Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Optimization of loudspeaker and microphone configurations for sound reproduction system based on boundary surface control principle

– An optimizing approach using Gram-Schmidt orthogonalization and its evaluation –

Seigo Enomoto (1), Yusuke Ikeda (1), Shiro Ise (2) and Satoshi Nakamura (1)

(1) National Insitutute of Information and Communications Technology, Kyoto, Japan
(2) Graduate School of Engineering, Kyoto University, Kyoto, Japan

PACS: 43.38.Md, 43.38.Vk, 43.60.Tj

ABSTRACT

We propose a 3-D sound reproduction system based on the boundary surface control principle (BoSC system) and evaluate its performance via demonstration and exhibition. The BoSC reproduction system, dome-shaped and constructed of wood, consists of 62 full-range loudspeakers and eight subwoofer loudspeakers. The BoSC recording system is designed from C_{80} fullerene consisting of 70 microphones of a 46-cm diameter. In the listening room, 62 full-range loudspeakers assisted by the designed inverse filters reproduce sound fields identical to the primary sound fields by reproducing sound pressure on the 70 microphones which surround the listener's head. The BoSC system requires huge numerical calculation to reproduce authentic 3-D sound fields. Consequently, a pre-convolution calculation of the inverse filters is required to reproduce and transmit these fields. Therefore, to realize a real-time 3-D sound field reproduction system, we investigated optimization of the loudspeaker and microphone configuration using Gram-Schmidt orthogonalization. In the BoSC system, the inverse filters are determined by an inverse system of a transfer function matrix measured between each loudspeaker and microphone pair. Therefore, a transfer function matrix with a huge condition number degrades the accuracy of the reproduced sound fields. The selection of loudspeakers in the active control system that includes the BoSC system is equal to the selection of a vertical vector in the transfer function matrix. This means that for the reduction of the number of loudspeakers the vertical vector is selected up to the required numbers. By applying Gram-Schmidt orthogonalization to the selection of loudspeakers, the loudspeaker is selected in the order of linear independence from highest to lowest. In this paper, the effect of the reduction of loudspeakers and microphones is evaluated by the subjective assessment of a sound image localization test.

INTRODUCTION

Stereophony is an important technology for the sensation of realistic sound. In recent years, the development of sound reproduction systems based on the Kirchhoff-Helmholtz integral equation (KHIE) is an attractive research field due to the development of digital signal processing and hardware performance. A 3-D sound field can be physically reproduced using such reproduction systems. Therefore, it is expected that these systems will be able to provide the listeners with greater sensation than the stereo or surround sound systems. For this purpose, the Boundary Surface Control principle (BoSC) based on the KHIE and the MIMO (Multiple-Input Multiple-Output) inverse system is proposed (Ise 1993; Ise 1999). We consider the BoSC to be an appropriate theory for the construction of a physics-based sound reproduction system. Consequently, we have developed the sound reproduction system based on the BoSC system and demonstrated its performance. Over 800 subjects participated in the demonstration and gave good ratings to the realistic sensation that was produced. We have also developed a "*sound field sharing*" (SFS) system, which comprises two remotely placed BoSC systems. Using the system, users receive the reproduced 3-D sound and converse with each other as if the conversation partner were in the same acoustic space. However, the system has several restrictions:

1. The 3-D sound signal should be convoluted with the inverse filters in advance.
2. Up to two users can participate in the system.

The lack of computer power causes these two restrictions. Therefore, to realize a real-time 3-D sound reproduction system and the three-party conversation system, we intend to reduce the number of channels in the microphone and loudspeaker arrays using Gram-Schmidt orthogonalization. An optimizing approach using this orthogonalization is proposed by Asano (Asano, Suzuki, and Swanson 1999). By using Gram-Schmidt orthogonalization, the configuration and number of loudspeakers can be optimized for an active control system. The candidates for a loudspeaker location with high independence in the transfer function matrix are chosen preferentially so that the system can prevent degradation in terms of sound reproduction accuracy.

In this paper, first the BoSC sound reproduction system is introduced. Next, the application of Gram-Schmidt orthogonalization for the BoSC system is described. By using Gram-Schmidt orthogonalization, the number of loudspeakers and microphones is reduced in the BoSC system. Finally, the influence of the reduction is confirmed by subjective assessment.

BOUNDARY SURFACE CONTROL PRINCIPLE (BOSC)

The region definition of the KHIE is illustrated in Fig. 1, where the interior volume V is completely enclosed by the boundary

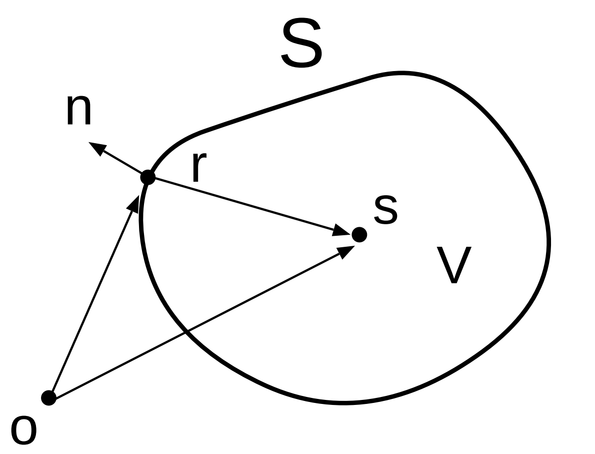

Figure 1: Region definition for the Kirchhoff-Helmholtz integral equation. The BoSC sound reproduction system is based on the KHIE. Volume V is completely bounded by surface S. $\mathbf{n}$ is the outward normal vector. $\mathbf{s}$ and $\mathbf{r}$ are the evaluation point and observation point, respectively. The interior sound pressure $p(\mathbf{s},\omega)$ is fully determined by the sound pressure $p(\mathbf{r},\omega)$ and particle velocity $v_n(\mathbf{r},\omega)$.

