Sound Field Reproduction System Tracking Environmental Variations With Deconvolution

Yuji Nobe and Yoshinobu Kajikawa
Kansai University, Suita 564-8680 Osaka Japan
E-mail: kaji@kansai-u.ac.jp Tel: +81-6-6368-0828

***Abstract*—In this paper, we propose a novel sound field reproduction system using the deconvolution. In the conventional sound reproduction systems, the preprocessing filters are generally determined and fixed based on transfer functions from loudspeakers to control points in advance. However, movements of control points result in severe localization errors. Therefore, we have already proposed a sound field reproduction system using the simultaneous perturbation (SP) method which updates the filter coefficients only using error signal. However the SP method suffers from the disadvantage of slow convergence. Hence, we newly propose a sound field reproduction system using Deconvolution Algorithm (DA) which has the advantage of the fast convergence. The DA updates the filter coefficients using the signals which input to the loudspeakers and the observing microphones. Simulation results demonstrate that the proposed method can track the movements of control points and has reasonable convergence speed.**

I. Introduction

Recently, sound field reproduction systems have been actively studied. Sound field control techniques with loudspeakers can be classified into two methods which control large areas and narrow points, respectively.

The method of controlling large areas uses the sound field control theory based on "the Kirchhoff-Helmholtz Integral Equation". A desired sound field can be reproduced in the controlled area by matching the sound pressure on the closed surface which encloses a certain area and the particle velocity of the normal direction to those of a target sound field by using "the Kirchhoff-Helmholtz Integral Equation". However, a great number of loudspeakers are needed according to the sampling theorem of the space if the sound field is controlled over the audio bandwidth of 20kHz[1]. Such an approach is impractical.

On the other hand, the latter method reproduces any desired sound at any desired point (e.g. listener's ears). To do this, the method of controlling points needs to remove the influence of transfer functions between loudspeakers and control points and crosstalk paths, i.e. the so-called crosstalk canceller. The crosstalk canceller can be realized using MINT(Multiple input-output INverse Theorem [2]), LNS(Least Norm Solution [3]), and so on. In this method, the number of loudspeakers can be generally reduced compared with the method of controlling areas. Hence, we focus on the method of controlling points because of the simple implementation.

In this method, the preprocessing filters are generally determined and fixed from transfer functions from loudspeakers to control points, which are measured in advance. Hence, the movements of control points and the changes of environments (e.g. temperature) result in severe localization errors, especially over higher frequency band.

To solve this problem, the preprocessing filters have to track the variation of transfer functions and consequently use an adaptive algorithm. The conventional adaptive algorithms must know the information of the variation for transfer functions from loudspeakers to control points. However, it would be difficult to obtain the precise information of the variation for the sound field reproduction systems in operation because of the crosstalk paths. Therefore, we previously have proposed a sound field reproduction system using the simultaneous perturbation (SP) method with delay control filters [4]. However, this method suffers from the disadvantage of slow convergence. Therefore, we propose a fast convergence algorithm which is called Deconvolution Algorithm (DA) using multi channel deconvolution and MINT. In DA, the multi channel deconvolution is used for the estimation of transfer functions' matrix for indefiniteness solutions and MINT updates the preprocessing filter coefficients using the transfer functions estimated by deconvolution.

II. Sound Reproduction System with Deconvoluiton Algorithm

A. Multi Channel Deconvolution

Figure 1 shows a conceptual diagram of the sound field reproduction system with the DA. Even if a listener moves, a desired sound field can be always reproduced by updating the preprocessing filters with the DA as shown in Fig. 1.

