Acoust. Sci. & Tech. **31**, 4 (2010)

PAPER

Robust transaural sound reproduction system based on feedback control

Toshiya Samejima[1], Yo Sasaki[2], Izumi Taniguchi[3] and Hiroyuki Kitajima[4]

[1]*Department of Acoustic Design, Faculty of Design, Kyushu University, 4–9–1, Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan*
[2]*Japan Broadcasting Corporation, 2–2–1, Jinnan, Shibuya-ku, Tokyo, 150–8001 Japan*
[3]*Cybozu Corporation, BizPort 1F, 2821–4, Minamiyoshida-machi, Matsuyama, 791–8042 Japan*
[4]*Nittobo Acoustic Engineering Co., Ltd., 1–21–10, Midori, Sumida-ku, Tokyo, 130–0021 Japan*

(*Received 20 April 2009, Accepted for publication 6 January 2010*)

Abstract: We present a new method for implementing transaural sound reproduction systems by using feedback control theory. The H_∞ control theory is employed to synthesize the feedback controller. The structure of the sound reproduction system is formulated such that the H_∞ norm of the system transfer function, which is to be minimized with feedback control, expresses the difference between the desired signals and the acoustic signals reproduced at the ears of a listener via loudspeakers. Modeling errors and plant perturbations resulting from the movement of the listener's head are also taken into account to ensure robust stability. Computer simulations indicate that the equalization and cross-talk cancellation by the proposed method are better preserved for deviations in position than are those by the conventional inverse filtering method.

Keywords: Transaural system, Feedback control, Robust control, H_∞ control theory

PACS number: 43.38.Md, 43.60.Pt [doi:10.1250/ast.31.251]

1. INTRODUCTION

Sound field reproduction is one of the most important purposes in the active control of sound fields. Sound field reproduction aims at reproducing the sound field of a space (primary sound field) in a different moment and place (secondary sound field), and much work has been reported on this technique up to now.

Sound field reproduction techniques that can reproduce sound fields precisely are classified into two approaches. One approach is based on Huygens' principle, i.e., the wave theory. Several methods of this approach have been proposed [1,2]; these methods use monopole and dipole sources installed in the boundary surface of a control area. The strength of those monopole and dipole sources is adjusted in proportion to the acoustic pressure and particle velocity on the boundary surface of a primary sound field, because the sound field in the control area is governed by the Kirchhoff-Huygens formula. Although it is possible for a primary sound field to be reproduced in a wide area by this technique, ideal monopole and dipole sources are necessary as control sources. It has been pointed out that, under a certain condition, a requested sound field can be generated only by adjusting the acoustic pressure on the boundary surface of a control area with monopole sources [3,4]. However, the wider a control area, the larger the number of control sources necessary to achieve a satisfactory control effect; this issue has been considered a common problem of this type of sound field reproduction technique.

Other approaches are a binaural system and a transaural system. Both systems aim at two-channel stereophonic sound reproduction, whereby acoustic signals are recorded at two points in a primary sound field, for example, at the ears of a dummy head. A binaural system uses a headphone to replay the recorded signals, and technically it is easy to realize. A transaural system is a technique for reproducing the recorded signals precisely at the ears of a listener via two loudspeakers with the help of inverse filtering of transfer functions from the loudspeakers to the ears. The basic arrangement of a transaural system is shown in Fig. 1. In order to make the signals $\{y_{p1}, y_{p2}\}^T$ produced at the ears of the listener equal to the recorded signals $\{r_1, r_2\}^T$, the controller $\boldsymbol{C}$ is designed such that it contains

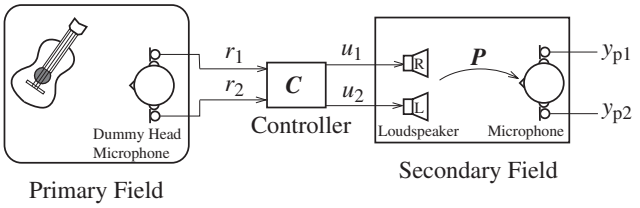

Fig. 1 Arrangement of transaural system.

the pseudo-inverse of the transfer function $\boldsymbol{P}$ from the loudspeakers to the ears. This type of system realized by analog technology has been reported from the 1960s to the early 1970s [5,6]. Since then, with the development of electronic devices for the digital processing of acoustic signals, several systems that make use of the advantage of digital signal processing have been developed [7–10]. Although a transaural system can accurately generate precisely defined signals at the ears of a listener with a control system of a realistic scale, its fundamental problem is that the control performance can be obtained only within a very tight space around the ears, and the movement of the listener's head results in inaccurate synthesis of the signals. Nelson *et al.* analyzed the properties of the reproduced sound fields with several geometrical arrangements of the loudspeakers in a transaural system and the influence of the geometrical arrangement of the loudspeakers on inverse filter design. As a result, they have presented the most suitable geometrical arrangement of the loudspeakers for overcoming the problem described above [11,12].

Furthermore, a hybrid type of sound field reproduction system, in which the advantage of binaural/transaural systems is combined with that of sound field reproduction systems based on Huygens' principle, has been proposed [13,14]. As described above, the improvement of each sound field reproduction system has been attempted by reexamining the geometrical structure of the control system, such as the arrangement of the loudspeakers or the control points. However, examinations in terms of the control theory, which is necessary to design the controller, have not been carried out. The issues that have been mainly researched concerning sound field reproduction techniques up to now basically belong to the open-loop control system. However, the sound field reproduction system may be realized as a feedback control system. There are advantages and disadvantages in using either type of control system. Among the advantages of a feedback control system are that, when perturbations of a plant exist, we can obtain a more accurate control effect with a feedback control system than with an open-loop control system. In other words, a feedback control system is relatively insensitive to system parameter variations. This favorable characteristic is theoretically deduced by transfer function analysis [15,16]. On the other hand, an inherent disadvantage of a feedback control system is the problem of stability. This is possibly the most important consideration when designing a feedback control system. In particular, as perturbations of a plant affect the stability of the closed-loop system, it is vital to design a controller such that the whole system remains stable when the perturbations occur within certain expected limits, i.e., the system possesses robust stability. The problem of designing such a controller has been investigated in terms of the robust control theory.

In the present study, we apply a feedback control scheme to a sound field reproduction system using loudspeakers, i.e., a transaural system. Since feedback control, as mentioned above, often exhibits better performance than open-loop control in tracking the output signals of a plant to desired signals when the model of the plant is imperfect, the control performance is expected to be less sensitive to the movement of the listener's head. For the synthesis of a controller that satisfies robust stability, the H_∞ control theory is employed. This is because the H_∞ control theory is an effective scheme for designing feedback controllers that accommodate both performance and stability in an optimal and robust manner.

