3-D Sound Reproduction System for Immersive Environments Based on the Boundary Surface Control Principle

Seigo Enomoto[1], Yusuke Ikeda[1], Shiro Ise[2], and Satoshi Nakamura[1]

[1] Spoken Language Communication Group, National Institute of Information and Communications Technology, 3-5 Hikaridai, Keihanna Science City, 619-0289, Japan
[2] Graduate school of engineering, Department of Architecture and architectural engineering, Kyoto University, C1-4-386 Kyotodaigaku-katsura, Nishikyo-ku, Kyoto, 615-8540, Japan
{seigo.enomoto,yusuke.ikeda,satoshi.nakamura}@nict.go.jp, ise@archi.kyoto-u.ac.jp

Abstract. We constructed a 3-D sound reproduction system containing a 62-channel loudspeaker array and 70-channel microphone array based on the boundary surface control principle (BoSC). The microphone array can record the volume of the 3-D sound field and the loudspeaker array can accurately recreate it in other locations. Using these systems, we realized immersive acoustic environments similar to cinema or television sound spaces. We also recorded real 3-D acoustic environments, such as an orchestra performance and forest sounds, by using the microphone array. Recreated sound fields were evaluated by demonstration experiments using the 3-D sound field. Subjective assessments of 390 subjects confirm that these systems can achieve high presence for 3-D sound reproduction and provide the listener with deep immersion.

Keywords: Boundary surface control principle, Immersive environments, Virtual reality, Stereophony, Surround sound.

1 Introduction

Stereophony is one of the primary factors in improving the sense of immersion in movies or television. Recent years have seen the emergence of surround sound systems with 5.1-channel or greater loudspeakers beyond movie theaters and into the home. Surround sound listeners can achieve a feeling as if they are in the actual places to which they are listening. Traditional surround systems, however, cannot reconstruct sound wavefronts that radiate in the actual environment. If a system can reconstruct the wavefront, it can recreate a fully immersive, rather than surrounding, environment. Since the hearing information can be obtained from all directions, the importance of this information is particularly increased in applications where many people in distance places communicate with each other.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part I, HCII 2011, LNCS 6773, pp. 174–184, 2011.

The Kirchhoff-Helmholtz integral equation (KHIE) is the theoretical basis of 3-D sound reproduction systems to record and reproduce sound fields, and this sort of reproduction has a long history. In the early 1930s, Fletcher and colleagues reported that ideal sound field recording and reproduction could be achieved by using numerous microphones and loudspeakers with acoustical transparency[3,4]. Steinberg et al. also confirmed in subjective experiments that stereophonic sound could be represented using only three loudspeakers[13]. A three-loudspeaker system cannot reconstruct a wavefront but can serve as a basis for surround sound.

Field or 3-D sound reproduction is, however, more attractive. For correctly reconstructing wavefronts, as compared to conventional surround systems, many technologies have been developed and their theoretical basis have been studied. In 1993, Berkhout et al. proposed Wave Field Synthesis (WFS) [1] as a 3-D sound field reproduction system based on the KHIE, and IOSONO [8] is a commercial application based on WFS. The theoretical basis of WFS is, however, the Rayleigh integral equation. The KHIE is not used directly; therefore, in the WFS an infinite plain boundary is considered explicitly. However, since the finite length loudspeaker array is applied to practical WFS, it is difficult to reproduce a 3-D sound field with no artifacts due to the truncation. Moreover, practical loudspeakers must be placed on the boundary instead of an ideal monopole source. An approximation of an ideal source by means of a loudspeaker causes other artifacts, especially in a higher frequency range of its own properties. Ambisonics [14] is another application of 3-D sound reproduction systems based on the wave equation. Ambisonic-based 3-D sound reproduction systems require higher order spherical harmonics to accurately reproduce 3-D sound fields, this they have to date been difficult to construct. It is difficult to extend the so-called sweet spot in ambisonics.