S. If there is no singular point in a given volume V and on a boundary S, the sound pressure at the evaluation point $\mathbf{s}$ within V can be written as

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

where ρ_0 is the medium density, ω is the angular frequency, $p(\mathbf{r},\omega)$ and $v_n(\mathbf{r},\omega)$ are the sound pressure and outward normal particle velocity at $\mathbf{r}$, respectively, and $G(\mathbf{r}|\mathbf{s},\omega)$ is the Green's function, where it is assumed that the suitable $G(\mathbf{r}|\mathbf{s},\omega)$ can be chosen for any impedance boundary (Williams 1999; Nelson and Elliot 1992). Similarly, the sound pressure $p(\mathbf{s}',\omega)$ at $\mathbf{s}'$ within V' enclosed by S' can be written as

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

where $\mathbf{r}'$ represents the observation point on the boundary S'. If it is supposed that $G(\mathbf{r}|\mathbf{s},\omega)$ is a *free-space* Green's function $g(\mathbf{r}|\mathbf{s},\omega)$ and the V' and S' are congruous to V and S, respectively, then Eq. (3) is satisfied.

$$\begin{aligned} g(\mathbf{r}|\mathbf{s},\omega) &= g(\mathbf{r}'|\mathbf{s}',\omega), \\ \frac{\partial g(\mathbf{r}|\mathbf{s},\omega)}{\partial n} &= \frac{\partial g(\mathbf{r}'|\mathbf{s}',\omega)}{\partial n'}. \end{aligned} \quad (3)$$

By substituting Eq. (3) in Eqs. (1) and (2), we get

$$\left\{ \begin{array}{c} p(\mathbf{r},\omega) = p(\mathbf{r}',\omega) \\ v_n(\mathbf{r},\omega) = v_n(\mathbf{r}',\omega) \end{array} \right\} \Rightarrow p(\mathbf{s},\omega) = p(\mathbf{s}',\omega). \quad (4)$$

Eq. (4) specifies the sound reproduction system when V and V' are considered the region of primary and reproduced sound fields. Therefore, if the sound pressure and the outward normal particle velocity on S and S' satisfy $p(\mathbf{r},\omega) = p(\mathbf{r}',\omega)$ and $v_n(\mathbf{r},\omega) = v_n(\mathbf{r}',\omega)$, then the sound pressure enclosed in the boundary satisfies $p(\mathbf{s},\omega) = p(\mathbf{s}',\omega)$. In this case, $p(\mathbf{r},\omega)$ and $v_n(\mathbf{r},\omega)$ are the signals recorded in the primary field, and $p(\mathbf{r}',\omega)$ and $v_n(\mathbf{r}',\omega)$ are the signals reproduced in the listening area using secondary loudspeakers.

Conventional physics-based sound reproduction systems imply that $g(\mathbf{r}|\mathbf{s},\omega)$ and $\partial g(\mathbf{r}|\mathbf{s},\omega)/\partial n$ are monopole and dipole

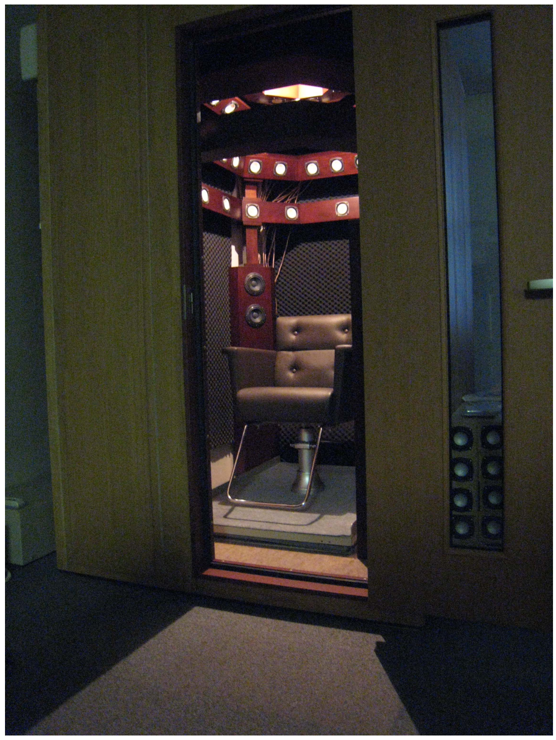

Figure 2: BoSC 3-D sound reproduction system. Loudspeaker array is installed in a soundproofed room with D-30 sound insulation level.