Figure 2 shows a transoral system which consists of 4 secondary sound sources and 2 control points. In this figure, i is the secondary sound source number, and j the control point number, $\mathbf{h}_{ij}$ the coefficient vector of the preprocessing filter, $\mathbf{g}_{ij}$ the coefficient vector of the transfer function, k the time, N the tap length of preprocessing filter, M the tap length of transfer function, and n the block number which indicates the period $k = N(n-1) \sim Nn-1$. Multi channel deconvolution involved in the DA is used for the estimation of the preprocessing filter coefficients' solution matrix and the transfer functions' matrix for indefiniteness solutions. To begin

with, we define the vectors using block number n in frequency domain for input signal $x_j(k)$, input signal to the observing microphones $y_j(k)$, $\mathbf{h}_{ij}$ and $\mathbf{g}_{ij}$ as follows:

$$\mathbf{X}_j(n) = diag\left\{\text{FFT}\left[x_j(nN-2N),\cdots,x_j(nN-1)\right]^T\right\} \quad (1)$$

$$\mathbf{Y}_j(n) = diag\left\{\text{FFT}\left[y_j(nN-2N),\cdots,y_j(nN-1)\right]^T\right\} \quad (2)$$

$$\mathbf{H}_{ij}(n) = diag\left\{\text{FFT}\left[h_{ij,0}(n),\cdots,h_{ij,N-1}(n),0\cdots,0\right]^T\right\} \quad (3)$$

$$\mathbf{G}_{ij}(n) = diag\left\{\text{FFT}\left[g_{ij,0}(n),\cdots,g_{ij,M-1}(n),0\cdots,0\right]^T\right\} \quad (4)$$

Besides, the estimation procedure at block n by the multi channel deconvolution is defined as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix}^T \quad (5)$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 & \mathbf{Y}_2 \end{bmatrix}^T \quad (6)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{21} & \mathbf{G}_{31} & \mathbf{G}_{41} \\ \mathbf{G}_{12} & \mathbf{G}_{22} & \mathbf{G}_{32} & \mathbf{G}_{42} \end{bmatrix} \quad (7)$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \\ \mathbf{H}_{31} & \mathbf{H}_{32} \\ \mathbf{H}_{41} & \mathbf{H}_{42} \end{bmatrix} \quad (8)$$

$$\mathbf{Y} = \mathbf{GHX} \quad (9)$$

$$\hat{\mathbf{H}} = \mathbf{HXY}^{-} \quad (10)$$

$$\hat{\mathbf{G}} = \mathbf{Y}(\mathbf{HX})^{-} \quad (11)$$

where $\mathbf{Y}^{-}$ is the generalized inverse matrix of $\mathbf{Y}$ and $\hat{\mathbf{G}}$ an estimation of $\mathbf{G}$. We may denote the relationship among input signal matrix $\mathbf{X}$, observing signal matrix $\mathbf{Y}$, preprocessing filter matrix $\mathbf{H}$, and transfer function matrix $\mathbf{G}$ defined by Eqs. (5)∼(8), as shown in Eq.(9). By transforming Eq.(9), we may express $\hat{\mathbf{H}}$ and $\hat{\mathbf{G}}$ as Eqs. (10) and (11), respectively. Finally, we may achieve $\hat{\mathbf{h}}_{ij}$ and $\hat{\mathbf{g}}_{ij}$ which are estimations of $\mathbf{h}_{ij}$ and $\mathbf{g}_{ij}$ in time domain, as follows:

$$\hat{\mathbf{h}}_{ij} = \left[\hat{h}_{ij,0},\hat{h}_{ij,1},\cdots,\hat{h}_{ij,N-1}\right]^T \quad (12)$$
$$= \text{first } N \text{ elements of IFFT } \left[\hat{\mathbf{H}}_{ij}\right]$$

$$\hat{\mathbf{g}}_{ij} = \left[\hat{g}_{ij,0},\hat{g}_{ij,1},\cdots,\hat{g}_{ij,N-1}\right]^T \quad (13)$$
$$= \text{first } N \text{ elements of IFFT } \left[\hat{\mathbf{G}}_{ij}\right]$$

Note that $\hat{\mathbf{h}}_{ij}$ and $\hat{\mathbf{g}}_{ij}$ with the multi channel deconvolution are indefiniteness solutions. Therefore, $\hat{\mathbf{g}}_{ij}$ is not the exact estimation of $\mathbf{g}_{ij}$ for all combinations of i and j, respectively. However, $\hat{\mathbf{G}}$ may come to an estimation of $\mathbf{G}$. Likewise, $\hat{\mathbf{H}}$ also has the same characteristic.