In contrast, Ise in 1997 proposed the boundary surface control principle (BoSC) [6,7]. By integrating the KHIE and inverse system, the BoSC can accurately reproduce a 3-D sound field surrounded by a closed surface. According to the BoSC, it is not necessary to place ideal monopole and dipole sources on the boundary. Therefore an approximation of such sources is also not required. There is also no restriction on the loudspeaker position; it would be located in an exterior area for the boundary.

Consequently, in our research we constructed 3-D sound reproduction systems that contain a 70-channel microphone array and 62-channel loudspeaker arrays. In this manuscript, we describe the results of subjective assessment to confirm the "presence" recreated by the BoSC-based 3-D sound field reproduction system.

2 Boundary Surface Control Principle

The BoSC is the theory for reproducing a 3-D sound field, and is now applied to the field of active noise control[10] or steering the directionality of a loudspeaker array[2]. This section describes the BoSC as an application of 3-D sound reproduction. Fig. 1 expresses 3-D sound field reproduction based on the BoSC. The

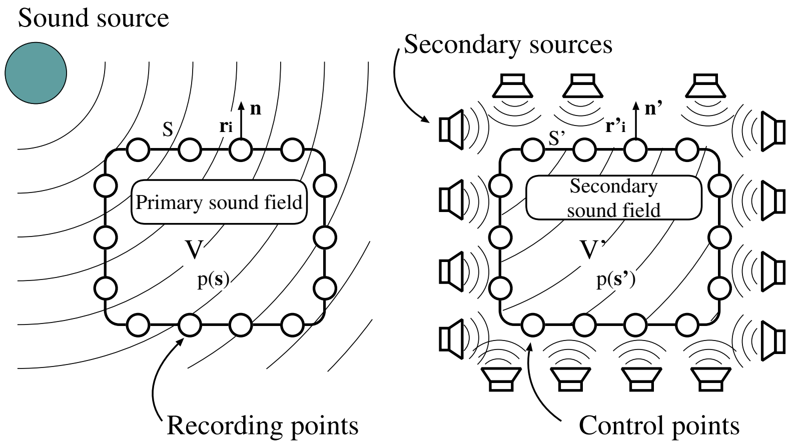


Fig. 1. 3-D sound field reproduction based on the BoSC. The left figure shows the primary sound field to be recorded e.g., concert hall. The right figure shows the secondary sound field where the recorded primary sound field is reproduced.

figure on the left shows the primary sound field to be recorded; e.g., a concert hall. On the right is the secondary sound field where the recorded primary sound field is reproduced; i.e., a listening room. The 3-D sound reproduction system based on the KHIE aims to record sound field V bounded by surface S and reproduce it into V' bounded by S'. Using the KHIE, the complex sound pressure of $p(\mathbf{s})$ and $p(\mathbf{s}')$ where $\mathbf{s} \in V$ and $\mathbf{s}' \in V'$ are the evaluation points is given by

$$p(\mathbf{s}) = \int_S \left\{ -j\omega\rho_0 G(r|s)v_n(\mathbf{r}) - p(\mathbf{r})\frac{\partial G(\mathbf{r}|\mathbf{s})}{\partial n} \right\} dS \quad , \tag{1}$$

$$p(\mathbf{s}') = \int_{S'} \left\{ -j\omega\rho_0 G(\mathbf{r}'|\mathbf{s}')v_n(\mathbf{r}') - p(\mathbf{r}')\frac{\partial G(\mathbf{r}'|\mathbf{s}')}{\partial n'} \right\} dS' \quad , \tag{2}$$

where ω is an angular frequency. ρ_0 is density of the medium. $p(\mathbf{r})$ and $v_n(\mathbf{r})$ are sound pressure and normal-outward particle velocity on the boundary, respectively. $G(\mathbf{r}|\mathbf{s}) = \frac{e^{-j\frac{\omega}{c}|\mathbf{r}-\mathbf{s}|}}{4\pi|\mathbf{r}-\mathbf{s}|}$ is a free-space Green's function [11,15]. For notational simplicity, angular frequency ω in equations is omitted. The free-space Green's function $G(\mathbf{r}|\mathbf{s})$ is explicitly defined by $\mathbf{r}$ and $\mathbf{s}$. Therefore, if the shape of boundary S is identical to S', the free-space Green's function $G(\mathbf{r}|\mathbf{s})$ is also identical to $G(\mathbf{r}'|\mathbf{s}')$. Consequently, if $p(\mathbf{r})$ and $v_n(\mathbf{r})$ are equal to $p(\mathbf{r})$ and $v_n(\mathbf{r})$, $p(\mathbf{s})$ is also equal to $p(\mathbf{s}')$.