Figure 3: Microphone array for BoSC 3-D sound reproduction system. Microphone array is designed from the structure of C80 fullerene. Seventy omnidirectional microphones are put on each node.

sound sources located on the boundary. In addition, $p(\mathbf{r},\omega)$ and $v_n(\mathbf{r},\omega)$ are driving functions for both sound sources. In contrast, the BoSC sound reproduction system states that $g(\mathbf{r}|\mathbf{s},\omega)$ and $\partial g(\mathbf{r}|\mathbf{s},\omega)/\partial n$ are the function defined by the boundary-shape. $p(\mathbf{r},\omega)$ and $v_n(\mathbf{r},\omega)$ are the variables to be measured

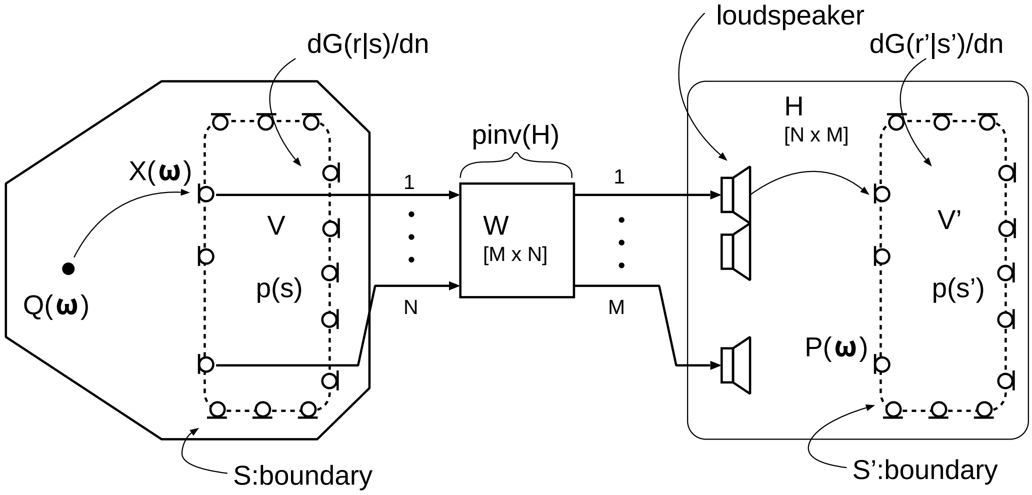

Figure 4: Inverse filter design for sound reproduction based on BoSC. The rooms depicted on the left and right side represent the primary sound field and reproduction room (listening room). The configuration of microphones on S' is congruous to it on S. As a consequence, V' and its bounding surface S' are congruous to V and its bounding surface S.

and reproduced on the boundary.

BoSC 3-D sound reproduction system

The loudspeaker and microphone arrays for the BoSC sound reproduction system are depicted in Figs. 2 and 3. The loudspeaker array is dome-shaped and constructed of wood. The dome structure consists of four layers; with 6, 16, 24, and 16 channel full-range loudspeakers respectively, hence the array comprises 62 full-range loudspeakers. In addition, the dome parts are supported by four pillars where two subwoofer loudspeakers are installed. The loudspeaker array is built in the simplified assembly sound-proofed room (sound insulation level is D-30) to reduce environmental noise. For regularly configuring microphones the microphone array is designed from C_{80} fullerene (Fowler and Manolopoulos 2007) that consisting of 70 microphones of a 46-cm diameter. Omnidirectional microphones are positioned on each node of the fullerene. The intervals of each microphone are around 8 cm between the shortest pair, and 16cm between the longest. The reproduction area in the BoSC system is the inner-region enclosed by the microphone array. Therefore, the system is supposed to reproduce the 3-D sound field surrounding the listener's head.

Using an omnidirectional microphone, only sound pressure is obtained on the surface bounding the reproduction area. In this case, applying the Dirichlet Green's function, $G_D(\mathbf{r}|\mathbf{s},\omega)$ (Williams 1999), Eq. (1) can be written as

$$p(\mathbf{s},\omega) \quad = \quad \int_S p(\mathbf{r},\omega)\frac{\partial G_D(\mathbf{r}|\mathbf{s},\omega)}{\partial n}dS. \tag{5}$$

Since Eq. (5) a the Dirichlet boundary value problem, a unique solution exists, with the exception of eigenfrequencies (Schenck 1967). Thus, only the sound pressure needs to be specified on the boundary in order to determine the interior sound pressure. Therefore, the BoSC system can reproduce the interior sound pressure by reproducing the sound pressure measured on the boundary. The value of $G_D(\mathbf{r}|\mathbf{s},\omega)$ is also not required in the BoSC system because the Green's function is considered to be invariant.

INVERSE FILTER DESIGN IN THE BOSC SYSTEM

The block diagram of the sound reproduction system based on BoSC is illustrated in Fig. 4. The volume V and its bounding surface S represent the region in the primary field and V' and S' represent the region in the reproduced field. Whereas the shape of S' is congruous to S, the reproduction environment is allowed to be different from the primary (recording) environment. Moreover, the BoSC system can be applied to the reverberant environment because the inverse filter can correctly compensate it.