2. H_∞ CONTROL THEORY

In this section, a brief review of H_∞ control theory is given. Since the details of the theory can be found in the literature [17,18], we present only the key points needed in the development of the proposed system: the standard H_∞ control problem is summarized; two typical control problems that can be solved by applying H_∞ control theory and that are related to problems of sound field reproduction are described.

2.1. Standard H_∞ Control Problem

Figure 2 shows a generalized control framework that is used to describe a control structure in H_∞ control theory. The framework is constructed of a generalized plant $\boldsymbol{G}(z)$ and a feedback controller $\boldsymbol{C}(z)$. In H_∞ control theory, the H_∞ norm of the transfer function $\boldsymbol{\Phi}(z)$ from the disturbance $\boldsymbol{W}(z)$ to the controlled variable $\boldsymbol{Z}(z)$ is adopted as a performance index, and $\boldsymbol{C}(z)$ that makes $\boldsymbol{\Phi}(z)$ asymptotically stable and satisfies

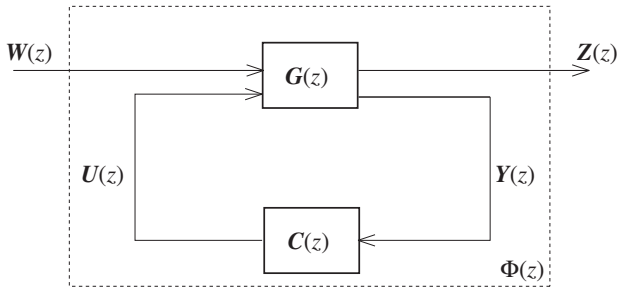

Fig. 2 General control framework.

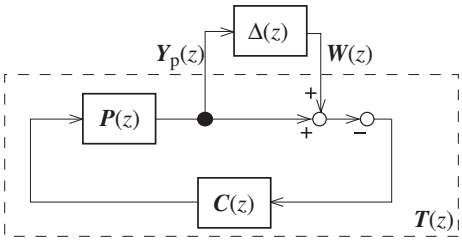

Fig. 3 Closed-loop system with multiplicative plant uncertainty.

$$\|\boldsymbol{\Phi}(z)\|_\infty < \gamma, \tag{1}$$

thereby minimizing $\boldsymbol{Z}(z)$, is to be found. This problem is called the standard H_∞ control problem, which can be analytically solved by several H_∞ algorithms.

2.2. Robust Stabilization Problem

Generally, it is difficult to model a plant accurately. Modeling errors are inevitable when a plant is identified either theoretically or experimentally. Perturbations of a plant may also occur owing to variations in physical conditions such as temperature, humidity, and geometry. These bring about deviations of the real plant from its nominal model; these are called plant uncertainties and affect the stability of the closed-loop system.

Figure 3 shows a closed-loop system containing plant uncertainty $\boldsymbol{\Delta}(z)\boldsymbol{P}(z)$. Here, $\boldsymbol{P}(z)$ is a nominal plant model, $\boldsymbol{C}(z)$ is a feedback controller, $\boldsymbol{T}(z)$ is the transfer function from $\boldsymbol{W}(z)$ to $\boldsymbol{Y}_\mathrm{p}(z)$, and $\boldsymbol{\Delta}(z)$ is a multiplicative uncertainty. To ensure the stability of the closed-loop system against this plant uncertainty, the following robustness condition derived from the small gain theorem must be satisfied:

$$\|\boldsymbol{W}_\mathrm{m}(z)\boldsymbol{T}(z)\|_\infty < 1, \tag{2}$$

where $\boldsymbol{W}_\mathrm{m}(z)$ satisfies

$$\sigma_{\max}[\boldsymbol{\Delta}(e^{j\Omega})] \le \sigma_{\max}[\boldsymbol{W}_\mathrm{m}(e^{j\Omega})], \quad 0 \le \Omega < 2\pi, \tag{3}$$

where $\sigma_{\max}[\cdot]$ denotes the maximum singular value. Finding $\boldsymbol{C}(z)$ that satisfies Eq. (3) is called the robust stability problem. If the closed-loop system shown in Fig. 3 is transformed into the system shown in Fig. 4 using $\boldsymbol{W}_\mathrm{m}(z)$, the transfer function from $\boldsymbol{W}(z)$ to $\boldsymbol{Z}_\mathrm{m}(z)$ is given by

$$\boldsymbol{Z}_\mathrm{m}(z) = \boldsymbol{W}_\mathrm{m}(z)\boldsymbol{T}(z)\boldsymbol{W}(z). \tag{4}$$

Thus, this problem can be treated as a standard H_∞ control problem, where $\boldsymbol{\Phi}(z) = \boldsymbol{W}_\mathrm{m}(z)\boldsymbol{T}(z)$ and $\gamma = 1$.

2.3. Disturbance Attenuation and Tracking Problem

Figure 5 shows a closed-loop system excited by disturbance $\boldsymbol{W}(z)$ and desired signal $\boldsymbol{R}(z)$. $\boldsymbol{S}(z)$ is the transfer function from $\boldsymbol{W}(z)$ to output $\boldsymbol{Y}(z)$ of the system.

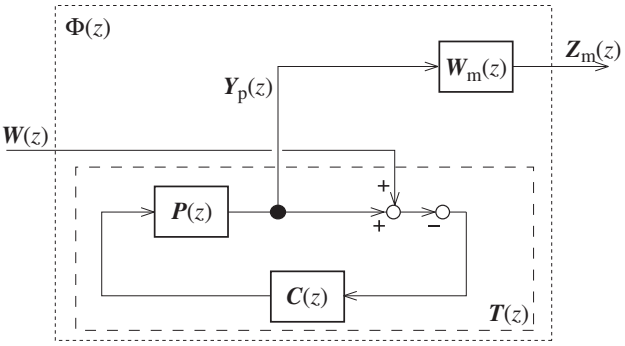

Fig. 4 System configuration for robust stability problem.

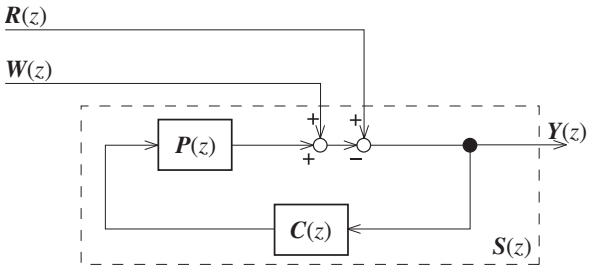

Fig. 5 Closed-loop system excited by disturbance and desired signal.