To equalize $p(\mathbf{r})$ and $v_n(\mathbf{r})$ with $p(\mathbf{r}')$ and $v_n(\mathbf{r}')$ respectively, the BoSC system employs the secondary loudspeakers. The output signal of the loudspeakers is determined by the convolutions of the recorded sound signal and the inverse system. The inverse system is computed to equalize the room transfer function between each loudspeaker and microphone.

3 3-D Sound Reproduction System

3.1 BoSC in Practice : Reproducing 3-D Sound Field from Recorded Sound Pressure on the Boundary

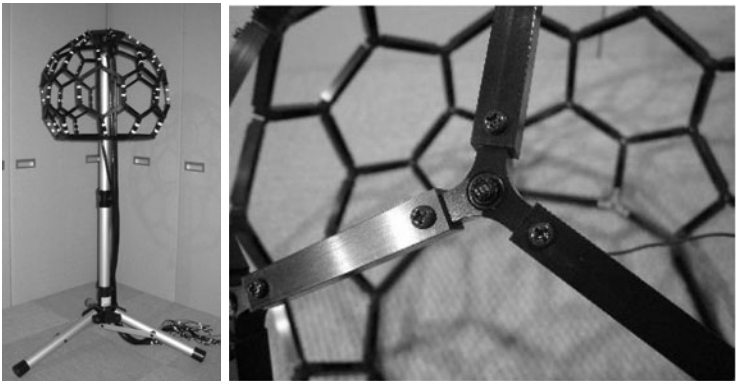
The 3-D sound reproduction system based on the KHIE theoretically requires measurements of the particle velocity on the boundary. It is well known that the particle velocity can be measured by using the sound pressure positioned with two points that intersect the boundary [11]. However, in this case, since the doubled record/control points are required, there is huge computational cost. Therefore we constructed a 3-D sound reproduction system that can record and reproduce only the sound pressure on the boundary. The Dirichlet Green's function $G_D(\mathbf{r}|\mathbf{s})$ can be used [15]. Substituting $G_D(\mathbf{r}|\mathbf{s})$ into equations (1) and (2), the first item of right-hand of these equations can be eliminated. Note that it is difficult to derive the exact value of the Dirichlet Green's function $G_D(\mathbf{r}|\mathbf{s})$, but it is not required in the BoSC system. The BoSC can assume that $G_D(\mathbf{r}|\mathbf{s})$ and $G_D(\mathbf{r}'|\mathbf{s}')$ are constants if the shape of the boundaries S is identical to S'. Therefore, if the sound pressure is recorded on boundary S and reproduced on boundary S', $p(\mathbf{s})$ is also reproduced in volume V'. Note, however, that 3-D sound fields at the natural frequency of a closed surface cannot be reproduced in the Dirichlet boundary condition.

3.2 Microphone Array for 3-D Sound Field Recording

As boundary S and S' depicted in Fig. 1, we presumed that the microphones should be distributed at regular intervals. To construct a microphone array of this shape, we designed it based on the C_{80} fullerene structure. The constructed array is shown in Fig. 2. Its diameter is around 46 cm. Omni-directional microphones (DPA 4060-BM) are installed on each node of the fullerene. Ten microphones located on the bottom of the fullerene are also eliminated to insert the head of the Head And Torso Simulator (HATS) or the subjects in Fig. 2. Therefore there are 70 nodes. Since the maximum and minimum interval of each microphone is respectively around 16 cm and 8 cm, the system can reproduce a frequency range up to 2 kHz. The system, however, aims to create immersive environments and have people feel the presence of other people or places. Therefore we did not limit the frequency range to below 2 kHz in the demonstration experiment. In the experiment we also aimed to evaluate 3-D sound fields that contain a frequency signal over 2 kHz.

