PAPER Special Section on Multi-channel Acoustic Signal Processing

On-Line Relaxation Algorithm Applicable to Acoustic Fluctuation for Inverse Filter in Multichannel Sound Reproduction System

Yosuke TATEKURA[†a)], *Member*, Shigefumi URATA[††], *Nonmember*, Hiroshi SARUWATARI[††b)], *and* Kiyohiro SHIKANO[††], *Members*

SUMMARY In this paper, we propose a new on-line adaptive relaxation algorithm for an inverse filter in a multichannel sound reproduction system. The fluctuation of room transfer functions degrades reproduced sound in conventional sound reproduction systems in which the coefficients of the inverse filter are fixed. In order to resolve this problem, an iterative relaxation algorithm for an inverse filter performed by truncated singular value decomposition (adaptive TSVD) has been proposed. However, it is difficult to apply this method within the time duration of the sound of speech or music in the original signals. Therefore, we extend adaptive TSVD to an on-line-type algorithm based on the observed signal at only one control point, normalizing the observed signal with the original sound. The result of the simulation using real environmental data reveals that the proposed method can always carry out the relaxation process against acoustic fluctuation, for any time duration. Also, subjective evaluation in the real acoustic environment indicates that the sound quality improves without degrading the localization.

key words: *sound reproduction, room transfer function, inverse filter, relaxation of inverse filter, on-line adaptation*

1. Introduction

To achieve a sound reproduction system with loudspeaker reproduction, for example, 'transaural stereo' [1], it is important to design inverse filters which cancel the effects of room transfer functions (RTFs). The RTFs vary depending on environmental variations (such as variations in the speed of sound due to temperature fluctuations and changes in reflection conditions due to changes in the locations of indoor items) and are not time invariant. Therefore, the reproduction accuracy is degraded by environmental variations in the sound reproduction system using fixed inverse filter coefficients. Also, if unstable inverse filters which enlarge the original signal are used, the variation of the RTFs causes a deterioration in sound quality, and it becomes necessary to either re-estimate the RTFs after the variations or to adaptively relax the inverse filters.

As an adaptive design procedure for the inverse filters, a method has been proposed for updating the inverse filter coefficients using reference microphones set up at several control points [2], [3]. In this procedure, it is possible to adaptively update the inverse filters to compensate for variations of room configuration or of temperature. However, since these reference microphones must be placed in close proximity to the ears of the listener, hearing is significantly impaired. For re-estimation of the RTFs that vary due to changes in room temperature, a method has been proposed for extending or contracting the time axis of the impulse responses [4]. However, only temperature variations are dealt with by this method; the variation of the reflection due to changes in room configuration cannot be handled.

One method for preventing the degradation of sound quality by environmental changes is to relax the inverse filters. In general, inverse filters are designed by deriving the inverse of the matrix consisting of the impulse responses of the RTFs. If the linear independence of the column vectors comprising this matrix is low, there is a danger that this inverse matrix may expand the solution in the presence of a small amount of error. Methods for resolving this problem include relaxation methods involving the regularization method [5] and the truncated singular value decomposition (TSVD) method [6]. In both methods, the parameters for performing the relaxation of the inverse filters (i.e., the regularization coefficients or truncation number) can be determined only by considering the RTF matrix. However, it is likely that sufficient control accuracy cannot be obtained as indicated by the fact that excessively relaxed inverse filters can be designed.

To resolve this problem, Nagata et al. have proposed a method for the relaxation of the inverse filters based on adaptive TSVD [7]. In this method, the relaxation of inverse filters is autonomously performed depending on the amount of expansion of the observed noise. Also, a monitoring microphone can be placed at a location that does not restrict the listener. However, it is difficult to adapt the inverse filters within the time period which contains the sound of speech or music in the original signals. Therefore, we develop the adaptive TSVD into an on-line type algorithm. In this paper, we propose a new on-line adaptive relaxation algorithm for inverse filters based on the observed signal at only one control point, which normalizes the observed signal with the original sound.

The organization of this paper is as follows. In Sect. 2, the configuration of the sound reproduction system is described and a design method for the inverse filter in a multichannel sound reproduction system is presented. An on-line

Manuscript received October 25, 2004.
Manuscript revised January 21, 2005.
Final manuscript received March 11, 2005.

[†]The author is with Faculty of Engineering, Shizuoka University, Hamamatsu-shi, 432-8561 Japan.
[††]The authors are with Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0101 Japan.
a) E-mail: tytatek@ipc.shizuoka.ac.jp
b) E-mail: sawatari@is.naist.jp

DOI: 10.1093/ietfec/e88–a.7.1747

relaxation method for an inverse filter is described in Sect. 3. In Sect. 4, the effectiveness of the proposed method is studied by a simulation experiment. In Sect. 5, a subjective evaluation of the sound quality and the sound localization accuracy is carried out by applying the proposed method to a real environment. Conclusions are presented in Sect. 6.

2. Design Method for Relaxed Inverse Filter

2.1 Inverse Filter Design with Moore-Penrose Generalized Inverse Matrix

In this section, we describe the design method for the inverse filter in the frequency domain. In the following, we assume a multichannel sound reproduction system with M secondary sound sources and N control points, as shown in Fig. 1. We define the matrices representing the RTF, inverse filter, original sound source signal, and reproduced sound as $\mathbf{G}(\omega)$, $\mathbf{H}(\omega)$, $\mathbf{X}(\omega)$, and $\hat{\mathbf{X}}(\omega)$, respectively. The matrices can be expressed as follows.

$$\mathbf{G}(\omega) = \begin{bmatrix} G_{11}(\omega) & G_{12}(\omega) & \dots & G_{1M}(\omega) \\ G_{21}(\omega) & G_{22}(\omega) & \dots & G_{2M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ G_{N1}(\omega) & G_{N2}(\omega) & \dots & G_{NM}(\omega) \end{bmatrix} \tag{1}$$

$$\mathbf{H}(\omega) = \begin{bmatrix} H_{11}(\omega) & H_{12}(\omega) & \dots & H_{1N}(\omega) \\ H_{21}(\omega) & H_{22}(\omega) & \dots & H_{2N}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ H_{M1}(\omega) & H_{M2}(\omega) & \dots & H_{MN}(\omega) \end{bmatrix} \tag{2}$$

$$\mathbf{X}(\omega) = [X_1(\omega), X_2(\omega), \cdots, X_N(\omega)]^T \tag{3}$$

$$\hat{\mathbf{X}}(\omega) = \left[\hat{X}_1(\omega), \hat{X}_2(\omega), \cdots, \hat{X}_N(\omega)\right]^T \tag{4}$$

Here, ω denotes the frequency, $G_{ji}(\omega)$ is the RTF and $H_{ij}(\omega)$ is the inverse filter coefficient. i $(= 1, 2, \cdots, M)$ is the order of the secondary sound source and j $(= 1, 2, \cdots, N)$ is the order of the control point. $X_j(\omega)$ is the original sound reproduced at control point j and $\hat{X}_j(\omega)$ is the reproduced sound at control point j.