However, $\hat{\mathbf{h}}_{ij}$ and $\hat{\mathbf{g}}_{ij}$ estimated from Eqs. (5)∼(13) are less-accurate estimations. This is because of fact that the estimations using Eqs. (10) and (11) depend on the initial values for $\mathbf{H}$ of Eq. (8). For instance, if the initial values of $\mathbf{H}_{ij}(i = 1 \sim 4, j = 1, 2)$ are the same, $\hat{\mathbf{H}}_{i1}(i = 1 \sim 4)$ becomes equal to $\hat{\mathbf{H}}_{i2}(i = 1 \sim 4)$. Moreover, this characteristic is the same for $\hat{\mathbf{G}}$. Additionally, if we set the initial values of $\mathbf{H}$ at random, the precise of $\hat{\mathbf{H}}$ would not vary greatly.

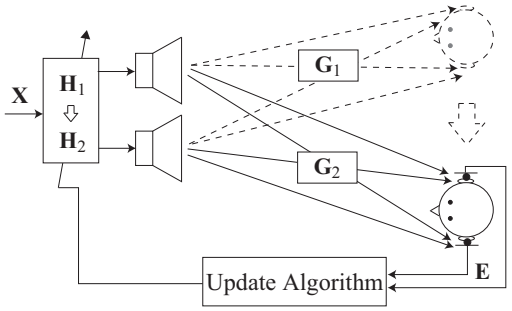


Fig. 1. Proposed transoral system.

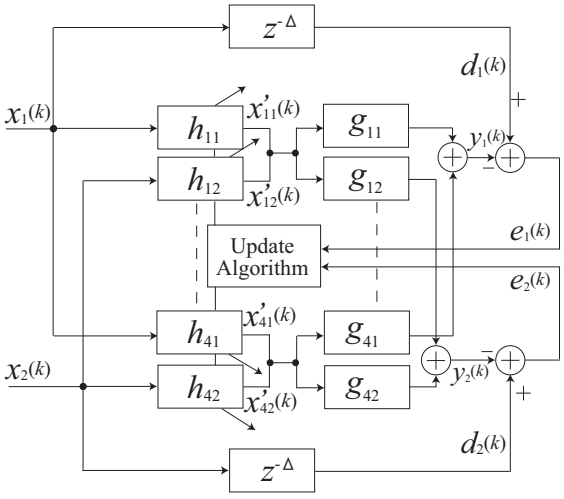


Fig. 2. Transoral system using the multi channel deconvolution algorithm.

B. Deconvolution Algorithm

In this section, we describe the Deconvolution Algorithm for sound reproduction system. This proposed method typically uses random values for the initial coefficients of the preprocessing filter $\mathbf{H}$. Besides, the following evaluation formulas are used for the split of the update process on the filter coefficients in this algorithm.

$$\text{Cor} = \sum_{j=1}^{2} \frac{\mathbf{y}_j^T \mathbf{e}_j}{||\mathbf{y}_j||\,||\mathbf{e}_j||} \quad (14)$$

$$\text{RA} = 10\log_{10}\frac{\sum\left(d_1(k)^2 + (d_2(k)^2\right)}{\sum\left(e_1(k)^2 + (e_2(k)^2\right)} \quad (15)$$

Eq. (14) expresses the sum of correlation factors between the observed signal vector $\mathbf{y}_j = [y_j(0), y_j(1), \cdots, y_j(N-1)]^T$ and error signal vector $\mathbf{e}_j = [e_j(0), e_j(1), \cdots, e_j(N-1)]^T$. Furthermore, $d_j(k)$ is the desired signal and Eq. (15) is Reproduction Accuracy (RA) which evaluates the precision of the preprocessing filters. Besides, Cor shown in Eq. (14) decreases as the update of the preprocessing filters proceeds to the convergence and becomes negative when the update proceeds to the divergence.