The reproduced sound pressure in vector form $\mathbf{P}(\omega)$ is defined as

$$\mathbf{P}(\omega) \quad = \quad [\mathbf{H}(\omega)][\mathbf{W}(\omega)]\mathbf{X}(\omega), \tag{6}$$

where

$$\begin{aligned}
\mathbf{X}(\omega) &= [X_1(\omega),\cdots,X_N(\omega)]^T, \\
\mathbf{P}(\omega) &= [P_1(\omega),\cdots,P_N(\omega)]^T, \\
[\mathbf{H}(\omega)] &= \begin{pmatrix} H_{11}(\omega) & \cdots & H_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ H_{N1}(\omega) & \cdots & H_{NM}(\omega) \end{pmatrix}, \text{and} \\
[\mathbf{W}(\omega)] &= \begin{pmatrix} W_{11}(\omega) & \cdots & W_{1N}(\omega) \\ \vdots & \ddots & \vdots \\ W_{M1}(\omega) & \cdots & W_{MN}(\omega) \end{pmatrix}.
\end{aligned} \tag{7}$$

$[\mathbf{H}(\omega)]$ represents the matrix form of the transfer function between mth loudspeaker and nth microphone. All components in Eqs.(6) and (7) are complex. Any signal $\mathbf{X}(\omega)$, $\mathbf{P}(\omega)=\mathbf{X}(\omega)$ is satisfied by

$$[\mathbf{W}(\omega)] \quad = \quad [\mathbf{H}(\omega)]^+, \tag{8}$$

where $^+$ represents the pseudo-inverse matrix (Golub and Loan 1996). Thus, $[\mathbf{W}(\omega)]$ is defined as the inverse system of $[\mathbf{H}(\omega)]$. It is well known that regularization is a reasonable approach for solving the inverse problem. It has also already been applied to a sound reproduction system (Tokuno et al. 1997; Corteel 2007). By using regularization, the calculated inverse matrix $[\widehat{\mathbf{W}}(\omega)]$ for Rank$([\mathbf{H}(\omega)])=N$ can be given as

$$[\widehat{\mathbf{W}}(\omega)] = ([\mathbf{H}(\omega)]^\dagger[\mathbf{H}(\omega)]+\beta(\omega)\mathbf{I}_M)^{-1}[\mathbf{H}(\omega)]^\dagger, \tag{9}$$

where $\dagger$ is a conjugate transpose, $\beta(\omega)$ is the regularization parameter, and $\mathbf{I}_M$ is an $M\times M$ unit matrix. On the other hand, the right inverse matrix (Golub and Loan 1996; Bauck and Cooper 1996) for Rank$([\mathbf{H}(\omega)])=M$ is derived as

$$[\widehat{\mathbf{W}}(\omega)] = [\mathbf{H}(\omega)]^\dagger([\mathbf{H}(\omega)][\mathbf{H}(\omega)]^\dagger+\beta(\omega)\mathbf{I}_N)^{-1}. \tag{10}$$

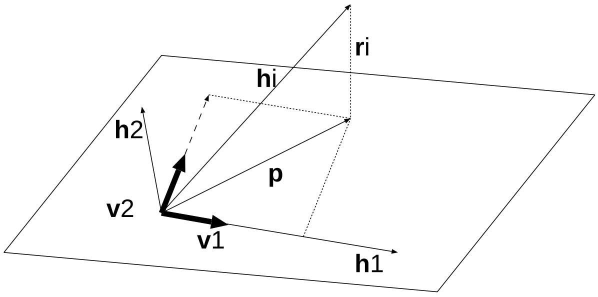

Figure 5: Column subspace of $[\mathbf{H}(\omega)]$. The subspace is spanned by $\mathbf{h}_1$ and $\mathbf{h}_2$. $\mathbf{p}$ denotes the projection of $\mathbf{h}_i$ onto the subspace. For $i = 3$, $\mathbf{h}_i$ (third source) is determined so that the 2-norm of $\mathbf{r}_i$ is maximized.

Eq. (9) and Eq. (10) are interpreted as the least square solution and minimum norm solution, respectively. For $\mathrm{Rank}([\mathbf{H}](\omega)) = N = M$, $[\mathbf{H}(\omega)]$ is also not singular and $[\mathbf{W}(\omega)]$ is given as $[\mathbf{H}(\omega)]^{-1}$. Finally, the time-domain inverse filter coefficients are derived from the inverse DFT of $[\widehat{\mathbf{W}}(\omega)]$. Note that both the loudspeaker and microphone array configurations affect the spatial sampling in the BoSC system (Enomoto and Ise 2005).

In Eqs. (9) and (10), a suitably chosen $\beta(\omega)$ reduces the instability of the inverse system. In this research, $\beta(\omega)$ was defined in each octave frequency band, heuristically. Additionally, the inverse filters were calculated by using the impulse responses measured between each loudspeaker and microphone pair in the soundproofed room beforehand. As a consequence, the fluctuations caused by environmental changes were not followed in later experiments. For a fluctuating real environment, MIMO adaptive inverse filtering can be applied to the BoSC system in which case the sensor microphones are placed on the boundary surface surrounding the reproduction area.

GRAM-SCHMIDT ORTHOGONARIZATION

This section describes the basic algorithm of loudspeaker selection using Gram-Schmidt orthogonalization for single frequency. If the linear independence of the N dimensional vertical vector which comprises the $N \times M$ matrix is low, the matrix is said to be *ill-conditioned*. The degradation of linear independence in $[\mathbf{H}(\omega)]$ causes instability of the BoSC system. $[\mathbf{H}(\omega)]$ in Eq. (7) can be written as Eq. (11).

$$\begin{aligned} \mathbf{P}(\omega) &= [\mathbf{H}(\omega)]\mathbf{Y}(\omega) \\ &= \{\mathbf{h}_1(\omega), \cdots, \mathbf{h}_M(\omega)\}\mathbf{Y}(\omega), \end{aligned} \tag{11}$$

where $\mathbf{Y}(\omega) = [\mathbf{W}(\omega)]\mathbf{X}(\omega)$ and $\mathbf{h}_{m,m=1,\cdots,M}(\omega)$ is the N dimensional vertical vector of $[\mathbf{H}(\omega)]$. The vertical vector $\mathbf{h}(\omega)$ represents the transfer function between a certain loudspeaker and each microphone at frequency ω. Therefore, selection of loudspeaker locations using Gram-Schmidt orthogonalization means the selection of the vertical vector $\mathbf{h}(\omega)$ pair with high linear independence from $[\mathbf{H}(\omega)]$.