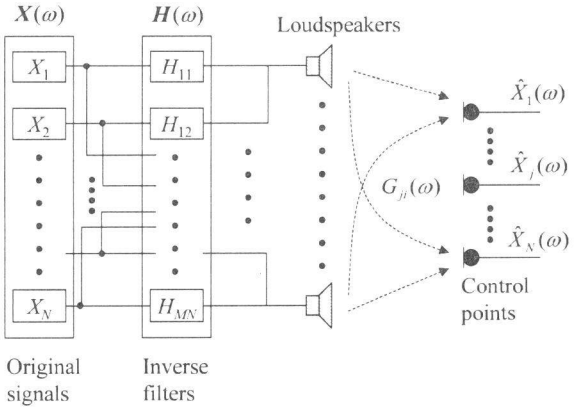

Fig. 1 Multichannel sound reproduction system.

The reproduced signal $\hat{\mathbf{X}}(\omega)$ can be expressed as follows in terms of the matrices given above:

$$\hat{\mathbf{X}}(\omega) = \mathbf{G}(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega). \tag{5}$$

Since our objective is to achieve control such that $\mathbf{X}(\omega) = \hat{\mathbf{X}}(\omega)$ in sound reproduction, the inverse filter $\mathbf{H}(\omega)$ can be obtained by deriving the inverse matrix of the room transfer function $\mathbf{G}(\omega)$. That is, the inverse filter design method is reduced to solving the following linear equation:

$$\mathbf{G}(\omega)\mathbf{H}(\omega) = \mathbf{I}_N, \tag{6}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. The inverse filter $\mathbf{H}(\omega)$ can be derived as the generalized inverse matrix of $\mathbf{G}(\omega)$ in the case of $M > N$. Since the solution becomes underdetermined if there is no rank reduction, the generalized Moore-Penrose (MP) inverse matrix with the least norm solution (LNS) [8] is used. In the following, the generalized MP inverse matrix of $\mathbf{G}(\omega)$ is expressed as $\mathbf{G}^{\dagger}$. In order to derive the generalized MP inverse matrix, the singular value decomposition (SVD) of $\mathbf{G}(\omega)$ is carried out as follows:

$$\mathbf{G}(\omega) = \mathbf{U}(\omega) \cdot [\mathbf{\Gamma}_N(\omega),\ \mathbf{O}_{N,M-N}] \cdot \mathbf{V}^{\mathrm{H}}(\omega) \tag{7}$$

$$\mathbf{\Gamma}_N(\omega) \equiv \mathrm{diag}\left[\mu_1(\omega), \dots, \mu_N(\omega)\right], \tag{8}$$

where $\mathbf{U}(\omega)$ is the $N \times N$ orthogonal matrix, $\mathbf{V}(\omega)$ is the $M \times M$ orthogonal matrix, $\mathbf{V}^{\mathrm{H}}(\omega)$ is the conjugate transposed matrix of $\mathbf{V}(\omega)$, and $\mathbf{O}_{N,M-N}$ is the $N \times (M - N)$ null matrix. Also, $\mu_k(\omega)$ $(k = 1, \dots, N,\ \mu_k(\omega) \geq \mu_{k+1}(\omega))$ denotes the singular values. By using Eq. (7), the generalized MP inverse matrix of $\mathbf{G}(\omega)$, $\mathbf{G}^{\dagger}(\omega)$, can be given by

$$\mathbf{G}^{\dagger}(\omega) = \mathbf{V}(\omega) \cdot \begin{bmatrix} \mathbf{\Lambda}_N(\omega) \\ \mathbf{O}_{M-N,N} \end{bmatrix} \cdot \mathbf{U}^{\mathrm{H}}(\omega), \tag{9}$$

where

$$\mathbf{\Lambda}_N(\omega) \equiv \mathrm{diag}\left[\xi_1(\omega), \dots, \xi_N(\omega)\right] \tag{10}$$

$$\xi_k(\omega) = \begin{cases} \frac{1}{\mu_k(\omega)} & \mu_k(\omega) \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{11}$$

By computing the inverse matrix $\mathbf{G}^{\dagger}(\omega)$ for each frequency, the inverse filter $\mathbf{H}(\omega)$ can be designed.

2.2 Relaxation of Inverse Filter

The inverse filter can be designed on the basis of the impulse response measurements of the RTFs. However, the measured impulse responses may contain infinitesimal noise. Also, due to environmental variations, the inverse filters may not accurately emulate the inverse characteristics of the actual transfer characteristics. For these reasons, the noise components contained in the original signal may be amplified in sound reproduction. This is caused mainly by a low linear independence of the column vectors composing the matrix $\mathbf{G}(\omega)$ which expresses the transfer characteristics used in the inverse filter design [9].

Such amplification of the noise components is not desirable from the viewpoint of sound quality, and therefore,

relaxation of the instability must be performed. One conventional relaxation method used in sound reproduction systems is regularization [10]. However, it is necessary to calculate the inverse matrix at each frequency every time the regularization parameter is determined. Also, the parameter must be optimized at each frequency.

Another relaxation method is truncated singular value decomposition (TSVD). When the generalized MP inverse matrix is derived by SVD, it is necessary to be cautious about the existence of small singular values. This is because they may contain round-off errors and may have a low linear independence, so that the norm of the solution may be expanded. Hence, the solutions of the inverse matrix obtained with all singular values have the potential to be unstable. Therefore, by limiting the number of singular values used in the SVD of the matrix, the inverse matrix is stabilized. The inverse matrix with relaxation by TSVD can be obtained by replacing $\xi_k(\omega)$ with 0 even if $\mu_k(\omega) \neq 0$ and $\mu_k(\omega) \cong 0$ in Eq. (11). Then, the number of singular values in the matrix $\mathbf{G}(\omega)$ replaced with 0 is called the truncation number on ω.

