June 1968.

[3] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. J. Acoust. Soc. Am., 93(5):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, number Th.P1.18 in IV, pages 3159–3162, Kyoto, Japan, April 2004.

[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, number RBA-15-011, pages 1–8, Madrid, Spain, September 2007.

[6] J. Blauert. Spatial Hearing, pages 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., 44(9):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., number ELE-04-003, pages 1–7, September 2007.

[9] M. Naoe, T. Kimura, Y. Yamakata, and M. Katsumoto. Performance evaluation of 3D sound field reproduction system using a few loudspeakers and wave field synthesis. In Proc. Second ISUC, number 2-2, pages 36–41, Osaka, Japan, December 2008.