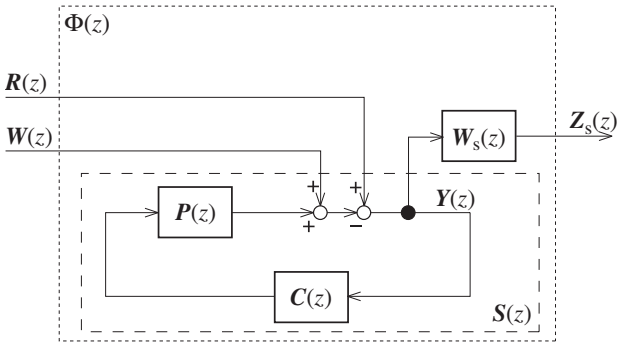

Fig. 6 System configuration for disturbance attenuation and tracking problem.

Note that the transfer function from $\boldsymbol{R}(z)$ to $\boldsymbol{Y}(z)$ is given by $-\boldsymbol{S}(z)$. By minimizing the gain of $\boldsymbol{S}(z)$, both the response of $\boldsymbol{Y}(z)$ to $\boldsymbol{W}(z)$ and the response of $\boldsymbol{Y}(z)$ to $\boldsymbol{R}(z)$ can be reduced. Thus, if the nominal performance condition expressed as

$$\|\boldsymbol{W}_\mathrm{s}(z)\boldsymbol{S}(z)\|_\infty < 1 \tag{5}$$

is satisfied under an appropriate weighting function $\boldsymbol{W}_\mathrm{s}(z)$, both disturbance attenuation and good tracking performance can be achieved. Finding $\boldsymbol{C}(z)$ that satisfies Eq. (5) is called the disturbance attenuation and tracking problem. If the closed-loop system shown in Fig. 5 is transformed into the system shown in Fig. 6 using $\boldsymbol{W}_\mathrm{s}(z)$, the transfer function from $\boldsymbol{W}(z)$ to $\boldsymbol{Z}_\mathrm{s}(z)$ is given by

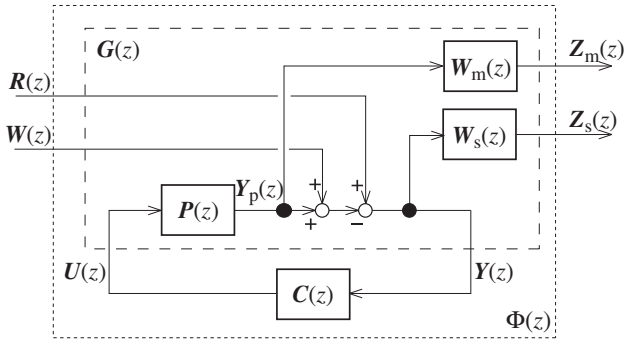

Fig. 7 System configuration for mixed sensitivity problem.

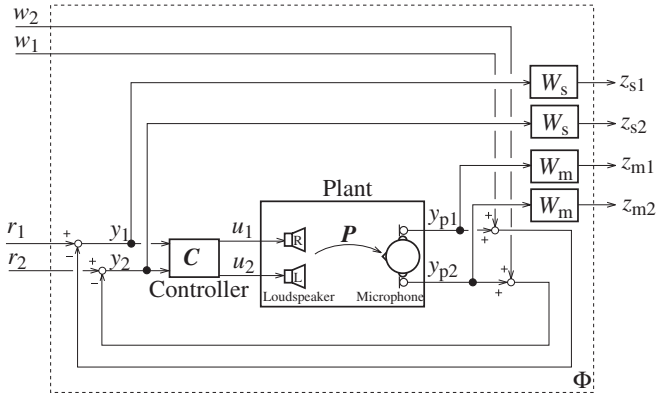

Fig. 8 Arrangement of proposed transaural sound reproduction system.

$$\boldsymbol{Z}_{\mathrm{s}}(z) = \boldsymbol{W}_{\mathrm{s}}(z)\boldsymbol{S}(z)\boldsymbol{W}(z). \tag{6}$$

Thus, one can also treat this problem as a standard H_∞ control problem, where $\boldsymbol{\Phi}(z) = \boldsymbol{W}_{\mathrm{s}}(z)\boldsymbol{S}(z)$ and $\gamma = 1$.

2.4. Mixed Sensitivity Problem

In most cases, the feedback controller is designed with regard to both robustness and performance. If the condition

$$\left\| \begin{bmatrix} \boldsymbol{W}_{\mathrm{m}}(z)\boldsymbol{T}(z) \\ \boldsymbol{W}_{\mathrm{s}}(z)\boldsymbol{S}(z) \end{bmatrix} \right\|_\infty < 1 \tag{7}$$

is satisfied, both the stability of the closed-loop system against plant uncertainty and tracking performance, i.e., robust performance, can be achieved. Finding $\boldsymbol{C}(z)$ that satisfies Eq. (7) is called the mixed sensitivity problem. Figure 7 shows the closed-loop system, in which the system shown in Fig. 4 is combined with the system shown in Fig. 6. It is easy to see that this problem can be solved as a standard H_∞ control problem by treating this system as the generalized control framework shown in Fig. 2.

3. PROPOSED TRANSAURAL SOUND REPRODUCTION SYSTEM

As described in Section 1, a transaural system makes the signals produced at the ears of the listener equal to the desired signals. Therefore, this system can clearly be treated as a tracking problem. The movement of the listener's head brings about changes in the transfer function from the loudspeakers to the ears, which is, in other words, a plant uncertainty. Thus, this issue can be treated as a robust stabilization problem. This consideration naturally suggests the arrangement of a transaural sound reproduction system shown in Fig. 8, to which the mixed sensitivity problem is introduced. Here, $\boldsymbol{P}$ is the plant to be controlled (a 2×2 matrix of transfer functions between inputs $\boldsymbol{u} = \{u_1, u_2\}^{\mathrm{T}}$ to the loudspeakers and outputs $\boldsymbol{y}_{\mathrm{p}} = \{y_{\mathrm{p1}}, y_{\mathrm{p2}}\}^{\mathrm{T}}$ of the microphones at the ears of the listener), $\boldsymbol{C}$ is the feedback controller (a 2×2 matrix of control filters), $\boldsymbol{r} = \{r_1, r_2\}^{\mathrm{T}}$ are the desired signals (binaural signals that produce the desired virtual auditory sensation when fed to the ears of the listener), W_{m} is the upper bound for multiplicative uncertainty, and W_{s} is an appropriate weighting function. Defining a generalized plant $\boldsymbol{G}(z)$ including not only $\boldsymbol{P}$ but also W_{m} and W_{s}, as shown in Fig. 7, enables this mixed sensitivity problem to be solved as a standard H_∞ problem.