3. On-Line Adaptive TSVD Method

3.1 Outline

In the adaptive TSVD which Nagata et al. have proposed [7], the relaxation of the inverse filter can be carried out using the noise signal amplified by variations of the RTFs. By applying this method, amplification of the noise is autonomously reduced and the sound quality can be improved. However, a silent duration in the original sound source is used for the adaptation of the inverse filters. Therefore, this method cannot adapt the inverse filters when the original sound source does not contain the silent duration.

To resolve this problem, we propose an on-line adaptive algorithm that can always adapt within any time duration. Figure 2 shows an overview of the multichannel sound reproduction system based on our method. The microphone for signal observation is set apart from the listener so that it does not prevent agreeable listening. In this section, our proposed algorithm, which is called "on-line adaptive TSVD" (OATSVD) in this paper, is described. This algorithm introduces the truncation number based on the observed signal and relaxes the inverse filters by TSVD. First, in order to observe the fluctuation of RTFs, a silent signal (zero signal) is reproduced at the control point for signal observation. To eliminate the effect of the original signal, the observed signal is normalized by the original signal. Next, the truncation number of the singular value is obtained using the normalized observed signal, and the relaxed inverse filters are designed.

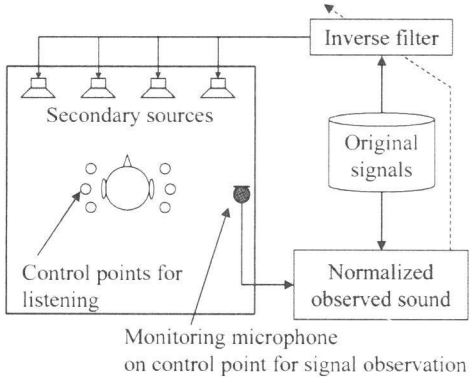

Fig. 2 Sound reproduction system based on proposed method.

3.2 Reproduction of Silent Signal

In OATSVD, the silent signal (zero signal) is reproduced at the control point for signal observation, which is set apart from the listener as shown in Fig. 2. In the cases in which the RTFs do not fluctuate, the observed signal $\hat{\mathbf{X}}(\omega)$ can be written as Eq. (5). If the RTFs $\mathbf{G}(\omega)$ are fluctuated into $\hat{\mathbf{G}}(\omega)$, we assume that the fluctuated RTFs $\hat{\mathbf{G}}(\omega)$ is given by

$$\hat{\mathbf{G}}(\omega) = \mathbf{G}(\omega) + \Delta\mathbf{G}(\omega), \tag{12}$$

where

$$\Delta\mathbf{G}(\omega) = \begin{bmatrix} \Delta G_{11}(\omega) & \Delta G_{12}(\omega) & \cdots & \Delta G_{1M}(\omega) \\ \Delta G_{21}(\omega) & \Delta G_{22}(\omega) & \cdots & \Delta G_{2M}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \Delta G_{N1}(\omega) & \Delta G_{N2}(\omega) & \cdots & \Delta G_{NM}(\omega) \end{bmatrix} \tag{13}$$

expresses the difference between the original RTFs $\mathbf{G}(\omega)$ and the fluctuated RTFs $\hat{\mathbf{G}}(\omega)$. In such a case, the observed signal $\hat{\mathbf{X}}(\omega)$ can be expressed as

$$\begin{aligned} \hat{\mathbf{X}}(\omega) &= \hat{\mathbf{G}}(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega) \\ &= \mathbf{X}(\omega) + \Delta\mathbf{G}(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega). \end{aligned} \tag{14}$$

Let the Nth control point be the control point for signal observation. The observed signal $\hat{X}_N(\omega)$ observed at the monitoring microphone is also given by

$$\hat{X}_N(\omega) = X_N(\omega) + \Delta\mathbf{G}_N(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega), \tag{15}$$

where $\Delta\mathbf{G}_N = [\Delta G_{N1}(\omega), \Delta G_{N2}(\omega), \cdots, \Delta G_{NM}(\omega)]$ and $X_N(\omega)$ is the original sound signal for the monitoring control point. Here, $\hat{X}_N(\omega)$ is regarded as the error signal obtained by the fluctuated RTFs if $X_N(\omega)$ is zero (silent signal).

3.3 Normalization of Observed Signal by Original Signals

From Eq. (15), the observed signal $\hat{X}_N(\omega)$ is influenced by the original signals $\mathbf{X}(\omega)$. Because the original signals are time variant, it is difficult to obtain only the fluctuation of the RTFs $\Delta\mathbf{G}(\omega)$ from the result of the observed signal. Therefore, we introduce the normalized power spectrum level (in dB scale) of the observed signal normalized by the original signals at the i-th iteration, $P^{[i]}_{\mathrm{norm}}(\omega_n)$, which is defined as

$$P_{\text{norm}}^{[i]}(\omega_n) = 10\log_{10}\left\{\frac{1}{L}\sum_{k=n-\lfloor\frac{L}{2}\rfloor}^{n+\lfloor\frac{L}{2}\rfloor} R^{[i]}(\omega_k)\right\}^2 \tag{16}$$

$$R^{[i]}(\omega_n) = \frac{\hat{X}_N^{[i]}(\omega_n)}{X_{\text{ave}}^{[i]}(\omega_n)} \tag{17}$$

$$X_{\text{ave}}^{[i]}(\omega_n) = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}\left|X_j^{[i]}(\omega_n)\right|^2}, \tag{18}$$

where $X_{\text{ave}}^{[i]}(\omega_n)$ is the averaged amplitude spectrum of the original signal at the control point for listening, $X_j^{[i]}(\omega_n)$ $(j = 1, \cdots, N-1)$. Here, L is the window length of the moving average, index $^{[i]}$ denotes the iteration time i, and ω_n is the discrete frequency with index n. Hereafter, the observed signal which is normalized by the original signal is called the normalized observed signal.

3.4 Update Algorithm of Inverse Filter

Figure 3 shows the flow for updating the inverse filter based on OATSVD. In the first frame, the original sound is reproduced with initial inverse filters in the sound reproduction system, and the reproduced sound is observed by the monitoring microphone. Using the observed sound signal, the inverse filters are updated. In the next frame, the original sound is reproduced with the updated inverse filters. Using above-mentioned procedure, we can obtain the relaxed inverse filters.