3.3 Loudspeaker Array and Sound Reproduction Room

As the secondary sound sources depicted in Fig. 1, we designed the loudspeaker array with a dome structure consisting of four wooden layers and supported by the four wood columns. Six, 16, 24, and 16 full-range loudspeakers (Fostex FE83E) were installed on each wood layer. We presumed that the height



(a) (b)

Fig. 2. BoSC-based 3D sound reproduction system : 70-channel microphone array in which omni-directional microphones are installed on every node. (a) 70-channel mic. array. (b) Omni-directional mic. installed.

of the third layer is as almost same as the center of the microphone array. To compensate for the lower frequency responses of full-range loudspeakers, two sub-woofer loudspeakers (Fostex FW108N) were installed on each wood column. However, we employed 62 full-range loudspeakers for the design of the inverse system. Though the minimum resonance frequency of a full-range loudspeaker is 127 Hz, the 3-D sound reproduction system we constructed can produce from 80 Hz by using the appropriate inverse system. We therefore employed the subwoofer loudspeaker only for below the 80 Hz frequency range. The loudspeaker array is shown in Fig. 3. The loudspeaker array was constructed in a soundproofed room (Yamaha Woodybox: sound insulation level $D_r = 30$; inside dimensions: 1,203 mm $\times$ 1,646 mm $\times$ 2,164 mm) to reduce the disturbance background noise. To reduce the reverberation in the soundproofed room, we attached sound absorbing sponge to each interior wall.

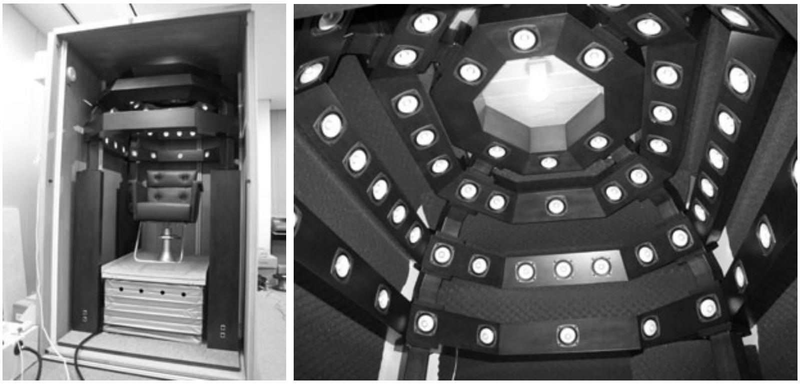
3.4 Design of the Inverse System

In Fig. 1, we presume that $\mathbf{X}(\omega)$ is a complex pressure recorded on boundary S in the primary sound field, $\mathbf{Y}(\omega)$ is a radiation signal from the loudspeaker in the secondary sound field, and $[\mathbf{G}(\omega)]$ is the impedance matrix of the transfer function between each loudspeaker and microphone pair. Complex pressure $\mathbf{Z}(\omega)$ measured on boundary S' in the secondary sound field then satisfies Equation (3).

$$\mathbf{Z}(\omega) = [\mathbf{G}(\omega)]\mathbf{Y}(\omega) \tag{3}$$

Therefore, Equation (4) is required to be $\mathbf{Z}(\omega) = \mathbf{X}(\omega)$.

$$\mathbf{Y}(\omega) = [\mathbf{G}(\omega)]^{+}\mathbf{X}(\omega) \tag{4}$$



(a) (b)

