Personal 3D Sound Field Reproduction Technique for Remote Control Using Wave Front Synthesis and Eight Directional Microphones

Toshiyuki Kimura
Universal Communication Research Institute,
National Institute of Information and Communications Technology, Japan
{t-kimura@nict.go.jp}

Abstract

In this paper, a personal 3D sound field reproduction technique using wave front synthesis and eight directional microphones is proposed for the remote control system. The localized performance of the proposed technique was evaluated by the localization test. As a result, it was indicated that the localized performance of the proposed technique was enough to construct the audio-visual remote control system.

1. Introduction

3D sound field reproduction techniques are recently investigated. It is expected that these techniques can be applied to the remote control systems of disaster vehicles. It is preferable that the number of transmission channels is little in the 3D sound field reproduction techniques for the remote control systems.

There are binaural [1] and transaural [2] techniques in conventional 3D sound field reproduction techniques in which the number of transmission channels is little. However, the effect of listeners' individuality is serious in the binaural technique. The transaural technique is not applied to the remote control systems since the long-term delay occurs in this technique by the inverse filtering to cancel the cross talk of acoustic transmission paths. Thus, it needs to apply the 3D sound field reproduction technique in which the effect of listeners' individuality is not serious and the inverse filtering is not needed.

Wave front synthesis technique using directional microphones [3-4] is a 3D sound field reproduction technique for reproducing wave fronts from a control area in a different area (the listening area), according to the Kirchhoff-Helmholtz integral equation [5]. In this technique, the original sound in a control area is recorded using the array of directional microphones and it is then replayed in a listening area using a loudspeaker array. The arrays are placed at the boundaries of their respective areas. The positions of the directional microphones and the loudspeakers are identical in their respective areas. In this technique, listeners can freely listen to a sound in a listening area and the inverse filtering is not needed although the number of transmission channels is not little.

The number of transmission channels used by the technique can be reduced by considering the auditory capability of the listeners, even if the wave fronts are reproduced in the low-frequency range [6]. Thus, by restricting the listening area to the neighborhood of a listener's head and performing a listening test, a practical technique can be constructed using only the minimum required number of transmission channels. In particular, since the number of transmission channels reproducing directional perception was more than that of transmission channels reproducing spatial impression when the listening tests are done for each realistic sensation parameter [6] (directional perception, distant perception and spatial impression [7]), it needs to perform a localization test.

In this paper, the personal 3D sound field reproduction technique using wave front synthesis and eight directional microphones, in which the number of transmission channels is eight, is proposed. The performance of the proposed technique is evaluated by the localization test.

2. Diagram of proposed technique

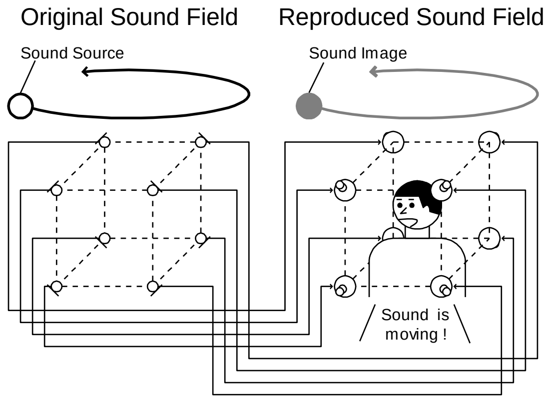

Figure 1 Proposed personal 3D sound field reproduction technique

The diagram of the proposed personal 3D sound field reproduction technique is shown in Figure 1. First, in the original sound field, eight directional microphones are set at the vertex of the control area of which the shape is cube and a sound is captured using eight directional microphones, as shown in the left-hand side of Figure 1. The directional microphones are then directed toward the outside of the control area. Second, in the reproduced sound field, the captured sound is replayed by eight loudspeakers, as shown in the right-hand side of Figure 1. The position of each loudspeaker is the same as that of each directional microphone. Note that the listener's field of vision in the horizontal direction is not prevented by loudspeakers because loudspeakers are not placed in the horizontal plane of the listener. As a result, because the 3D sound field captured by eight directional microphones is reproduced in the cubic loudspeaker array, the listener, who is in the cubic loudspeaker array, feels that sound is moving above his/her head when the sound is moving above the cubic microphone array, as shown in the right-hand side of Figure 1.