The update algorithm falls into three steps: (1) perform sound reproduction with the inverse filters obtained by TSVD, and observe the reproduced sound, (2) normalize reproduced sound and determine the truncation number, and (3) back to first step.

Let the truncation number at each stage of the iterative updating process be $l_i(\omega_n)$, the variation sign of the normalized observed signal be $p_i(\omega_n)$, and the variation magnitude

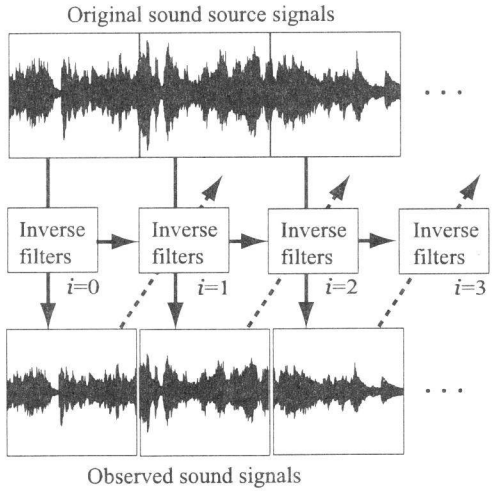

Fig. 3 Flow of proposed algorithm.

of the truncation number in the stages between $(i-1)$ and i be $a_i(\omega_n)$.

[Step 0] Let us initialize as $l_0(\omega_n) = 0$, $p_0(\omega_n) = 1$ and $a_0(\omega_n) = 0$.

[Step 1] Using the relationship expressed by Eq. (9), an inverse filter $\hat{\mathbf{H}}^{[i]}(\omega_n)$ relaxed using $l_i(\omega_n)$ is designed:

$$\hat{\mathbf{H}}^{[i]}(\omega_n) = \mathbf{V}(\omega_n)\cdot\begin{bmatrix}\hat{\mathbf{\Lambda}}_N^{[i]}(\omega_n)\\ \mathbf{O}_{M-N,N}\end{bmatrix}\cdot\mathbf{U}^{\mathrm{H}}(\omega_n), \tag{19}$$

where

$$\hat{\mathbf{\Lambda}}_N^{[i]}(\omega_n) \equiv \text{diag}\left[\frac{1}{\mu_1(\omega_n)},\cdots,\frac{1}{\mu_{N-l_i(\omega_n)}(\omega_n)},\mathbf{o}_{l_i(\omega_n)}(\omega_n)\right]. \tag{20}$$

Here, $\mathbf{o}_{l_i(\omega_n)}$ indicates the $1\times l_i(\omega_n)$ zero matrix. Let us assume that $\mathbf{G}(\omega_n)$, which is used for designing the inverse filter $\hat{\mathbf{H}}^{[i]}(\omega_n)$ is a full (row) rank matrix. The sound is reproduced using the inverse filter $\hat{\mathbf{H}}^{[i]}(\omega_n)$, and it is then observed with the monitoring microphone.

[Step 2] The observed signal is normalized using Eq. (16). Also, $p_i(\omega_n)$ is defined as

$$\begin{cases} p_i(\omega_n) = +1 & \text{when} \quad T < P_{\text{diff}}^{[i]}(\omega_{n'}) \\ p_i(\omega_n) = -1 & \text{when} \quad P_{\text{diff}}^{[i]}(\omega_{n'}) < -T \\ p_i(\omega_n) = 0 & \text{otherwise} \end{cases}, \tag{21}$$

where

$$P_{\text{diff}}^{[i]}(\omega_n) = P_{\text{norm}}^{[i]}(\omega_n) - P_{\text{norm}}^{[i-1]}(\omega_n). \tag{22}$$

$P_{\text{diff}}^{[i]}(\omega_n)$ is the difference in the normalized observed signal between iteration time i and $i-1$. Let n' be

$$n - k \le n' \le n + (V - k), \tag{23}$$

where k is assumed to satisfy $0 \le k \le V$. Also, V expresses the signal length needed to prevent the sharp transition of $p_i(\omega_n)$, and T denotes the threshold for the quantization of $p_i(\omega_n)$. Here, T must be tuned moderately because a very large T causes only a small effect of truncation and a very small T aggravates the convergency of the proposed algorithm.

Using the result of $p_i(\omega_n)$ and the variation magnitude of the truncation number at iteration time $i-1$, $a_{i-1}(\omega_n)$, $a_i(\omega_n)$ is obtained on the basis of Table 1. Here, the point regarding the decision rule of $a_i(\omega_n)$ is as follows.

1. $(a_{i-1}(\omega_n) = -1,\ \ p_i(\omega_n) = -1)$ To avoid excessive truncation, the truncation number will be remained if $p_i(\omega_n)$ is equal to zero in with decreasing truncation number.
2. $(a_{i-1}(\omega_n) = 0,\ \ p_i(\omega_n) = -1)$ If $P_{\text{diff}}^{[i]}(\omega_n)$ decreases with remaining truncation number, the truncation number will be remained in order to avoid excessive truncation.
3. $(a_{i-1}(\omega_n) = +1,\ \ p_i(\omega_n) = -1)$ If $P_{\text{diff}}^{[i]}(\omega_n)$ decreases with increasing truncation number, the truncation can be regarded as successful. So the truncation number will be increased in the next iteration time.

Table 1 Decision rule of $a_i(\omega_n)$.

	$a_{i-1}(\omega_n)$		
	-1	0	$+1$
$p_i(\omega_n) = -1$	0	0	$+1$
$p_i(\omega_n) = 0$	0	0	-1
$p_i(\omega_n) = +1$	$+1$	$+1$	-1