Fig. 3. BoSC-based 3D sound reproduction system: 70-channel loudspeaker array which 62 full-range loudspeakers and eight subwoofer loudspeakers installed. Only 62 full-range loudspeakers are used to render a wavefront. (a) Loudspeaker array constructed in the soundproofed room; (b) Dome structure of the loudspeaker array

where, $[\cdot]^+$ represents pseudo-inverse matrix. In addition,

$$\begin{aligned}\mathbf{X}(\omega) &= [X_1(\omega), \cdots, X_N(\omega)]^T,\\ \mathbf{Y}(\omega) &= [Y_1(\omega), \cdots, Y_M(\omega)]^T,\\ \mathbf{Z}(\omega) &= [Z_1(\omega), \cdots, Z_N(\omega)]^T,\end{aligned} \qquad [\mathbf{G}(\omega)] = \begin{bmatrix} G_{11}(\omega) & \cdots & G_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ G_{N1}(\omega) & \cdots & G_{NM}(\omega) \end{bmatrix}, \quad (5)$$

where, $[\cdot]^T$ represents transpose, the number of microphones is $N = 70$, and the number of loudspeakers is $M = 62$ in this manuscript. Therefore, by using the left inverse matrix, $[\mathbf{G}(\omega)]^+$ can be given as

$$[\mathbf{G}(\omega)]^+ = ([\mathbf{G}(\omega)]^\dagger [\mathbf{G}(\omega)] + \beta(\omega)\mathbf{I}_M)^{-1} [\mathbf{G}(\omega)]^\dagger, \quad (6)$$

where, $[\cdot]^\dagger$ represents conjugate transpose, $\beta(\omega)$ represents the regularization parameter, and $\mathbf{I}_M$ is the unit matrix with order M. An appropriate regularization parameter can reduce instabilities of $([\mathbf{G}(\omega)]^\dagger [\mathbf{G}(\omega)])$. We determined the parameters in each octave frequency band heuristically. We also presume the center of each frequency band is 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz, and 16 kHz.

The transfer impedance matrix $[\mathbf{G}(\omega)]$ is measured by using a 2^{17} points swept-sine signal experimentally in advance. Therefore, the inverse matrix only contains the inversion of the measured transfer function and is not the correct inversion in the reproducing circumstance. To determine $[\mathbf{G}(\omega)]^+$ as it can compensate for the fluctuations and time variants of the transfer function, we can employ adaptive signal processing. Many adaptive signal processing methods for the MIMO inverse system have been proposed and applied in the WFS system to compensate for reverberations in a listening room [9,12,5]. Almost all algorithms can be applied into the BoSC system. However, such adaptive algorithms

(a) (b)

Fig. 4. Recording of 3-D sound field data: (a) Orchestra with two microphone arrays are located in the auditorium and in front of the conductor during playing of Beethoven's Symphony No. 7. (b) Forest sounds which contains air-plane, hamming bird, singing of insects, footsteps, conversational voice, and so on.

with 62 inputs and 70 outputs require huge computational complexity. We therefore assumed in this manuscript that the instabilities caused by fluctuations and time variants of the transfer impedance matrix can be reduced by using the appropriate regularization parameters in this manuscript.

4 Demonstration Experiments

4.1 3-D Sound Field Recording

To demonstrate the performance of the 3-D reproduction system based on BoSC, we recorded real 3-D sound environments. The orchestra and forest sounds were reproduced in the subjective assessments described in the next section. The recording environments are shown in Fig. 4. The recordings of 3-D sound sources were carried out with 48 kHz sampling frequency and 24-bit quantization bit depth. In the recording of the orchestra, we employed two microphone arrays; one located in the auditorium and another in front of the conductor. Providing the visual information supplementally in the demonstration, we carried out video recording at the same position as 3-D field recording.