Basic algorithm

In the lth step, $(l-1)$ loudspeakers already chosen. $\tau = \{\mathbf{h}_1, \cdots, \mathbf{h}_M\}$ denotes the set of the vertical vector including $[\mathbf{H}]$, S_{l-1} denotes the subset of vector selected at $(l-1)$ steps, and τ_{l-1} denotes the subset of the unused vector. $\nu_{l-1} = \{\nu_1, \cdots, \nu_{l-1}\}$ denotes the orthonormal basis of the subspace spanned by S_{l-1}. Fig. 5 is an example of the subspace spanned by S_{l-1}. In the lth step, $\hat{\mathbf{h}}_l$ is selected so that the perpendicular component of $\hat{\mathbf{h}}_l$ to the subspace spanned by S_{l-1} is the largest. The perpendicular component $\mathbf{r}_i$ of an arbitrary vector $\mathbf{h}_i \in \tau_{l-1}$ is expressed as

$$\mathbf{r}_i = \mathbf{z}_i - \mathbf{p} \tag{12}$$

where $\mathbf{p}$ denotes the projection of $\mathbf{h}_i$ onto the subspace spanned by S_{n-1}. The nth loudspeaker is determined so that 2-norm of $\mathbf{r}_i$ is maximized, i.e.,

$$\hat{\mathbf{h}}_l = \arg\max_{\mathbf{h}_i \in \tau_{l-1}} J(\mathbf{h}_i) \tag{13}$$

where $J(\mathbf{h}_i)$ is the performance index defined as

$$J(\mathbf{h}_i) = ||\mathbf{r}_i||. \tag{14}$$

If the perpendicular component of $\hat{\mathbf{h}}_l$ is denoted as $\hat{\mathbf{r}}_l$, then nth orthonormal basis ν_l is determined as

$$\nu_l = \frac{\hat{\mathbf{r}}_l}{||\hat{\mathbf{r}}_l||}. \tag{15}$$

The maximized performance index at the lth step is denoted as

$$\hat{J}_l = J(\hat{\mathbf{h}}_l) \tag{16}$$

This process is iterated until $\hat{J}_l$ becomes smaller than J_{thr}. For the frequency range of interest $[\omega_h, \omega_j]$, the following two performance indexes are evaluated:

$$\begin{aligned} J_{avg}(\bar{\mathbf{h}}_i) &= \frac{1}{K}\left(a_h||\mathbf{r}_i(\omega_h)||+, \cdots, +a_j||\mathbf{r}_i(\omega_j)\right) \\ J_{min}(\bar{\mathbf{h}}_i) &= \min\left(a_h||\mathbf{r}_i(\omega_h)||, \cdots, a_j||\mathbf{r}_i(\omega_j)\right) \end{aligned} \tag{17}$$

where $\bar{\mathbf{h}}_i = \{\mathbf{h}_i(\omega_l), \cdots, \mathbf{h}_i(\omega_h)\}$, K denotes the number of discrete frequencies, and a_k denotes an arbitrary weight at the discrete frequency ω_k. The perpendicular component $\mathbf{r}_i(\omega_k)$ and orthonormal basis $\mathbf{v}_i(\omega_k)$ are calculated for every discrete frequency separately in the same manner as a case of single frequency. During the optimization process (Asano, Suzuki, and Swanson 1999), J_{avg} is maximized. On the other hand, J_{min} is judged as terminating the optimization process.

Selection of loudspeaker configuration

In (Asano, Suzuki, and Swanson 1999), selection was continued if the performance index was larger than the threshold. However, an approach to determine the appropriate threshold was not confirmed. Thus, this research investigated the greatest number of loudspeakers and microphones we can use in the real-time 3-D sound transmission system and three-party conversation system. Then, we determined the loudspeaker configuration up to those numbers using Gram-Schmidt orthogonalization. The results are strongly influenced by the first chosen loudspeaker location according to Gram-Schmidt orthogonalization, because the loudspeaker location is determined by the previously selected configuration of loudspeakers. The performance indexes are represented in Fig. 6. In this figure 24 loudspeakers were selected. The free-space Green's function was used to obtain the transfer function between each loudspeaker and microphone. Although upper limit frequency for stimulus described in the next section was not restricted, the loudspeaker configuration was determined in frequencies ranging from 20 Hz to 1 kHz every 20 Hz. Many loudspeakers located on the upper layer were chosen if the upper limit frequency was not restricted. Synthesizing a wave front that comes from a direction where there is no loudspeaker is difficult in 3-D sound reproduction systems. Therefore the loudspeakers should be located in every possible directions around the microphone array. In Fig. 6, the blue and red line show the performance index J_{avg} and J_{min} at the 24th steps. As a consequence, we selected the #60 loudspeaker as the initial location so that the largest performance index was obtained at the 24th loudspeaker selection. Fig. 7 shows the performance index in each step when the #60 loudspeaker was selected as the initial location. As can be seen from the figure, J_{avg} decreases for each iteration. The configuration of selected loudspeakers is shown in Fig. 8 where 24 loudspeakers were selected and the loudspeaker #60 was selected as the initial location. The green point shows the #60 loudspeaker and blue points