4. $(a_{i-1}(\omega_n) = -1, \quad p_i(\omega_n) = 0)$ If $P_{\text{diff}}^{[i]}(\omega_n)$ does not change with decreasing truncation number, the truncation can be regarded as convergent. So the truncation number will be remained in the next iteration time.
5. $(a_{i-1}(\omega_n) = 0, \quad p_i(\omega_n) = 0)$ If $p_i(\omega_n)$ is equal to zero with remaining the truncation number, the truncation number will be remained because the truncation can be regarded as an propriety.
6. $(a_{i-1}(\omega_n) = +1, \quad p_i(\omega_n) = 0)$ To avoid excessive truncation, the truncation number will be decreased if $p_i(\omega_n)$ is equal to zero in spite of increasing truncation number.
7. $(a_{i-1}(\omega_n) = -1, \quad p_i(\omega_n) = +1)$ If $P_{\text{diff}}^{[i]}(\omega_n)$ increases with decreasing truncation number, the truncation number will be increased in the next iteration time because increase of $P_{\text{diff}}^{[i]}(\omega_n)$ is prevented.
8. $(a_{i-1}(\omega_n) = 0, \quad p_i(\omega_n) = +1)$ If $P_{\text{diff}}^{[i]}(\omega_n)$ increases with remaining truncation number, the truncation number will be increased in the next iteration time because increase of $P_{\text{diff}}^{[i]}(\omega_n)$ is prevented.
9. $(a_{i-1}(\omega_n) = +1, \quad p_i(\omega_n) = +1)$ If $P_{\text{diff}}^{[i]}(\omega_n)$ increases with increasing truncation number, the truncation number will be decreased in the next iteration time because the truncation can be regarded as an impropriety.

Table 1 is obtained under the above-mentioned assumption and preliminary experiment. And this table is tuned in order to achieve sufficient effect with few truncation number.

Therefore, we can obtain $l_{i+1}(\omega_n)$ as

$$l_{i+1}(\omega_n) = l_i(\omega_n) + a_i(\omega_n). \tag{24}$$

To prevent sharp truncation, which is often the cause of the degradation of the reproduced sound, a projection on ranks within a short signal length W is smoothed. The parameters V and W also must be tuned to small values, which do not cause sharp truncation.

[Step 3] The process returns to **Step 1** and the original sound is reproduced.

4. Numerical Evaluation

4.1 Experimental Setup

To investigate the effect of the proposed method, a numerical simulation was carried out with real environmental data. The sound reproduction system used in this experiment consisted of 16 secondary sound sources and 7 control points (6 for listening and 1 for monitoring) within a room with a reverberation time of 0.15 seconds. Figure 4 shows a plan view of the secondary sources and the control points. The

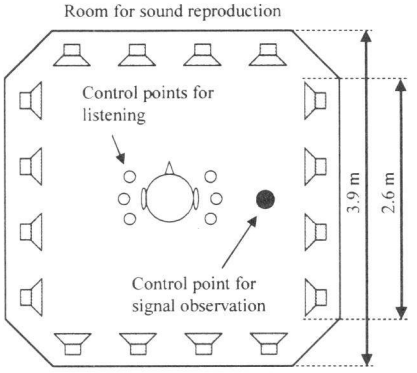

Fig. 4 Plan view of room with sound reproduction system used in numerical simulation.

Table 2 Measurement conditions of room impulse responses.

TSP signal length	65536 points
sampling frequency	48000 Hz
quantization	16 bits
addition for averaging	4 times

Table 3 Experimental condition.

FFT point length (inverse filter length)	32768 points
control bandwidth	150–4000 Hz
L in Eq. (16)	81 points
T in Eq. (21)	2 dB
V in Eq. (23)	50 points
W shown in Sect. 3.4	50 points
original signal	speech, piano

control points for listening were placed at both ears of the head and torso simulator (HATS), and 0.05 m in front and to the rear. The control point for signal observation is placed 0.3 m away from the HATS.

The impulse response used in the present experiment is measured with a time stretched pulse signal (TSP signal) [11], [12]. The measurement conditions are listed in Table 2.

In this simulation, the temperature variations in the room are considered as the environmental change of the transfer system. The impulse responses measured at a room temperature of 18.0°C are used for the inverse filter design, and those measured at a room temperature of 28.0°C are used as the room impulse responses after variations of the transfer system. The inverse filters with 32768 points are designed on the basis of the room impulse responses with 9600 points. With these points, the effect of circular convolution in the inverse filter can be neglected.

The experimental conditions are listed in Table 3, in which L, T, V and W are obtained empirically.

To compare with the conventional method, LNS-based inverse filters are also designed using Eq. (9), where the regularization method is not carried out at all. The inverse filter length and the control bandwidth are the same as those shown in Table 3.

4.2 Results

Figures 5–8 show the normalized power spectrum of the observed signal and the number of ranks of the inverse matrix

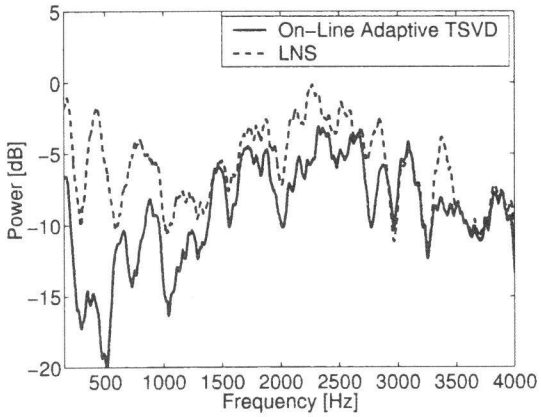

Fig. 5 Normalized power spectrum of normalized observed signal with one-minute adaptation (speech).

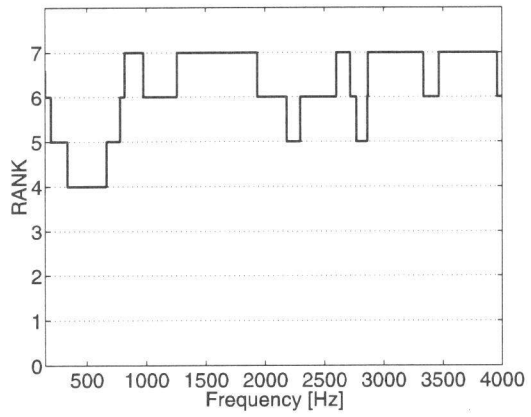

Fig. 6 Number of ranks with one-minute adaptation (speech).

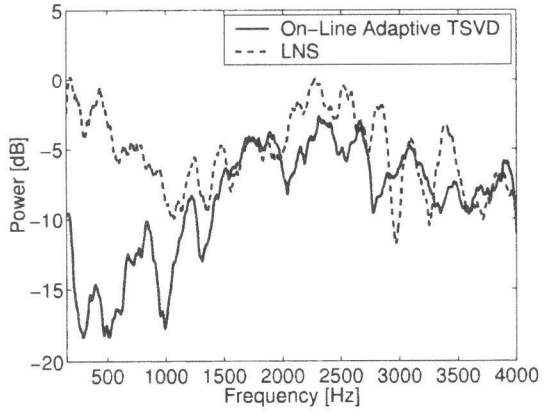

Fig. 7 Normalized power spectrum of normalized observed signal with one-minute adaptation (piano).

before and after 1-minute adaptation. Considering these results, it is apparent that because the shapes of the spectra are almost alike, this algorithm does not depend on the type of original sound, as confirmed by comparing the speech case and the piano case.