When RA $< T$ or Cor < 0, the update processes for the preprocessing filter coefficients and the estimation of transfer functions are conducted according to the following equations:

TABLE I
SIMULATION CONDITIONS IN SOUND REPRODUCTION SYSTEM.

Sampling frequency	44100 Hz
Frequency Range	400-10000 Hz
Tap length of, h_{ij}, g_{ij} (i=1,2,3,4, j=1,2)	256
Stepsize Parameter μ(Proposed)	3.5×10^{-2}
Stepsize Parameter μ (SP method)	3.5×10^{-5}
Threshold level T	3[dB]
Number of average of Deconvolution A	50
Forgetting factor λ	0.85
Delay u, w_1	128

$$\hat{\mathbf{H}}(n+A) = \lambda\mathbf{H}(n) + \frac{(1-\lambda)}{A}\sum_{m=n}^{n+A-1}\mathbf{H}(n)\mathbf{X}(m)\mathbf{Y}^{-}(m) \quad (16)$$

$$\hat{\mathbf{G}}(n+A) = \lambda\hat{\mathbf{G}}(n) + \frac{(1-\lambda)}{A}\sum_{m=n}^{n+A-1}\mathbf{Y}(m)(\mathbf{H}(n)\mathbf{X}(m))^{-} \quad (17)$$

where T is the threshold level, λ the forgetting factor, and A the average sample number for the input signals. Note that these update processes are block processing. That is, n is the block number in case that RA $< T$ or Cor < 0, and the preprocessing filters are fixed throughout the calculations of Eqs. (16) and (17). On the other hand, when RA $> T$ and Cor > 0, the update process for the preprocessing filter coefficients is conducted using MINT [2]. MINT using the transfer functions estimated by the deconvolution is defined, as follows:

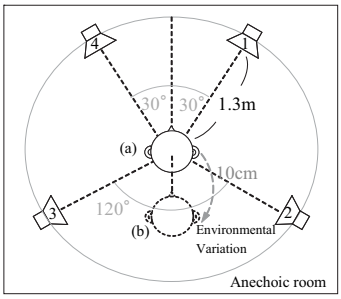
$$\mathbf{h}_{ij}(k+1) = \mathbf{h}_{ij}(k) + \mu\sum_{l=1}^{2} e_l(k)\mathbf{r}_{il}(k) \quad (18)$$

where μ is the step-size parameter, and $\mathbf{r}_{il}$ the input signals filterd with the estimations of transfer functions. The precision of the filter coefficients can be greatly improved by using Eq. (18) rather than only deconvolution.

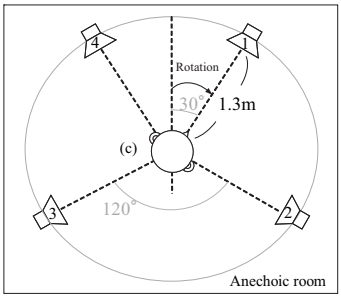
III. COMPUTER SIMULATION

In this section, the validity of the proposed method (DA) is verified through some computer simulations. This verification is conducted through comparing the DA with the conventional method (the SP method) and the SP method with delay control filters[4]. The simulation conditions are shown in Table I. The transfer functions measured in the loudspeaker arrangement as shown in Figure 3 are used for the computer simulations. In this figure, the position (b) is 10 cm backward from the position (a), and the position (c) is rotated rightward 30 degrees at the position (a). In the simulations, Case1 is the case where the listener moves from (a) to (b), and Case2 is the case where the listener's head rotates from (a) to (c). White noise which is limited within 400 - 10000 Hz is used for the input signal, and the listener moves in 30 seconds. Besides, the initial values of the preprocessing filter coefficients are set to LNS derived from transfer functions measured in practice.