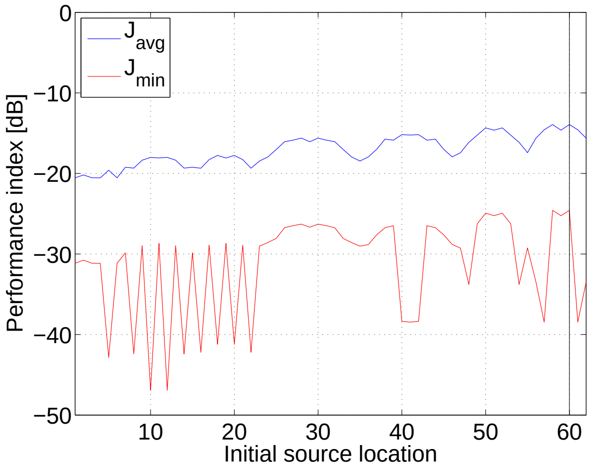

Figure 6: Performance indexes when 24 loudspeakers are selected. Blue line shows the J_{avg}, and red line shows the J_{min} for each initial selected loudspeaker location.

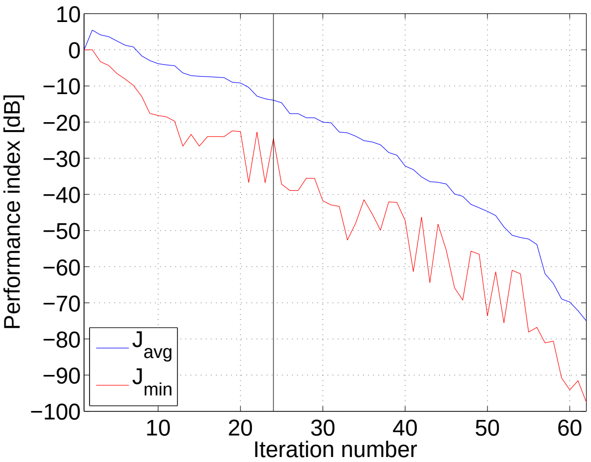

Figure 7: Performance indexes. #60 loudspeaker is chosen for the initial location. Blue line shows the J_{avg}, and red line shows the J_{min}.

show the chosen loudspeaker location as a result of the iteration based on Gram-Schmidt orthogonalization. White points show the configuration of entire loudspeaker array. From the figure, the loudspeakers distributed in each direction and height are regularly observed.

Selection of microphone configuration

To realize real-time 3-D sound transmission system, the number of microphones should be reduced. By replacing the configuration of the loudspeakers and microphones, Gram-Schmidt orthogonalization can be applied to the reduction of the microphone array. The results of the preliminary experiment show that the number of loudspeakers (M) and microphones (N) should be decided so that the number of components in $[\mathbf{W}(\omega)]$ is less than $M \times N = 192$. Hence, the number of microphones is eight for 24 loudspeakers. The eight microphones configuration was determined in the same manner as the case of the loudspeaker selection. Fig. 9 shows the configuration of eight microphones for the configuration of 24 loudspeakers depicted in Fig. 8. Fig. 9 shows the constructed microphone array frame and the green point represents the location selected initially. The blue points represent the chosen microphone locations. The selected microphones were assembled in the upper half of the microphone array. However, every direction in the horizontal plane

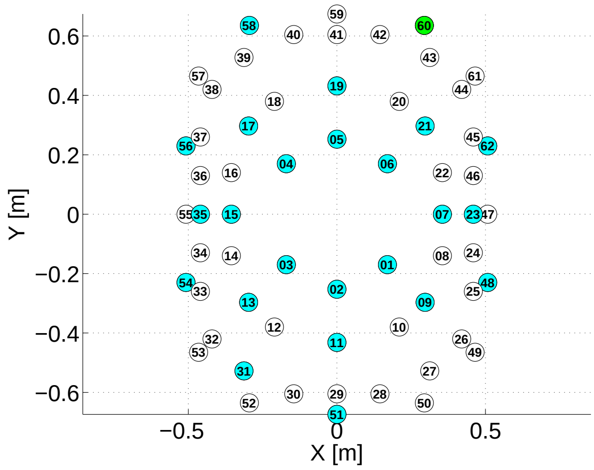

Figure 8: Selected loudspeaker location based on Gram-Schmidt orthogonalization. Loudspeaker #1 to #6 composes the first layer of the BoSC loudspeaker, #7 to #22 is the second, #23 to #46 is the third, and #47 to #62 is the fourth. Green point shows initially selected loudspeaker. Blue points show the selected loudspeakers as a result of the iteration based on Gram-Schmidt orthogonalization.

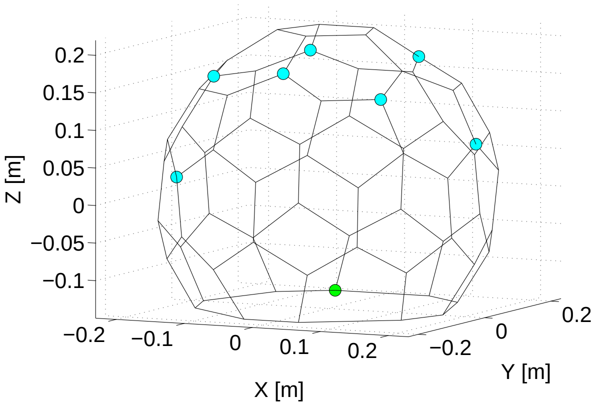

Figure 9: Selected microphone location based on Gram-Schmidt orthogonalization. Green point shows initially selected microphone. Blue points show the selected microphones.