In the proposed method, the silent signal is reproduced at the control point for signal observation. So, the normalized observed signal indicates the leakage of the reproduced sounds which reproduced on the control points for listening. Therefore, the reduction of the normalized power spectrum means that the proposed method is effective. From the results shown in Figs. 5 and 7, the relaxation of the inverse filters is particularly effective at low frequencies. This is due to the numerical instability of the matrix $\mathbf{G}(\omega_n)$. Hence, *the condition number* relevant to the instability of $\mathbf{G}(\omega_n)$ is calculated. Figure 9 shows the condition number of $\mathbf{G}(\omega_n)$ at every ω_n, which can be computed by the following equation:

$$\mathrm{cond}\big(\mathbf{G}(\omega_n)\big) = \frac{\mu_{\max}(\omega_n)}{\mu_{\min}(\omega_n)}, \tag{25}$$

where $\mu_{\max}(\omega_n)$ is the maximum singular value and $\mu_{\min}(\omega_n)$ is the minimum singular value. Because the condition number is large at low frequencies, the inverse filters are instable

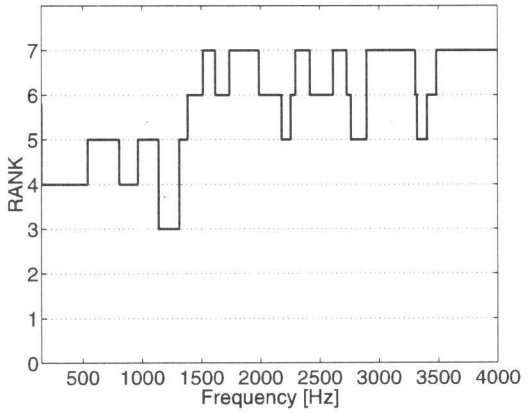

Fig. 8 Number of ranks with one-minute adaptation (piano).

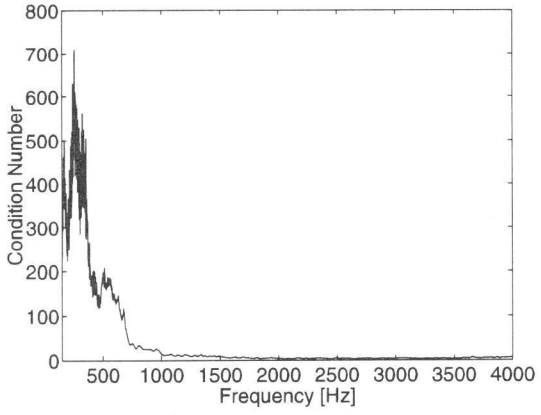

Fig. 9 Condition number of room transfer functions.

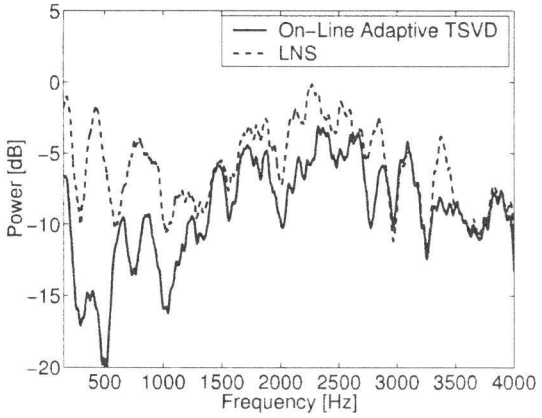

Fig. 10 Normalized power spectrum of the normalized observed signal with five-minute adaptation (speech).

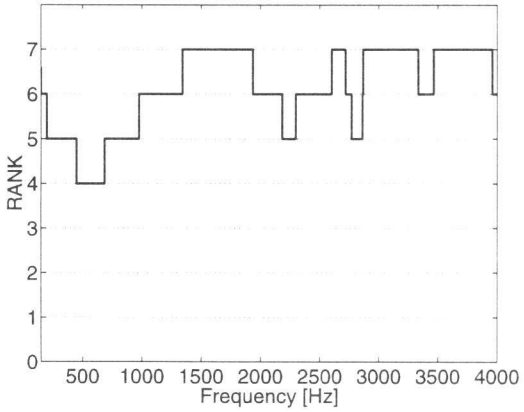

Fig. 11 Number of ranks with five-minute adaptation (speech).

at these frequencies. The reason for this is that the values of the RTF are very similar because the wavelength of the reproduced sound is longer than the intervals of the control points.

To investigate the convergency of the proposed method, Figs. 10 and 11 show the results of 5-minute adaptation, where the original signal is a speech signal. By comparing these with the results of the 1-minute adaptations shown in Figs. 5 and 6, it is found that the algorithm can converge into a constant value, and also it implies that the convergence was complete after 1-minute adaptation.

5. Subjective Evaluation

In the previous section, we demonstrated the efficiency of OATSVD using numerical simulations. It is found that the LNS method can reproduce the sound with sufficient sound localization accuracy [8], [13]. On the other hand, in the adaptive TSVD method, there is a trade-off relationship between the improvement in sound quality and the sound localization accuracy when the RTFs are fluctuated [7]. In this section, we describe the results of a subjective evaluation test for the comparison of the conventional LNS method and

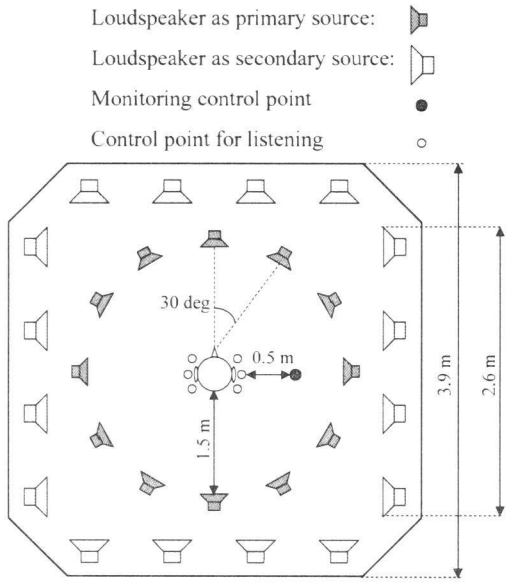

Fig. 12 Sound reproduction system for subjective evaluation.