First, we show simulation results on the estimation for the preprocessing filters and transfer functions. Figure 4 shows the impulse responses and frequency spectra on the path $\mathbf{g}_{11}$ which is from loudspeaker No.1 to the right ear at position (a). This figure shows $\mathbf{g}_{11}$ and $\hat{\mathbf{g}}_{11}$ estimated when RA is 25dB

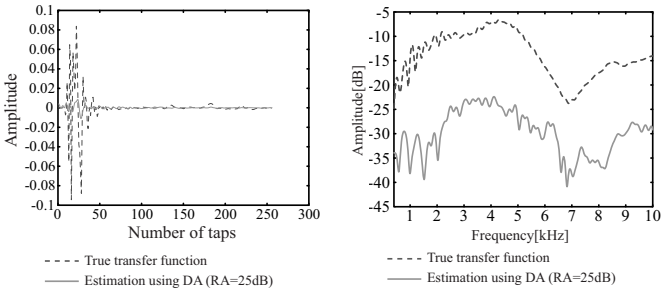


(A) Before and after backward movement



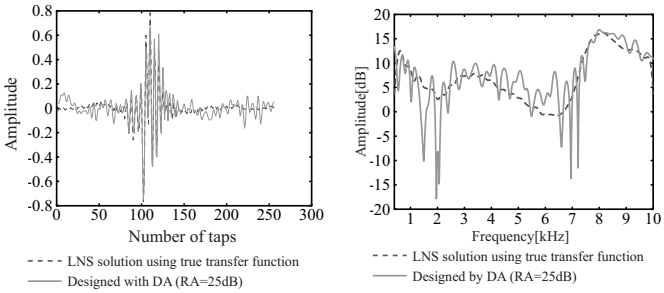
(B) After rotational movement

Fig. 3. Loudspeaker arrangement.



(A) Time domain (B) Frequency domain

Fig. 4. Comparison between true transfer function and estimated one.



(A) Time domain (B) Frequency domain

Fig. 5. Preprocessing filters designed via LNS and DA.

by the DA. Furthermore, Figure 5 shows the filter coefficients and frequency spectra of $\mathbf{h}_{11}$. This figure shows LNS and the estimation by DA when RA is 25dB. It can be seen from Fig. 4 that $\hat{\mathbf{g}}_{11}$ differs from $\mathbf{g}_{11}$ despite the fact RA=25dB. Likewise, LNS differs from the corresponding one estimated by the DA from Fig. 5. However, those estimations have good accuracy in overall matrices sense. This characteristic will be described later.

Next, we show simulation results on the convergence properties of RA. Figure 6 shows the convergence properties in

Case1 and Case2. These figures include three convergence properties using the SP method, the improved SP method, and the DA. It can be seen from Case1 of Fig. 6 that the convergence speed can be improved by using the delay control filters in the SP method. However, the convergence speed of the DA is farther improved than those of the other methods. It can be seen from Case2 of Fig. 6 that the improved SP method is invalid for rotational variations. On the other hand, the DA has still fast convergence speed in Case2. These facts indicate that the DA is superior to any SP methods without complicated processes.

Finally, we show that the preprocessing filters estimated by the DA are able to compensate any distortional sound field. Now, if the optimum preprocessing filter matrix $\mathbf{H}_{opt}$ is designed, $\mathbf{G}_f\mathbf{H}_{opt,f} = \mathbf{I}_f$ would be obvious at each frequency f. Note that $\mathbf{I}$ is the unit matrix. Now let $\mathbf{G}_f\hat{\mathbf{H}}_f = \mathbf{M}_f$. In this case, we can evaluate the estimation precision of the preprocessing filters by evaluating whether $\hat{\mathbf{M}}_f$ is near to 2×2 unit matrices at all frequencies. Hence, we define Diagonal Power of Matrix Difference (DPMD) which is an evaluating metrics of the sound field compensation at each frequency, as follows:

$$\mathbf{M}_f = \begin{bmatrix} a_f & b_f \\ c_f & d_f \end{bmatrix} \quad (19)$$