Table 1: Experimental conditions.

Cond.	Num. of Sp.	Num. of Mic.
1	24	8
2	32	6
3	16	12
4	24	70
5	62	70

are covered.

SOUND LOCALIZATION TEST

Experimental conditions

In order to evaluate the influence of reduction using Gram-Schmidt orthogonalization, a sound localization test in a horizontal plane was conducted. Stimuli reproduced using the BoSC system were generated from the convolution of pink noise and the impulse response. The impulse responses were simulated from the free-space Green's function. The source location was assumed as one meter away from the center of the microphone array in the simulation. The inverse filters for 3-D sound reproduction in the BoSC system were calculated by using the

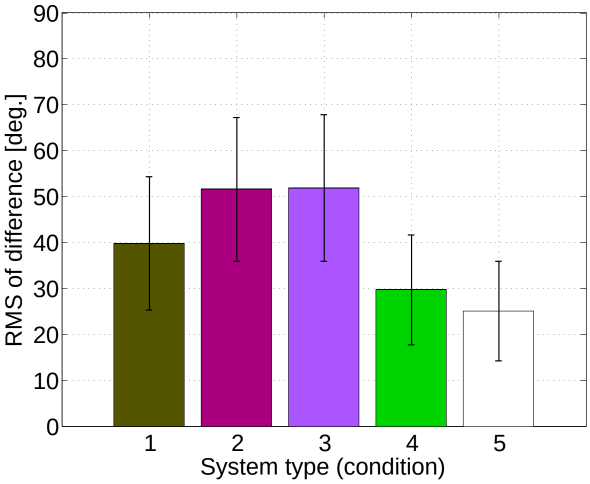

Figure 10: Averaged RMS values of errors between the subjects' perceived angle. Error-bar represents 95% CI.

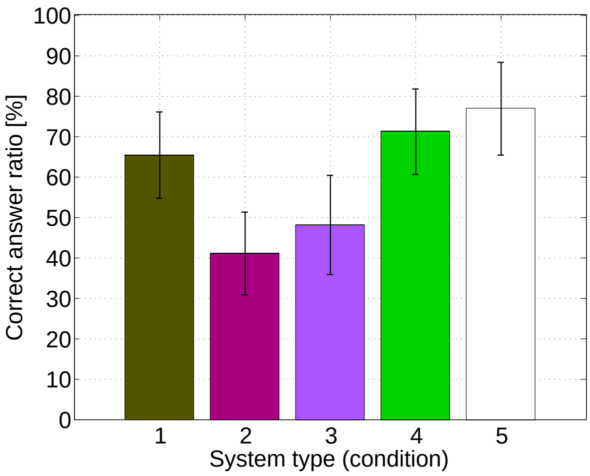

Figure 11: Correct answer ratio of subject's perceived angle for the presented sound localization. Error-bar represents 95% CI.

impulse responses measured in advance with 48 kHz sampling frequency, and the inverse filter length was 4,096 points. The sound pressure level of stimulus was adjusted to $L_{A,Fmax} = 55$ dB at the center of the microphone array to eliminate the level difference between each condition and direction. The experimental conditions with respect to the number of loudspeakers and microphones are summarized in Table 1. Whole loudspeakers and microphones were used in condition 5. In condition 4, the number of loudspeakers was reduced to 24. Conditions 1, 2, and 3 were consistent with the reduction for the number of the elements in the inverse matrix, $[\mathbf{W}(\omega)]$. Thirteen subjects aged 20s to 50s (five males and eight females) indicated their perceived angles after listening to the reproduced stimuli. Stimuli were presented from 0 - 330 degrees at every 30 degrees. They were continued with two seconds and repeated two times at each angle. The presentation order was randomized using the Latin square method. Subjects were permitted to move their head and body during the listening to stimulus.

Experimental results

The results of the sound localization test are represented in Figs. 10 and 11. Fig. 10 shows the RMS values of the difference between the reproduced sound angle and the perceived one. In the figure, the error-bar represents a 95% confidence interval (CI). Condition 5 obtained the lowest RMS value, and as a consequence the highest accuracy of sound localization was achieved. A total of 24 loudspeakers and 70 microphones were used in condition 4, and the RMS value was around 5 degrees more than condition 5. The RMS value of condition 1 was around 15 degrees more than condition 5. On the other hand, the RMS value of conditions 2 and 3 are almost the same, but their values were around 25 degrees more than condition 5.