Let $\boldsymbol{P}(z)$, $\boldsymbol{W}_{\mathrm{m}}(z)$, and $\boldsymbol{W}_{\mathrm{s}}(z)$ be represented in state space as follows:

$$\begin{cases} \boldsymbol{x}_{\mathrm{p}}(n+1) = \boldsymbol{A}_{\mathrm{p}}\boldsymbol{x}_{\mathrm{p}}(n) + \boldsymbol{B}_{\mathrm{p}}\boldsymbol{u}(n) \\ \boldsymbol{y}_{\mathrm{p}}(n) = \boldsymbol{C}_{\mathrm{p}}\boldsymbol{x}_{\mathrm{p}}(n) + \boldsymbol{D}_{\mathrm{p}}\boldsymbol{u}(n) \end{cases}, \tag{8}$$

$$\begin{cases} \boldsymbol{x}_{\mathrm{m}}(n+1) = \boldsymbol{A}_{\mathrm{m}}\boldsymbol{x}_{\mathrm{m}}(n) + \boldsymbol{B}_{\mathrm{m}}\boldsymbol{y}_{\mathrm{p}}(n) \\ \boldsymbol{z}_{\mathrm{m}}(n) = \boldsymbol{C}_{\mathrm{m}}\boldsymbol{x}_{\mathrm{m}}(n) + \boldsymbol{D}_{\mathrm{m}}\boldsymbol{y}_{\mathrm{p}}(n) \end{cases}, \tag{9}$$

$$\begin{cases} \boldsymbol{x}_{\mathrm{s}}(n+1) = \boldsymbol{A}_{\mathrm{s}}\boldsymbol{x}_{\mathrm{s}}(n) + \boldsymbol{B}_{\mathrm{s}}\{-\boldsymbol{y}_{\mathrm{p}}(n) - \boldsymbol{w}(n)\} \\ \boldsymbol{z}_{\mathrm{s}}(n) = \boldsymbol{C}_{\mathrm{s}}\boldsymbol{x}_{\mathrm{s}}(n) + \boldsymbol{D}_{\mathrm{s}}\{-\boldsymbol{y}_{\mathrm{p}}(n) - \boldsymbol{w}(n)\} \end{cases}, \tag{10}$$

with

$$\boldsymbol{u}(n) = \begin{Bmatrix} u_1(n) \\ u_2(n) \end{Bmatrix}, \ \boldsymbol{y}_{\mathrm{p}}(n) = \begin{Bmatrix} y_{\mathrm{p1}}(n) \\ y_{\mathrm{p2}}(n) \end{Bmatrix},$$
$$\boldsymbol{z}_{\mathrm{m}}(n) = \begin{Bmatrix} z_{\mathrm{m1}}(n) \\ z_{\mathrm{m2}}(n) \end{Bmatrix}, \ \boldsymbol{z}_{\mathrm{s}}(n) = \begin{Bmatrix} z_{\mathrm{s1}}(n) \\ z_{\mathrm{s2}}(n) \end{Bmatrix},$$
$$\boldsymbol{w}(n) = \begin{Bmatrix} w_1(n) \\ w_2(n) \end{Bmatrix}.$$

Combining Eqs. (8), (9), and (10) yields a state-space description for $\boldsymbol{G}(z)$ expressed as follows:

$$\begin{cases} \boldsymbol{x}(n+1) = \boldsymbol{A}\boldsymbol{x}(n) + \boldsymbol{B}_1\boldsymbol{w}(n) + \boldsymbol{B}_2\boldsymbol{u}(n) \\ \boldsymbol{z}(n) = \boldsymbol{C}_1\boldsymbol{x}(n) + \boldsymbol{D}_{11}\boldsymbol{w}(n) + \boldsymbol{D}_{12}\boldsymbol{u}(n) \\ \boldsymbol{y}(n) = \boldsymbol{C}_2\boldsymbol{x}(n) + \boldsymbol{D}_{21}\boldsymbol{w}(n) + \boldsymbol{D}_{22}\boldsymbol{u}(n) \end{cases}, \tag{11}$$

with

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{\mathrm{m}} & \boldsymbol{0} & \boldsymbol{B}_{\mathrm{m}}\boldsymbol{C}_{\mathrm{p}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathrm{s}} & -\boldsymbol{B}_{\mathrm{s}}\boldsymbol{C}_{\mathrm{p}} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\mathrm{p}} \end{bmatrix}, \ \boldsymbol{B}_1 = \begin{bmatrix} \boldsymbol{0} \\ -\boldsymbol{B}_{\mathrm{s}} \\ \boldsymbol{0} \end{bmatrix},$$

$$\boldsymbol{B}_2 = \begin{bmatrix} \boldsymbol{B}_m \boldsymbol{D}_p \\ -\boldsymbol{B}_s \boldsymbol{D}_p \\ \boldsymbol{B}_p \end{bmatrix},\ \boldsymbol{C}_1 = \begin{bmatrix} \boldsymbol{C}_m & \boldsymbol{0} & \boldsymbol{D}_m \boldsymbol{C}_p \\ \boldsymbol{0} & \boldsymbol{C}_s & -\boldsymbol{D}_s \boldsymbol{C}_p \end{bmatrix},$$

$$\boldsymbol{D}_{11} = \begin{bmatrix} \boldsymbol{0} \\ -\boldsymbol{D}_s \end{bmatrix},\ \boldsymbol{D}_{12} = \begin{bmatrix} \boldsymbol{D}_m \boldsymbol{D}_p \\ -\boldsymbol{D}_s \boldsymbol{D}_p \end{bmatrix},\ \boldsymbol{C}_2 = [\boldsymbol{0} \quad \boldsymbol{0} \quad -\boldsymbol{C}_p],$$

$$\boldsymbol{D}_{21} = -\boldsymbol{I},\ \boldsymbol{D}_{22} = -\boldsymbol{D}_p,$$

$$\boldsymbol{x}(n) = \begin{Bmatrix} \boldsymbol{x}_m(n) \\ \boldsymbol{x}_s(n) \\ \boldsymbol{x}_p(n) \end{Bmatrix},\ \boldsymbol{z}(n) = \begin{Bmatrix} \boldsymbol{z}_m(n) \\ \boldsymbol{z}_s(n) \end{Bmatrix},\ \boldsymbol{y}(n) = \begin{Bmatrix} y_1(n) \\ y_2(n) \end{Bmatrix}.$$

State-space description (11) allows us to synthesize $\boldsymbol{C}(z)$ by using H_∞ algorithms based on the state space approach [19–22].