Table 4 Types of presentation sound.

Original:	original signal directly from the primary sources
LNS:	sound reproduced by the inverse filters based on LNS
Low:	sound reproduced by OATSVD at the applied band of 150–1000 Hz
High:	sound reproduced by OATSVD at the applied band of 1000–4000 Hz
Full:	sound reproduced by OATSVD at the applied band of 150–4000 Hz

the proposed method in terms of sound quality and sound localization accuracy in an actual sound reproduction system.

5.1 Experimental Setup

The configuration of the sound reproduction system is shown in Fig. 12. Six control points for listening are placed, one at each of the two ears and the others 0.05 m in front and behind them, to allow effective sound reproduction even during rotation of the head [13], [14]. The monitoring microphone is placed 0.3 m from the listener. These setup conditions are the same as those described in the previous section. The 12 loudspeakers as primary sound sources are placed at intervals of 30 degrees on a circle with a radius of 1.5 m, with the frontal direction of the listener taken as 0 degrees. The conditions of the environmental variations are identical to those described in the previous section. The speech sound and the piano sound are used as the original sound source signals. The adaptation time for the relaxation of inverse filters by the proposed method is one minute, using the normalized observed signal averaged every four frames, where the frame length is 32768 points (by approximately 0.7 sec.). The types of presentation sound are listed in Table 4. For discussion of the effectiveness of the applied frequency band, inverse filters relaxed on the basis of

three different passbands are prepared in the inverse filters adapted to the proposed method.

The listeners are 10 males and females with normal hearing capabilities. For each presentation sound in Table 4, the speech sound and the piano sound are prepared for each direction and are randomly rearranged. The evaluation of the sound quality is judged using a 5-point opinion scale (5: very good, 4: good, 3: fair, 2: poor, 1: very poor). With regard to the evaluation of sound localization accuracy, any one of the 12 directions is judged. First, the listener turns his or her head toward the direction of the arrival of the presentation sound. Then, the perceived direction is evaluated.

5.2 Results

The sound quality evaluation results are shown in Figs. 13 and 14. The error bars in the figure indicate the 95% confidence interval. When an analysis of variance is carried out for these results, there is a significant difference with a significance level of 5%. However, there is no significant difference between **LNS** and **High**, or between **Low** and **Full** in the speech sound case. Also, there is no significant difference between **LNS** and **High**, or between **High** and **Low** in the piano sound case. This reveals that the sound quality is improved when we apply the proposed method to the low band in both cases.

Figures 15 and 16 show the percentages of correct answers for the perceptual directions. When an analysis of variance is carried out for these results, there is no significant difference with a significance level of 5% between **LNS** and the proposed methods. The results show that the sound localization accuracy of the sound reproduced by the proposed method is similar to that of the sound reproduced by the conventional LNS method. The 95% confidence interval of the sound localization accuracy is higher than that of the sound quality. Moreover, the piano sound case has percentages of correct answers lower than those in the speech sound case. This probably indicates the variation in human capability for sound localization according to the direction of sound and the frequency band.

On the basis of these results, we expect to improve the sound quality of the reproduced sound without degradation of the sound localization.

We listened to the reproduced sounds. In the case of **LNS**, the sound quality was degraded by the amplification of background noise in the low-frequency band. Moreover, particularly in the speech sound case, there was echoic noise prior to the actual reproduction in the case of **LNS**, whereas in the case of the proposed method, these noise and echo

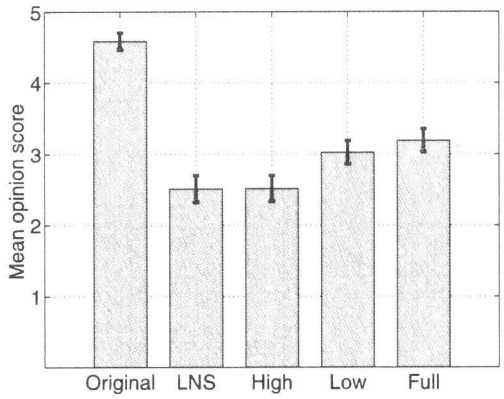

Fig. 13 Result of subjective evaluation for sound quality (speech).

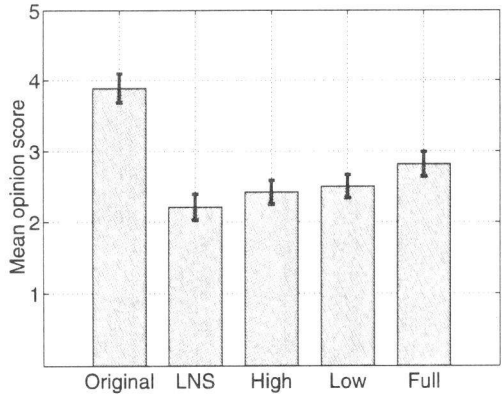

Fig. 14 Result of subjective evaluation for sound quality (piano).

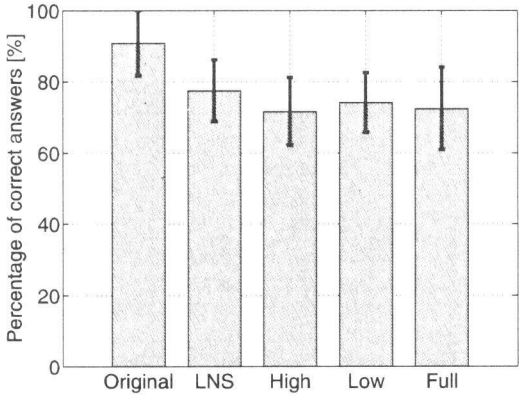

Fig. 15 Percentages of correct answers for perceptual direction (speech).