Table 2: The results of Tukey's multiple comparison test for the RMS value of error.

	cond. 2	cond. 3	cond. 4	cond. 5
cond. 1	2.04	2.04	1.55	2.28
cond. 2		0.00	3.59	4.32*
cond. 3			3.59	4.32*
cond. 4				0.73

Table 3: The results of Tukey's multiple comparison test for the correct answer ratio.

	cond. 2	cond. 3	cond. 4	cond. 5
cond. 1	4.79*	3.40	1.13	2.27
cond. 2		1.39	3.59	7.06**
cond. 3			4.54*	5.67**
cond. 4				1.13

Fig. 11 shows the correct answer ratio and the error-bar represents 95% CI. Condition 5 again obtained the highest ratio and the ratio was around 5% more than condition 4, and around 10% more than condition 1. Since a Hartley's test ensured that all conditions have a similar variance, Tukey's multiple comparison test was applied to find which system means were significantly different from the others. The results of statistical test were summarized in Table 2 and 3. The marker "$*$" and "$**$" in the tables represent a respective 1% and 5% significant difference between each condition. A Tukey's multiple comparison test ensured that there were no significant differences between conditions 5 and 4, and conditions 5 and 1. It is therefore confirmed that the number of loudspeakers and microphones composing the BoSC system can be reduced using Gram-Schmidt orthogonalization. In contrast, a significant difference between conditions 5 and 2, and conditions 5 and 3 was observed. Therefore, using these conditions, the realistic sensation reproduced by the BoSC system was impaired.

CONCLUSION

This paper described a reduction method of the number of loudspeakers and microphones using Gram-Schmidt orthogonalization. To realize the real-time 3-D sound transmission system and three-party conversation system, the number of loudspeakers and microphones should be reduced. We consider these reductions a requirement for practical application of BoSC. Using Gram-Schmidt orthogonalization, the vertical vector group with high independence was selected from the matrix comprising the transfer functions between each loudspeaker and microphone. The selected vectors correspond to the loudspeaker and microphone configurations in the BoSC system. Therefore, such a system can overcome acoustical environmental changes compared to the systems composed by the configurations selected using other criteria. In the selection procedure, a restriction of the frequency range of interest from 20 Hz to 1 kHz satisfied the regularly distributed configuration of loudspeakers in the simulation. Gram-Schmidt orthogonalization was also applied to the reduction for the number of microphones. As a result of the selection, we obtained a reduced number of microphones distributed regularly in all horizontal directions. Although microphones were selected in the same manner as in the loudspeaker selection, the reduction procedures for the configuration of loudspeaker were already carried out by using Gram-Schmidt orthogonalization. Since the stability of the BoSC system relies on the configuration of both loudspeakers and microphones, further investigation to determine both configurations should be conducted in future studies.

To evaluate the degradation caused by the reduction of loudspeakers and microphones, a sound localization test in a horizontal plane was conducted. The results of subjective assessments showed there is no statistically significant difference between the BoSC system consisting of 62 loudspeakers and 24 loudspeakers. Moreover, with respect to the number of microphones, there is no statistically significant difference between 70 and eight microphones for 24 loudspeakers. To develop a three-party conversation system, we apply the configuration of 24 loudspeakers (Ikeda et al. 2010). Again, we apply the configurations of 24 loudspeakers and eight microphones to realize a real-time 3-D sound transmission system. The intervals of each loudspeaker and microphone are related to the frequency range of interests in the BoSC system. In future studies, the investigation for the median and frontal planes should be conducted since the configuration in these planes is sparse compared with the horizontal plane.

REFERENCES

Asano, F., Y. Suzuki, and D.C. Swanson (1999). "Optimization of control source configuration in active control systems using Gram-Schmidt orthogonalization". *Speech and Audio Processing, IEEE Transactions on* 7.2, pp. 213–220. ISSN: 1063-6676. DOI: 10.1109/89.748126.

Bauck, Jerry and Duane H Cooper (1996). "Generalized Transaural Stereo and Applications". *Journal of the Audio Engineering Society* 44.9, pp. 683–705.

Corteel, E. (2007). "Synthesis of Directional Sources Using Wave Field Synthesis, Possibilities, and Limitations". *EURASIP Journal on Advances in Signal Processing* 2007, Article ID 90509.

Enomoto, Seigo and Shiro Ise (2005). "A proposal of the directional speaker system based on the boundary surface control principle". *John Wiley & Sons, Electronics and Communications in Japan* 88.2, pp. 1–9.

Fowler, P. W. and D. E. Manolopoulos (2007). *An Atlas of FULLERENES*. Dover Publications.

Golub, Gene H. and Charles F. Van Loan (1996). *Matrix computations*. Baltimore, Maryland: The Johns Hopkins University Press.

Ikeda, Yusuke et al. (2010). "Three-party sound field sharing system based on the boundary surface control principle: Subjective assessment of voice reproduction with speaker's facing angle". *Proc. 2010 20th International Congress on Acoustics, ICA2010*. Sydney, Australia.

Ise, Shiro (Oct. 1993). "A study on the sound field reproduction in a wide area (1) -Based on Kirchhoff-Helmholtz integral equation-". *Proc. of Autumun meeting of the Acoustical Society of Japan*. (in Japanese), pp. 479–480.

— (1999). "A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems". *Acustica* 85, pp. 78–87.

Nelson, P. A. and S. J. Elliot (1992). *Active Control of Sound*. San Diego, CA: ACADEMIC PRESS.

Schenck, Harry A. (1967). "Improved Integral Formulation for Acoustic Radiation Problems". *Journal of the Acoustical Society of America* 44.1, pp. 41–58.

Tokuno, Hironori et al. (1997). "Inverse Filter of Sound Reproduction Systems Using Regularization". *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer* E80-A.5, pp. 809–820.

Williams, E. G. (1999). *Fourier Acoustics:Sound Radiation and Nearfield Acoustical Holography*. London: Academic Press.