4. NUMERICAL STUDIES OF PROPOSED SYSTEM

4.1. Design of H_∞ controller

We applied the proposed method to two-channel sound reproduction and evaluated its robustness to the movement of the listener's head by a computer simulation. Figure 9 shows the geometrical arrangement used for the simulation. The objective is to produce both an equalized response at one ear of the dummy head and cross-talk cancellation at the other ear. This two-input two-output acoustic system was first modeled with measured head-related transfer functions (HRTFs) from L and R loudspeakers to the left and right ears. We used a database comprising directionally discrete HRTFs on a virtual spherical surface 1.4 m from a KEMAR dummy head, which can be obtained from MIT Media Lab. The measurements were made in MIT's anechoic chamber, using a small two-way loudspeaker with a 4-inch woofer and 1-inch tweeter, at a sampling frequency of 44.1 kHz [23].

In general, an efficient model that can represent an acoustic system to be controlled with few parameters is preferred because the scale of the controller depends on the number of parameters. The proposed method is such a case. Thus, the measured HRTFs, i.e., all-zero models, were transformed into a more efficient model such that the scale of the feedback controller to be implemented becomes as small as possible. In this work, we employed the singular value decomposition (SVD) method for efficient modeling. The SVD method is the state-space model-based system identification technique and has been developed in the outline of the modern control theory. The authors introduced this method for modeling enclosed sound fields and verified that it has several advantages, such as direct identification of a state-space description of a multi-input multi-output acoustic system, as expressed by Eq. (8), and high precision of identification with few parameters [24].

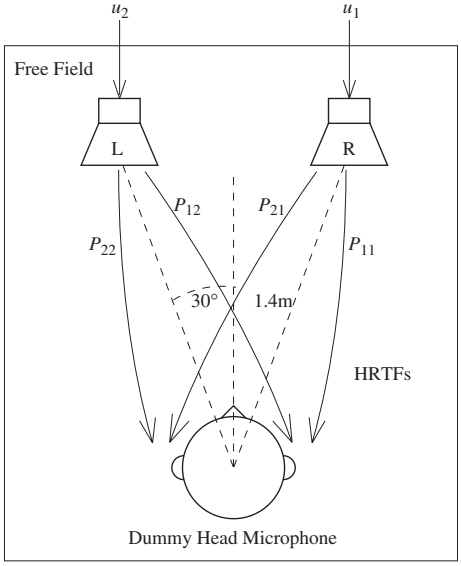

Fig. 9 Geometrical arrangement of loudspeakers and dummy head for cross-talk cancellation.

This fact naturally suggests that the SVD method is also suitable for HRTF representation because HRTFs have the same kind of properties as enclosed sound fields such as acoustical resonance. The dimension N_p of the state-space description (8) of the two-input two-output acoustic system to be identified by the SVD method was determined to achieve the normalized mean-squared modeling error of -35 dB. The resultant dimension was $N_p = 256$. If the state-space description (8) is transformed into its controllability canonical form, one finds that the number of arbitrary coefficients (degree-of-freedom) of the acoustic system is given by $4N_p + 5$ [25]. Thus the total number of arbitrary coefficients is 1,029. Note that, if each HRTF is modeled by the conventional all-zero model, i.e., a truncated version of the impulse response of the HRTF, the model needs about 375 coefficients to achieve the same precision. In this case, as the two-input two-output acoustic system is composed of four HRTFs, the total number of its arbitrary coefficients is clearly about 1,500. Figure 10 shows an example of the identification results. The dashed curve is a measured HRTF from the R loudspeaker to the right ear, and the solid curve is its SVD model. It can be seen that the HRTF is accurately identified with the SVD model of the chosen dimension. This mathematical model of the HRTFs was taken as a nominal plant $\boldsymbol{P}(z)$.

The real plant frequency response varies as a function of the head position. These variations were modeled using a multiplicative uncertainty as follows:

$$\Delta(z) = \frac{P'(z) - P(z)}{P(z)}, \tag{12}$$

where $P'(z)$ is the transfer function of a perturbed plant and $P(z)$ is that of a nominal plant. In the two-channel case

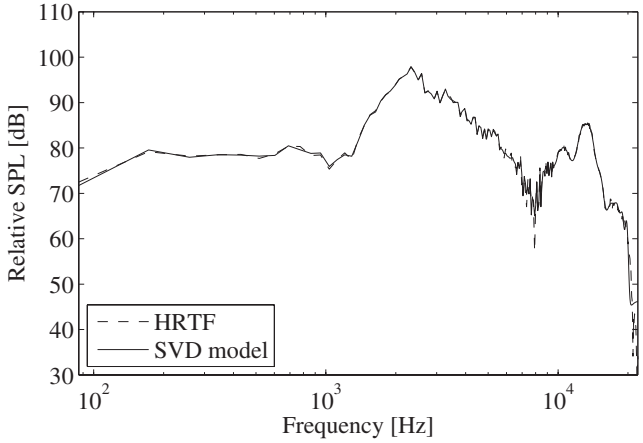

Fig. 10 Frequency response functions of HRTF and its SVD model.

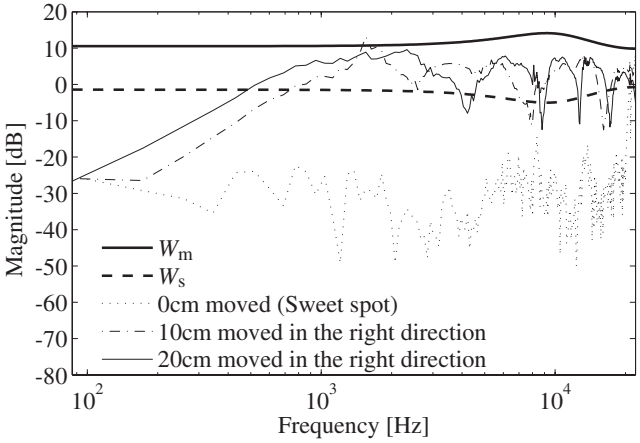

Fig. 11 Multiplicative plant uncertainties and the weighting functions.