3. Localization test

3.1. Synthesis of multichannel signals

The multichannel signals replayed by the cubic 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 *i*th microphone (i = 1...8), $g_i(n)$, is denoted as follows:

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

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 *i*th microphone. The values of $\mathbf{r}_0$ and $\mathbf{r}_i$ (position vectors of the sound source and the *i*th microphone, respectively) were set as follows:

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

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

where d (=1 and 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. A denotes the size of the cubic arrays. It needs to make the size smaller according to the condition in which wave fronts are accurately reproduced. In this paper, the array size A was set to two conditions (0.4 and 0.5 m).

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

Index	θ	ϕ	Index	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°	---	---	---

A sound source was white noise synthesized using MATLAB. If the sound source signal is represented by $s(n)$, then $x_i(n)$, which represents the channel signals recorded by the *i*th microphone, is denoted as follows:

$$x_i(n) = D_i\{g_i(n) * s(n)\} = \frac{D_i}{d_i} s\left\{n - \mathrm{round}\left(\frac{d_i F_S}{c}\right)\right\}, \quad (4)$$

where * is the convolution. Previous studies have indicated that the sound is only recorded from outside the control area according to D_i (directivity of the *i*th microphone) [4]. In this paper, D_i was set to two conditions, unidirectional and shotgun directivities as shown in Figure 2, as follows:

$$\text{(Unidirectional)} \quad D_i = (1 + \cos\theta_i)/2$$
$$\text{(Shotgun)} \quad D_i = \begin{cases} \cos\theta_i & (\theta_i \le 90^\circ) \\ 0 & (\theta_i > 90^\circ) \end{cases}, \quad (5)$$

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

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

Total four synthesis conditions used in the localization test are shown in Table 2.

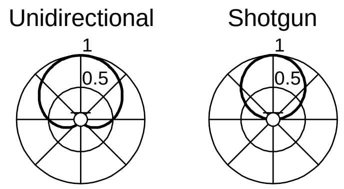

Figure 2 Directivity patterns of microphones in the localization test

Table 2 Synthesis conditions in the localization test

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

3.2. 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. 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.

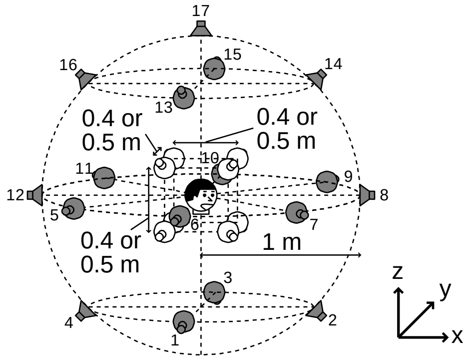

Figure 3 Position of a listener and the loudspeakers in the localization test

The five experimental conditions in the localization test are shown in Figure 4. 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 4(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 4(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 4(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).

3.3. Experimental procedure

Seven males and three females participated as listeners in this test. The flowchart of the localization test is shown in Figure 5. The test was divided into two sessions for each array size. 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.