4.2 Subjective Assessment

The demonstration experiments were conducted to evaluate the performance of the BoSC-based sound reproduction system. The reproduced sound fields are listed in Table 1. Three kinds of nature sound (A) recorded in forests, and an orchestra performance (B and C) were employed in the demonstration. The performance in (B) corresponds to (A). The demonstration was limited to around five minutes. The recorded video was shown on an LCD monitor while the orchestra

Table 1. Reproduced 3-D sound field data

		Feature sound/Location	Time [sec.]
Forest sound	A	airplane	70.0
		conversing voices	37.5
		footsteps	22.5
Orchestra	B	in auditorium	67.0
	C	in front of conductor (stage)	67.0

Table 2. Questionnaire entries for the demonstration of the 3-D sound field reproduction

Comprehensive feeling for the reproduced 3-D sound field
A. Did you feel as if you were in a forest? 1. Very poor 2. Poor 3. Average 4. Good 5. Very good
B. Did you feel as if you were in the auditorium? 1. Very poor 2. Poor 3. Average 4. Good 5. Very good
C. Did you feel as if you were in front of the conductor? 1. Very poor 2. Poor 3. Average 4. Good 5. Very good
What was the most impressive sound? (description)

Table 3. Averaged scores of each age and total subject (1. very poor, 2. poor, 3. average, 4. good, and 5. very good)

	-9y/o	10's	20's	30's	40's	50y/o-	total
A. Forest sound	4.54	4.69	4.48	4.66	4.55	4.10	4.55
B. Auditorium	4.50	4.53	4.29	4.24	4.36	4.28	4.37
C. Stage	4.26	4.43	4.59	4.57	4.68	4.24	4.50

performance was being reproduced. Subjective assessment was conducted for the audience of the 3-D sound reproduction system. The questionnaire entries of the subjective assessment were listed in Table 2. For the orchestra performance (B and C), the subwoofer loudspeakers were employed to raise the lower frequency range. Each subwoofer loudspeaker was assigned to radiate the sound signals, which were measured by using the C_{80} microphone array with delay-and-sum.

4.3 Experimental Results and Discussions

In the demonstration, we obtained questionnaire responses from 390 subjects. The results of the subjective assessment for each age are shown in Fig. 5 and the averaged scores are shown in Table 3. Fig. 5 (a) shows that almost all subjects for all ages felt as if they were in the nature. This confirmed that the BoSC-based sound reproduction system reproduced a 3-D sound field and can create immersive environments.

Figs. 5(b) and (c) also show that many subjects rated the reproduced sound field as "good" or "very good." The system therefore can be said to yield a

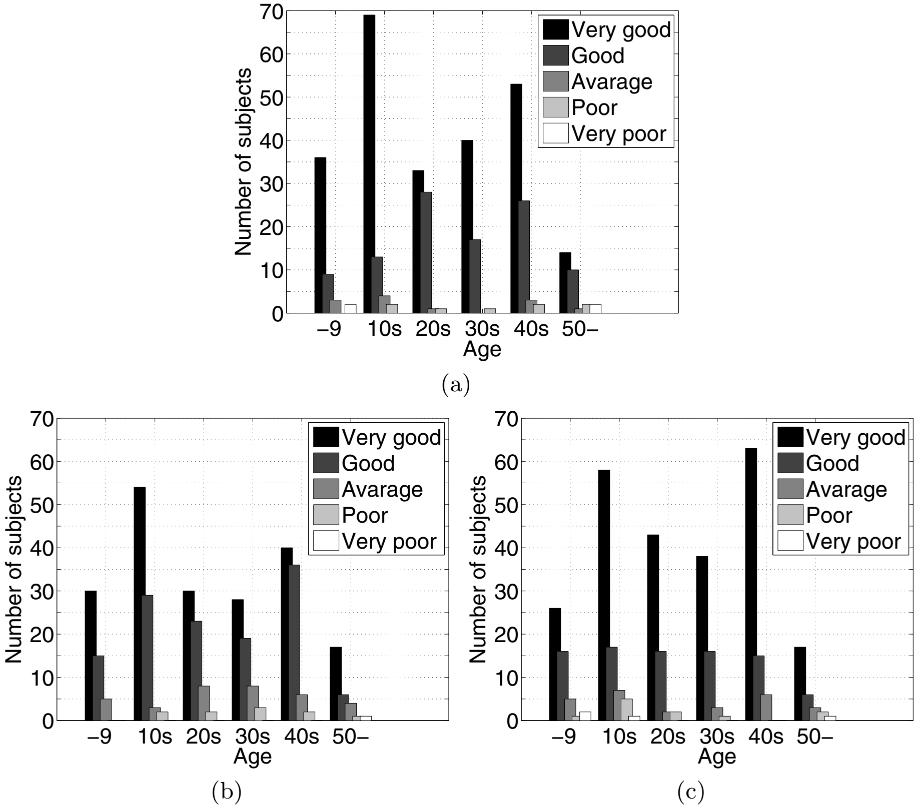


Fig. 5. Experimental results: Subjective assessments for (a) forest sounds, (b) auditorium, (c) stage.