concerned, the sum of the HRTFs from L and R loudspeakers to the right ear and the sum of those to the left ear with the dummy head placed at three different positions were taken as the perturbed plants. When the position of the dummy head was between sampled directions of the used HRTF database [23], its HRTF was obtained by bilinear interpolation of magnitude and phase spectra in the frequency domain. When the dummy head was positioned at a different distance from the database, its HRTF was obtained by extrapolation with an appropriately chosen fractional delay [26] and spherical attenuation. The calculated plant uncertainties $\Delta(z)$ are illustrated in Fig. 11. The upper bound $W_m(z)$ for the multiplicative uncertainties and the weighting function $W_s(z)$ for the disturbance attenuation and tracking problem were determined as follows. Firstly we chose a candidate of $W_m(z)$ such that Eq. (3) was satisfied for the small gain theorem, i.e., $\Delta(z)$ was bounded by $W_m(z)$. In order for the scale of the feedback controller to be synthesized as small as possible, $W_m(z)$ was realized as a rather low-order IIR filter. After such $W_m(z)$ was chosen for robustness, we then chose the weighting function $W_s(z)$ so that $W_m(z)$ and $W_s(z)$ could satisfy a trade-off between the robustness and performance. In this work, $W_s(z)$ was simply set to the reciprocal of $W_m(z)$ except for its gain. On the basis of these procedures, we searched for $W_m(z)$ and $W_s(z)$ with which the mixed sensitivity problem could be solved, i.e., the feedback controller could be obtained, by trial and error. The resultant $W_m(z)$ and $W_s(z)$ were eighth-order IIR filters, and are plotted in Fig. 11. Finally, by using an H_∞ algorithm based on the state space approach, as mentioned in Section 3, the feedback controller was synthesized as the following state-space description:

$$\begin{cases} \boldsymbol{x}_c(n+1) = \boldsymbol{A}_c\boldsymbol{x}_c(n) + \boldsymbol{B}_c\boldsymbol{y}(n) \\ \boldsymbol{u}(n) = \boldsymbol{C}_c\boldsymbol{x}_c(n) + \boldsymbol{D}_c\boldsymbol{y}(n) \end{cases}. \tag{13}$$

In addition to the feedback controller, an open-loop controller based on the conventional inverse filtering method was also designed for comparison. The open-loop controller was composed of four FIR filters calculated by the least-squares method in the time domain. The number of taps of each FIR filter was determined as follows. As described above, the number of arbitrary coefficients of the state-space description (13), i.e., the scale of the feedback controller, is given by $4N_c + 5$ [25], where N_c is the dimension of the state vector $\boldsymbol{x}_c$. On the other hand, when the number of taps of each FIR filter is L, the number of arbitrary coefficients of the open-loop controller is clearly given by $4L$. In order to make the scale of the open-loop controller nearly equal to that of the feedback controller, it is required that L be set to $N_c + 1$. Since the feedback controller synthesized in this computer simulation had 272 state variables ($N_c = 272$), the open-loop controller was realized such that each FIR filter had 273 taps. It is well known that an inverse filter with a larger number of taps, i.e., a larger open-loop controller, gives a more accurate inversion of an acoustic system. In this work, as a preliminary study, we assess the control performance of the proposed method under the condition that the scale of the designed controller is as small as possible. This condition can be regarded as reasonable because, in practical situations, hardware constraints will usually limit the maximum filter length possible for real-time operation. This is the reason why we designed the open-loop controller to have such small taps.

A modeling delay included to ensure that the FIR filters were causal was set to 38 points, which was identical to the number of initial delay points of the net impulse response of the proposed transaural sound reproduction system. Note that, the proposed method intrinsically does not involve a modeling delay in its design parameters, while the conventional inverse filtering method requires it. Since the value

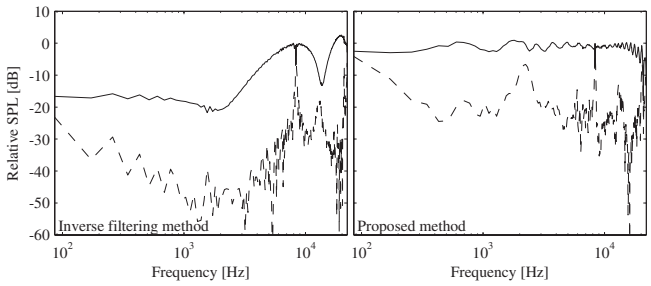

Fig. 12 Frequency response functions from the desired signal r_1 to the right ear y_{p1} (solid line) and left ear y_{p2} (dashed line) without movement of the head.

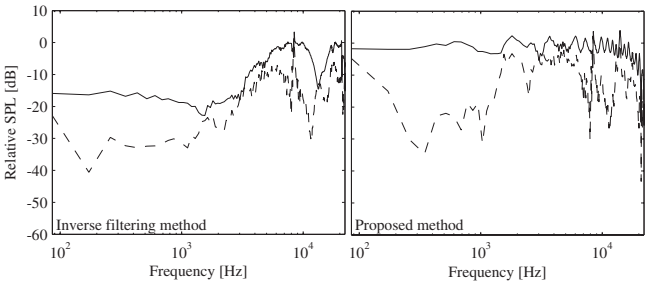

Fig. 13 Frequency response functions from the desired signal r_1 to the right ear y_{p1} (solid line) and left ear y_{p2} (dashed line) with movement of 3 cm of the head.

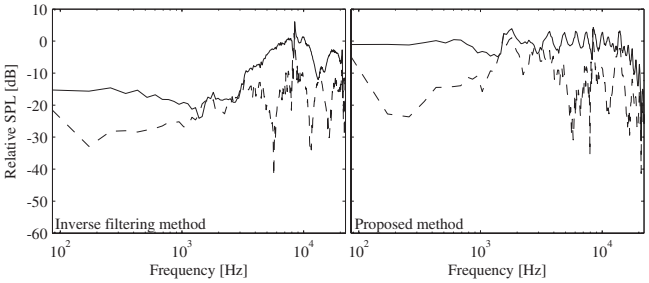

Fig. 14 Frequency response functions from the desired signal r_1 to the right ear y_{p1} (solid line) and left ear y_{p2} (dashed line) with movement of 6 cm of the head.

of a modeling delay corresponds to the sample number at which the peak of the impulse response of the convolution of an acoustic system with its inverse filter occurs, i.e., the initial delay time of the net response of the system, we set the modeling delay such that the initial delay time of the net response of the open-loop system is equal to that of the proposed system. It is also well known that a relatively long modeling delay in the design of an inverse filter assists in the accurate inversion of an acoustic system. However, since making the initial delay time as small as possible is meaningful for some applications of a sound field reproduction system (for example, applications including synchronization of the acoustic signal and video signal in real-time reproduction) from the practical point of view, we designed the open-loop controller to have such a small modeling delay.