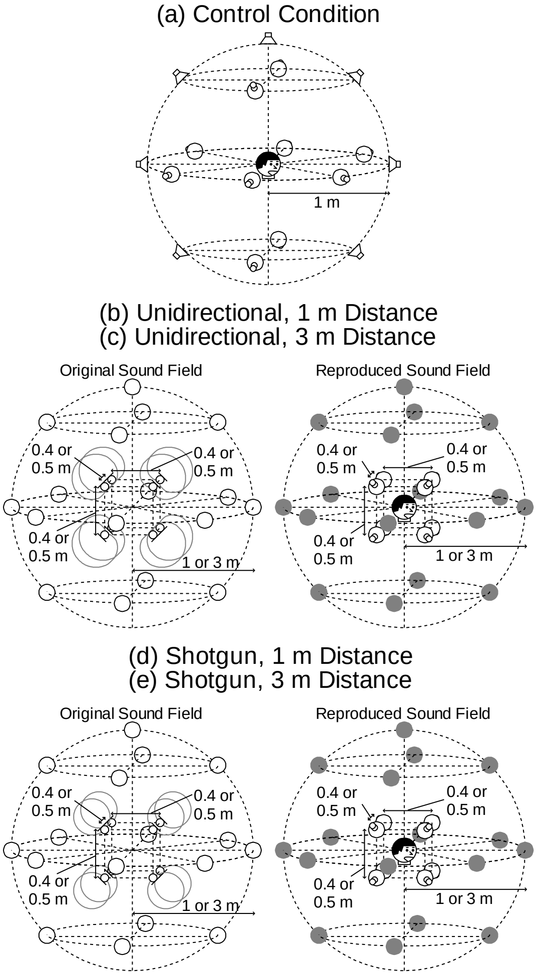

Figure 4 Five experimental conditions in the localization test

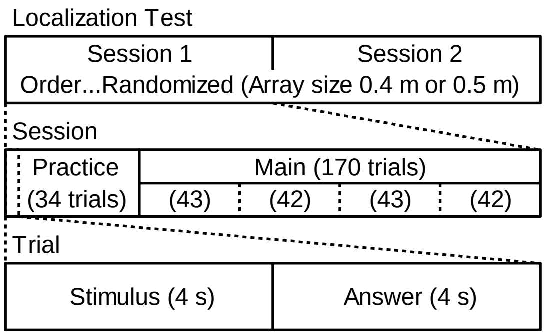

Figure 5 Flowchart of the localization test

Table 3 Practice and main trials in the localization test

	Element	Note
Practice (34)	= 17 directions × 2 conditions	(a) and (b) of Figure 4
Main (170)	= 17 directions × 5 conditions × 2 repetitions	(a)-(e) of Figure 4

The listeners were instructed to report the perceived direction of sound by listing the number of directions in an answer sheet. The listeners were allowed to turn their heads freely while listening to the sounds.

4. Result and discussion

4.1. Front-back confusion

In conventional personal 3D sound field reproduction technique such as binaural and transaural techniques, front-back confusions occur in the localization test if individual head-related transfer functions are not applied [8]. Thus, in order to evaluate the effect of listeners' individuality in the proposed technique, front-back confusions are firstly analyzed.

In this paper, the number of front-back confusions was counted from the result of the localization test. The presented direction pairs of front-back confusions in the localization test are listed as follows:

- 2 (θ = 0°, ϕ = -45°) ↔ 4 (θ = 180°, ϕ = -45°),
- 5 (θ = -135°, ϕ = 0°) ↔ 7 (θ= -45°, ϕ = 0°),
- 8 (θ = 0°, ϕ = 0°) ↔ 12 (θ = 180°, ϕ = 0°),
- 9 (θ = 45°, ϕ = 0°) ↔ 11 (θ = 135°, ϕ = 0°),
- 14 (θ = 0°, ϕ = 45°) ↔ 16 (θ = 180°, ϕ = 45°).

Results of the number of front-back confusions are shown in Table 4. The rates of front-back confusions are very low (0.25%) in the control condition. Thus, it was indicated that listeners could accurately localize sound sources of all directions.

The rates of front-back confusions are less or equal than 1.5% in eight other conditions. Chi-square tests were performed between the control condition and eight other conditions. As a result, there were no significant differences between the control condition and eight other conditions. Thus, it was indicated that the front-back confusions of the proposed technique were comparable in those of the control condition in this localization test. This is due to fact that the listeners are allowed to turn their heads freely while listening to the sounds.

Table 4 Results of front-back confusions in the localization test

Condition	Distance	Number(Rate)
Control	---	1/400(0.25%)
(i)	1 m	3/200(1.50%)
	3 m	2/200(1.00%)
(ii)	1 m	0/200(0.00%)
	3 m	1/200(0.50%)
(iii)	1 m	0/200(0.00%)
	3 m	0/200(0.00%)
(iv)	1 m	0/200(0.00%)
	3 m	1/200(0.50%)

4.2. Perceived direction error

Although it is most common to evaluate the result of the localization test by the error of azimuth and elevation angles, the resolution of the azimuth and elevation angles is not constant in the spherical surface. On the other hand, listeners turned their heads toward the perceived direction when they listened to the sounds in the localization test. Thus, if the presented directions in the localization test are transformed to the frontal direction of listeners, it is considered that the result of the localization test can be evaluated at the constant resolution because the perceived directions in the localization test are converged to the frontal direction of listeners.