presence similar to as if subjects were in the concert hall or on the stage. The scores for B, however, were smaller than for A or C, especially in those aged 20's to 40's since mismatches of reproduced sound fields and video information due to the monitor's size and position caused odd sensations. On the other hand, because only the conductor was shown on the monitor, the subjects did not have such sensations in C. In addition, subjects with experience playing instruments gave more "very good" scores in C compared to B.

5 Conclusions

We constructed a 3-D sound reproduction system based on the BoSC, consisting of a 70-channel microphone array designed from a C_{80} fullerene structure, and a loudspeaker array in which 62 full-range loudspeakers and eight subwoofer loudspeakers were installed.

To evaluate the performance of the BoSC-based system, we conducted subjective evaluation through the demonstration. The assessment results confirm that the system can provide immersive environments and the presence of other people or places. However, we did not limit the frequency range of the reproduced sound field, though theoretically it is limited up to 2 kHz. The accuracy of reproduced sound fields should be physically evaluated in the future.

Acknowledgments

Part of this study was supported by the Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Strategic Information and Communications R&D Promotion Program commissioned by the Ministry of Internal Affairs and Communications of Japan.

References

1. Berkhout, A.J., de Vries, D., Vogel, P.: Acoustic control by wave field synthesis. Journal of the Acoustical Society of America 93(5), 2764–2778 (1993)
2. Enomoto, S., Ise, S.: A proposal of the directional speaker system based on the boundary surface control principle. Electronics and Communications in Japan 88(2), 1–9 (2005)
3. Fletcher, H.: Auditory perspective - basic requirement. In: Symposium on Wire Transmission of Symphony Music and its Reproduction in Auditory Perspective, vol. 53, pp. 9–11 (1934)
4. Fletcher, H.: The stereophonic sound film system - general theory. The J. Acoust. Soc. Am. 13(2), 89–99 (1941)
5. Gauthier, P.A., Berry, A.: Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory. Journal of the Acoustical Society of America 119(5), 2721–2737 (2006)
6. Ise, S.: A principle of sound field control based on the kirchhoff-helmholtz integral equation and the inverse system theory. Journal of the Acoustical Society of Japan 53(9), 706–713 (1997) (in Japanese)
7. Ise, S.: A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems. Acustica 85, 78–87 (1999)
8. Lee, N.: Iosono. ACM Computers in Entertainment (CIE) 2(3), 3–3 (2004)
9. Lopez, J., Gonzalez, A., Fuster, L.: Room compensation in wave field synthesis by means of multichannel inversion. In: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2005)
10. Nakashima, T., Ise, S.: A theoretical study of the descretization of the boundary surface in the boundary surface control principle. Acoustical Science and Technology 27(4), 199–205 (2006)
11. Nelson, P.A., Elliot, S.J.: Active Control of Sound. Academic Press, San Diego (1992)

12. Spors, S., Buchner, H., Rabenstein, R., Herbordt, W.: Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering. Journal of the Acoustical Society of America 122(1), 354–369 (2007)
13. Steinberg, J., Snow, W.: Auditory perspective - physical factors. In: Symposium on Wire Transmission of Symphony Music and its Reproduction in Auditory Perspective, vol. 53, pp. 12–17 (1934)
14. Ward, D.B., Abhayapala, T.D.: Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Transactions on Speech and Audio Processing 9(6), 697–707 (2001)
15. Williams, E.G.: Fourier Acoustics:Sound Radiation and Nearfield Acoustical Holography. Academic Press, London (1999)