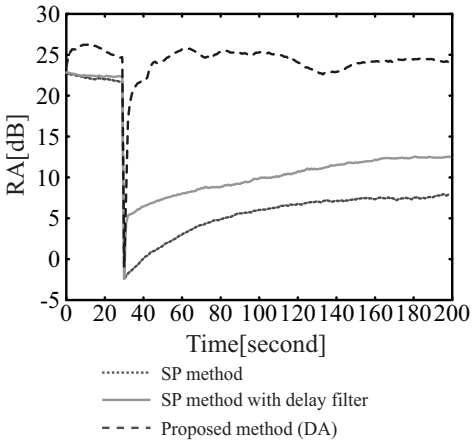
$$\mathrm{DPMD}(\mathbf{M}_f) = \frac{|a_f| + |d_f| - |b_f| - |c_f|}{2} \quad (20)$$

DPMD approaches 1 if $\mathbf{M}_f$ is near to the unit matrix. In addition, DPMD approaches 1 at all frequencies if $\mathbf{H}_{opt}$ is designed. We show DPMD frequency characteristics when $\mathbf{H}$ is designed by various algorithm in Fig. 7. In this figure, LNS and MINT are designed using the true tranfer function matrix $\mathbf{G}$ and cannot track any environmental variations. Note that RA in the parenthesis shows Reproduction Accuracy of the preprocessing filters designed by each algorithm. From Fig. 7, the values of DPMD are almost 1 at all frequencies on MINT because of using the true transfer functions. On the other hand, the DA which is the proposed method is inferior to MINT, but the precision can be improved more than only deconvolution. As a result, the precision of the preprocessing filter matrix designed by the DA may achive the same level as LNS.

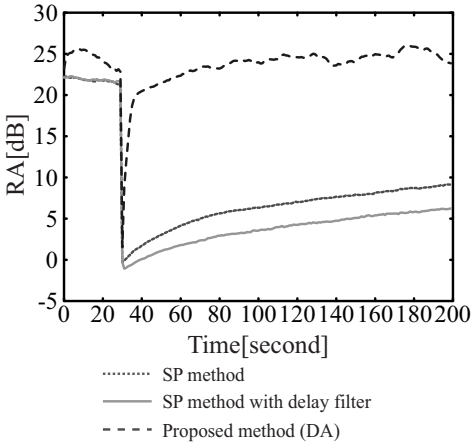
The proposed method is therefore effective for the sound field reproduction system with environmental variations.

IV. Conclusions

In this paper, we have proposed a novel sound field reproduction system using the DA. The proposed system can fast track any movements of controlling points and consequently has wide controlled areas compared with the conventional one. However, when the strongly colored signals are used, the proposed method becomes unstable. We will explore an improved algorithm for the colored signals in the future.



(A) Case1



(B) Case2

Fig. 6. Convergence properties of RA.

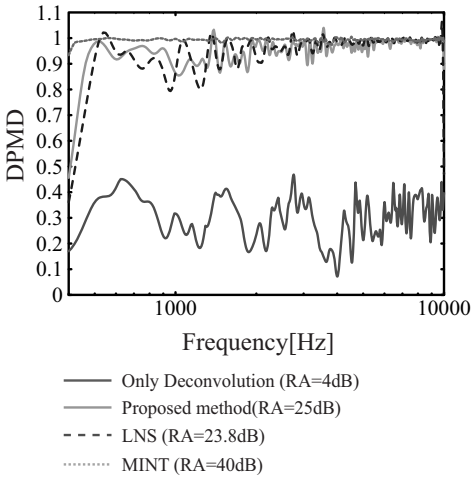


Fig. 7. DPMD characteristics.

References

[1] Y. A. Huang and J. Benesty, *Audio Signal Processing*. London: Kluwer Academic Publishers, 2004.

[2] M. Miyoshi and Y. Kaneda, "Inverse Filtering of Room Acoustics," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 142–152, Feb. 1988.

[3] Y. Tatekura, H. Saruwatari, K. Shikano, "An Iterative Inverse Filter Design Method for the Multichannel Sound Field Reproduction System.", *IEICE Trans. on Fundamentals*, vol. E84-A, no.4, pp. 991–998, Apr. 2001.

[4] K. Tsukamoto, Y. Kajikawa and Y. Nomura, "Sound Reproduction System with Simultaneous Perturbation Method." in *Proc. EUSIPCO 2006*, Florence, Italy, Sep. 2006.