4.2. Results and Discussion

Figure 12 shows the frequency response functions from the desired signal r_1 to the left and right ears without the movement of the dummy head (at the sweet spot). In other words, this demonstrates the frequency responses at the left and right ears when the input signal to the right channel is an impulse and that to the left channel is 0, i.e., the effect of cross-talk cancellation. The left graph shows the frequency responses with the conventional inverse filtering method, and the right one shows those with the proposed method. It can be seen that the proposed method effectively produces not only equalization at the right ear, but also cross-talk cancellation. The inverse filtering method cannot equalize the response at the right ear well. This is because the modeling delay used to design the inverse filters was not long enough. Note that the arrival time of the sound reproduced by the proposed method is equal to that in the case of the inverse filtering method; thereby, comparison of the performance is done under the identical condition. For cross-talk cancellation, the inverse filtering method exhibits better performance than the proposed method.

Figure 13 shows the frequency response functions with the dummy head moved 3 cm away from the sweet spot in the right direction. In this case, the proposed method is more effective than the inverse filtering method, particularly in the low-frequency range. The result when the dummy head was moved 6 cm away is shown in Fig. 14. One can see that, although the control performance with the inverse filtering method is degraded, that with the proposed method does not deteriorate very much.

To investigate quantitatively the spatial extent of the effect of the transaural sound reproduction system, a performance index defined as

$$\beta = 20 \sum_{k=k_1}^{k=k_2} \frac{1}{k} \log\left(\frac{|Y_{p1}(k)|}{|Y_{p2}(k)|}\right) \Bigg/ \sum_{k=k_1}^{k=k_2} \frac{1}{k} \quad \text{[dB]} \qquad (14)$$

was calculated. $Y_{p1}(k)$ denotes the discrete Fourier transform of the equalized signal $y_{p1}(n)$, and $Y_{p2}(k)$ denotes that of the canceled signal $y_{p2}(n)$. This performance index was used in a previous work [8], and represents the average amount of suppression of $Y_{p2}(k)$ to $Y_{p1}(k)$, which is equally weighted for any octave band. The larger this performance index, the more effective the cross-talk cancellation. Figure 15 shows the relationship of the performance index to the lateral and fore-and-aft movement of the dummy head for the inverse filtering method. Figure 16 shows the same for the proposed method. Note that these density plots are clipped for values over 10 dB and under 0 dB. The performance index for the inverse filtering method decreases with an increase of lateral deviation in position.

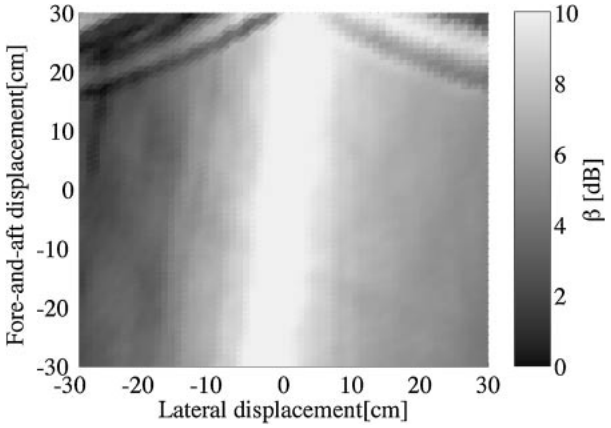

Fig. 15 Relationship of performance index β to movement of the head for the inverse filtering method.

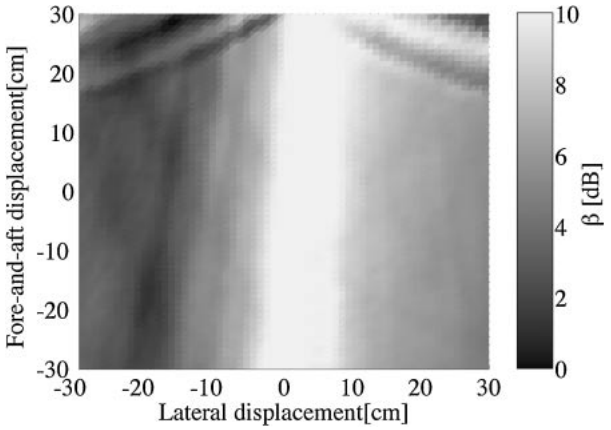

Fig. 16 Relationship of performance index β to movement of the head for the proposed method.

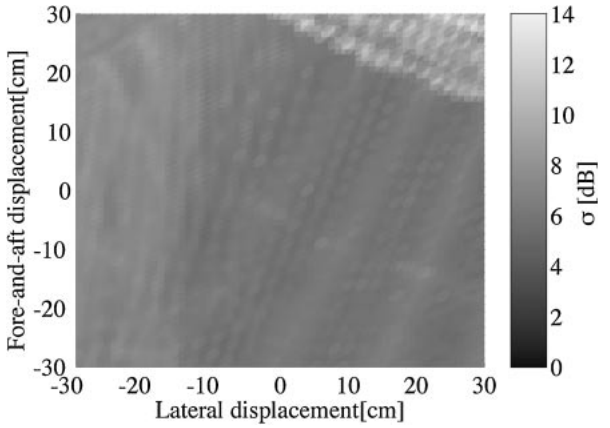

Fig. 17 Relationship of performance index σ to movement of the head for the inverse filtering method.

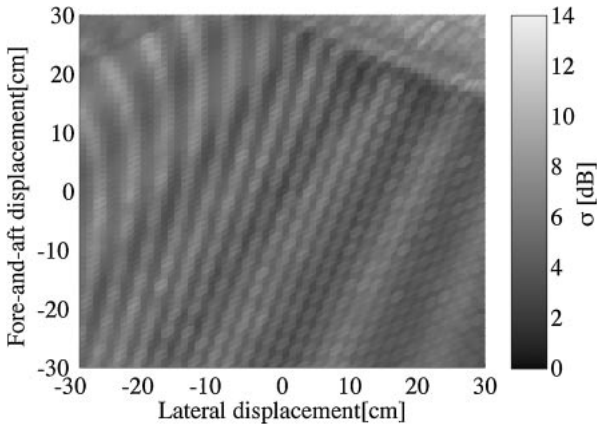

Fig. 18 Relationship of performance index σ to movement of the head for the proposed method.

The width of the lateral displacement that maintains β larger than 10 dB (width of the white belt) is roughly 7.5 cm. The performance index for the proposed method also decreases with an increase of lateral deviation in position. However, it can be seen that much larger displacement is allowed for the proposed method. The width of the lateral displacement that maintains β larger than 10 dB is roughly 10 cm. This indicates that the cross-talk cancellation with the proposed method is better preserved for deviations in position than it is with the conventional inverse filtering method. At the sweet spot, however, the value of β for the proposed method is 16.9 dB, whereas that for the inverse filtering method is 24.2 dB; the inverse filtering method can perform more effective cross-talk cancellation than the proposed method. This feature is also presented by the frequency response function reported in Fig. 12.