In this paper, in order to evaluate the result of the localization test at the constant resolution, the azimuth and elevation angles of the perceived directions (θ and ϕ) are transformed to the horizontal and vertical angles (θ' and ϕ') according to following equation:

$$\theta' = \tan^{-1}\frac{y'}{x'}, \quad \phi' = \sin^{-1}\frac{z'}{\sqrt{x'^2+y'^2+z'^2}}, \tag{7}$$

where x', y' and z' (three-dimensional coordinates of transformed perceived directions) are defined as follows:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{R}_y(-\phi_0)\mathbf{R}_z(-\theta_0)\begin{pmatrix} \cos\theta\cos\phi \\ \sin\theta\cos\phi \\ \sin\phi \end{pmatrix},$$
$$\mathbf{R}_y(-\phi_0) = \begin{pmatrix} \cos\phi_0 & 0 & \sin\phi_0 \\ 0 & 1 & 0 \\ -\sin\phi_0 & 0 & \cos\phi_0 \end{pmatrix}, \tag{8}$$
$$\mathbf{R}_z(-\theta_0) = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 & 0 \\ -\sin\theta_0 & \cos\theta_0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where θ_0 and ϕ_0 denote the azimuth and elevation angles of presented direction. Note that the data of transformed horizontal angles θ' was not counted when front-back confusions occur or the transformed vertical angles ϕ' become 90°.

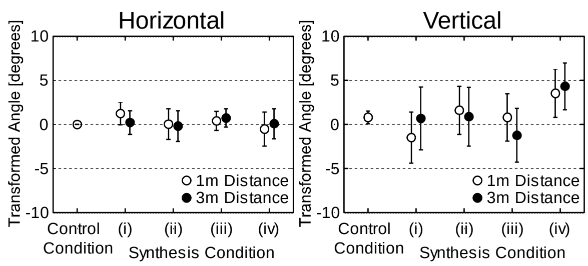

Figure 6 Results of transformed angles in control and synthesis conditions (Left: horizontal angles, Right: vertical angles)

Results of transformed angles are shown in Figure 6. Error bars denote the 95% confidential interval of average angles. In the transformed horizontal and vertical angles, analysis of variance (ANOVA) and multiple comparisons between the control condition and eight other conditions were performed. As a result, there were no significant differences between the control condition and eight other conditions in the transformed horizontal angle. On the other hand, there were no significant differences between the control condition and seven other conditions except the condition (synthesis condition (iv), 3 m distance) in the transformed vertical angle. Thus, it is indicated that the localized performance of the proposed technique is enough to construct the auditory remote control system except the condition (synthesis condition (iv), 3 m distance).

However, in the condition (synthesis condition (iv), 3 m distance), the value of average angles in the transformed vertical angle is 4.32°. This value is less than that of the difference limen of the expert in the front direction in the ventriloquism effect (about 11° [9]). Thus, it is indicated that the localized performance of the proposed technique was enough to construct the audio-visual remote control system.

5. Conclusions

In this paper, for the remote control systems, the personal 3D sound field reproduction technique using wave front synthesis and eight directional microphones was proposed. Eight directional microphones and loudspeakers are set at the vertex of a cube placed around a listener's head in the proposed technique. In order to evaluate the localized performance of the proposed technique, the localization test was performed. The result of the localization test was analyzed based on the transformed horizontal and vertical angles. As a result, it was indicated that the localized performance of the proposed technique was enough to construct the audio-visual remote control system since the errors of transformed angles are always less than or equal to 4.32°.

A subject for future work is the development of the audio-visual remote control system using the proposed technique and the glasses-free 3D video display technique (e.g., [10]). It also needs to evaluate the effect of the developed system.

Acknowledgements

The author would like to thank Mr. M. Naoe for performing the localization test. The localization test in this paper was performed according to the approval of the ethical committee of National Institute of Information and Communications Technology (NICT), Japan.

References

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

[2] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, "Comparative study of european concert halls: Correlation of subjective preference with geometric and acoustic parameters," *J. Acoust. Soc. Am.*, vol. 56, no. 4, pp. 1195–1201, October 1974.

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

[4] 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. 1–6.

[5] B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle*, pp. 23–26, Oxford University Press, London, UK, second edition, 1950.

[6] 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, pp. 3159–3162.

[7] M. Morimoto, "The relation between spatial impression and the precedence effect," in *Proc. Int. Conf. Auditory Display*, Kyoto, Japan, July 2002, number SS#2-1, pp. 297–306.

[8] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)*, vol. 1, no. 3, pp. 167–174, July 1980.

[9] S. Komiyama, "Subjective evaluation of angular displacement between picture and sound directions for HDTV sound systems," *J. Audio Eng. Soc.*, vol. 37, no. 4, pp. 210–214, April 1989.

[10] M. Kawakita, S. Iwasawa, S. Gurbuz, and N. Inoue, "Development of glasses-free 3D video system," in *Proc. Int. Univ. Comm. Symp.*, Gumi, South Korea, October 2011, pp. 323–327.