In addition to the effect of cross-talk cancellation, we assessed the degree of flatness of the equalized frequency response precisely. For that purpose, the following performance index was calculated:

$$\sigma = \left[\frac{1}{N} \sum_{k=k_1}^{k_2} \{20 \log(|Y_{\mathrm{p1}}(k)|) - r\}^2 \right]^{\frac{1}{2}} \quad [\mathrm{dB}], \qquad (15)$$

with

$$r = \frac{1}{N} \sum_{k=k_1}^{k_2} 20 \log(|Y_{\mathrm{p1}}(k)|),$$

where N is the number of frequency samples between k_1 and k_2. This index is the standard deviation of the modulus of $Y_{\mathrm{p1}}(k)$, and represents the scale of the peaks and dips of the equalized frequency response. The smaller the performance index, the flatter the equalized frequency response. Figure 17 shows the relationship of the performance index to the movement of the dummy head for the inverse filtering method. Figure 18 shows that for the proposed method. For both cases, although there is a slight fluctuation in the performance index, its value remains almost constant, on average, over the spatial range shown in the figure. However, it can be seen that, at most positions, the performance index for the proposed method is smaller than that for the inverse filtering method. In fact, the average

value of σ for the proposed method is 5.4 dB, whereas that for the inverse filtering method is 7.2 dB. This demonstrates that the proposed method can perform more effective equalization, on average, in space than the inverse filtering method. In this computer simulation, neither filter length nor modeling delay of the inverse filters was large enough to produce the satisfactory performance of inversion. However, as described in Subsection 4.1, in practical situations, it is normal that hardware constraints will limit the maximum scale of the controller and/or that the specification of an applied system will require the reduction of the latency of the system to the lowest possible level. Even in such a situation, the proposed method can provide an acceptable performance of inversion.

5. CONCLUSIONS

A transaural sound reproduction system adopting the feedback control approach was proposed in this work. This method involves modeling HRTFs as a control plant by the SVD method and introducing H_∞ control theory to synthesize the feedback controller. Performance, stability, and robustness of the feedback system were met by using H_∞ control theory. Computer simulations demonstrated that the proposed system exhibited better performance than the conventional system based on the inverse filtering method in both the equalization of acoustic transfer functions and cross-talk cancellation when the listener's head moved. Thus, the proposed scheme is useful for extending the "sweet spot" of transaural systems.

REFERENCES

[1] M. J. M. Jessel and G. A. Mangiante, "Active sound absorbers in an air duct," *J. Sound Vib.*, **23**, 383–390 (1972).

[2] G. A. Mangiante, "Active sound absorption," *J. Acoust. Soc. Am.*, **61**, 1516–1523 (1977).

[3] K. Furuya and H. Ichinose, "Sound field control in a closed space using boundary-pressure-method," *IEICE Tech. Rep.*, **90**(26), pp. 25–32 (1990) (in Japanese).

[4] A. J. Berkhout, D. de Vries and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, **93**, 2764–2778 (1993).

[5] B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," *J. Audio Eng. Soc.*, **9**, 148–151 (1961).

[6] P. Damaske, "Head-related two-channel stereophony with loudspeaker reproduction," *J. Acoust. Soc. Am.*, **50**, 1109–1115 (1971).

[7] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Int. Conv. Rec.*, Part 7, pp. 150–155 (1963).

[8] H. Hamada, "Construction of orthostereophonic system for the purposes of quasi-insitu recording and reproduction," *J. Acoust. Soc. Jpn.* (*J*), **39**, 337–348 (1983) (in Japanese).

[9] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust. Speech Signal Process.*, **36**, 145–152 (1988).

[10] P. A. Nelson, H. Hamada and S. J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," *IEEE Trans. Signal Process.*, **40**, 1621–1632 (1992).

[11] O. Kirkeby, P. A. Nelson and H. Hamada, "Local sound field reproduction using two closely spaced loudspeakers," *J. Acoust. Soc. Am.*, **104**, 1973–1981 (1998).

[12] T. Takeuchi and P. A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *J. Acoust. Soc. Am.*, **112**, 2786–2797 (2002).

[13] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *Acustica — Acta Acustica*, **85**, 78–87 (1999).

[14] S. Takane, Y. Suzuki, T. Miyajima and T. Sone, "A new theory for high definition virtual acoustic display named ADVISE," *Acoust. Sci. & Tech.*, **24**, 276–283 (2003).

[15] M. Ito, *Jido Seigyo Gairon* (Shokodo, Tokyo, 1991), Chap. 4 (in Japanese).

[16] T. Samejima, "Algorithms for feedback active control," *J. Acoust. Soc. Jpn.* (*J*), **59**, 407–413 (2003) (in Japanese).

[17] B. A. Francis, *A Course in H_∞ Control Theory* (Springer-Verlag, Berlin, 1987).

[18] B. M. Chen, *H_∞ Control and Its Applications* (Springer-Verlag, Berlin, 1998).

[19] K. Glover and J. C. Doyle, "State space formulae for all stabilizing controllers that satisfy an H_∞-norm bound and relations to risk sensitivity," *Syst. Control Lett.*, **11**, 167–172 (1988).

[20] J. C. Doyle, K. Glover, P. P. Khargonekar and B. A. Francis, "State space solutions to standard H_2 and H_∞ control problems," *IEEE Trans. Autom. Control*, **34**, 831–847 (1989).

[21] M. G. Safonov, D. J. N. Limebeer and R. Y. Chiang, "Simplifying the H^∞ theory via loop-shifting, matrix-pencil and descriptor concepts," *Int. J. Control*, **50**, 2467–2488 (1989).

[22] K. Glover, D. J. N. Limebeer, J. C. Doyle, E. M. Kasenally and M. G. Safonov, "A characterization of all solutions to the four block general distance problem," *SIAM J. Control Optimization*, **29**, 283–324 (1991).

[23] http://sound.media.mit.edu/resources/KEMAR.html

[24] Y. Sasaki and T. Samejima, "A study of a multi-input/output system identification of a sound field based on experimental modal analysis," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 637–638 (2009) (in Japanese).

[25] T. Yoshikawa and J. Imura, *Modern Control Theory* (Shokodo, Tokyo, 1995), Chap. 3 (in Japanese).

[26] T. I. Laakso, V. Välimäki, M. Karjalainen and U. K. Laine, "Splitting the unit delay: Tools for fractional delay filter design," *IEEE Signal Process. Mag.*, **13**, 30–60 (1996).