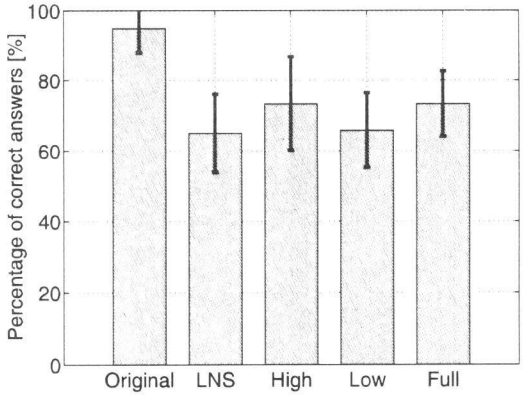

Fig. 16 Percentages of correct answers for perceptual direction (piano).

sensations were markedly reduced.

6. Conclusion

We presented an on-line algorithm for the adaptive relaxation of inverse filters by normalizing the observed signal with the original signals, called OATSVD. In the simulation using real environmental data, it was found that the proposed method can always carry out adaptation processing to acoustic fluctuation. According to the subjective evaluation of sound quality and sound localization of the reproduced sound, the proposed method can improve the reproduced sound quality while maintaining sound localization. Summarizing these results, the proposed method, OATSVD, is applicable for achieving a sound reproduction system that is robust against variations in the listening environment.

In this method, we assume that the causes of environmental variation are temperature fluctuation or object movement. However, in this paper, we discuss only the situation of temperature fluctuation. Therefore, we must further investigate the performance of the proposed method assuming various other environmental fluctuations such as object movement. On the other hand, because ill-condition of the RTFs concentrates at low frequencies, OATSVD is effective for low frequencies in particular. On the basis of this assumption, in future studies, we intend to combine this method and the method for compensation of temperature fluctuation introduced [4], to achieve a sound reproduction system which is more robust against environmental fluctuation.

References

[1] J. Bauck and D.H. Cooper, "Generalized transaural stereo and applications," J. Audio Eng. Soc., vol.44, no.9, pp.683–705, 1996.

[2] S.J. Elliott, I.M. Stothers, and P.A. Nelson, "A multiple error LMS algorithm and its application to the active control of sound and vibration," IEEE Trans. Acoust. Speech Signal Process., vol.35, no.10, pp.1423–1434, 1987.

[3] P.A. Nelson, H. Hamada, and S.J. Elliott, "Adaptive inverse filters for stereophonic sound reproduction," IEEE Trans. Signal Process., vol.40, no.7, pp.1621–1632, 1992.

[4] Y. Tatekura, H. Saruwatari, and K. Shikano, "Sound reproduction system including adaptive compensation of temperature fluctuation effect for broad-band sound control," IEICE Trans. Fundamentals, vol.E85-A, no.8, pp.1851–1860, 2002.

[5] A.N. Tihonov, "Solution of incorrectly formulated problems and the regularization method," Soviet Math., vol.4, pp.1035–1038, 1963.

[6] P.C. Hansen, "Computation of the singular value expansion," Computing, vol.40, pp.185–199, 1988.

[7] Y. Nagata, Y. Tatekura, H. Saruwatari, and K. Shikano, "Iterative inverse filter relaxation algorithm for adaptation to acoustic fluctuation in sound reproduction system," Electron. Commun. Jpn., part 3, vol.87, no.7, pp.15–26, 2004.

[8] A. Kaminuma, S. Ise, and K. Shikano, "A method of designing inverse system for multi-channel sound reproduction system using least-norm-solution," Proc. Active99, vol.2, pp.863–874, Florida, USA, Dec. 1999.

[9] Y. Nagata, Y. Tatekura, H. Saruwatari, and K. Shikano, "Adaptive relaxation algorithm to acoustic fluctuation for inverse filter of sound reproduction system," IEICE Technical Report, EA2001-42, 2001.

[10] H. Tokuno, O. Kirkeby, P.A. Nelson, and H. Hamada, "Inverse filter of sound reproduction system using regularization," IEICE Trans. Fundamentals, vol.E80-A, no.5, pp.809–820, 1997.

[11] N. Aoshima, "Computer-generated pulse signal applied for sound measurement," J. Acoust. Soc. Am., vol.69, no.5, pp.1484–1488, 1981.

[12] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," J. Acoust. Soc. Am., vol.97, no.2, pp.1119–1123, 1995.

[13] A. Kaminuma, S. Ise, and K. Shikano, "Robust sound-reproduction-system design against the head movement," Proc. WESTPRAC VII, vol.1, pp.489–492, Kumamoto, Japan, Oct. 2000.

[14] K. Abe, F. Asano, Y. Suzuki, and T. Sone, "Sound field reproduction by controlling the transfer functions from the source to multiple points in close proximity," IEICE Trans. Fundamentals, vol.E80-A, no.3, pp.574–581, March 1997.

Yosuke Tatekura was born in Kyoto, Japan, on May 17, 1975. He received the B.E. degrees in precision engineering from Osaka University in 1998, and received the M.E. and Ph.D. degrees in information science from Nara Institute of Science and Technology (NAIST) in 2000 and 2002, respectively. He is currently a research associate of Shizuoka University. His research interests include sound field control and virtual sound source synthesis. He is a member of the Acoustical Society of Japan, and the VR Society of Japan.

Shigefumi Urata was born in Nara, Japan on 1979. He received the B.E. degrees in electronic engineering from Osaka City University in 2001 and received the M.E. degrees in information science from Nara Institute of Science and Technology (NAIST) in 2003. He is a member of the Acoustical Society of Japan.

Hiroshi Saruwatari was born in Nagoya, Japan, on July 27, 1967. He received the B.E., M.E. and Ph.D. degrees in electrical engineering from Nagoya University, Nagoya, Japan, in 1991, 1993 and 2000, respectively. He joined Intelligent Systems Laboratory, SECOM CO.,LTD., Mitaka, Tokyo, Japan, in 1993, where he engaged in the research and development on the ultrasonic array system for the acoustic imaging. He is currently an associate professor of Nara Institute of Science and Technology (NAIST). His research interests include array signal processing, blind source separation, and sound field reproduction. He received the Paper Award from IEICE in 2000, and TAF in 2004. He is a member of the IEEE, the VR Society of Japan, and the Acoustical Society of Japan (ASJ).

Kiyohiro Shikano received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nagoya University in 1970, 1972, and 1980, respectively. From 1972, he had been working at NTT Laboratories. During 1990–1993, he was the executive research scientist at NTT Human Interface Laboratories. During 1986–1990, he was the Head of Speech Processing Department at ATR Interpreting Telephony Research Laboratories. During 1984–1986, he was a visiting scientist in Carnegie Mellon University, where he was working on distance measures, speaker adaptation, and statistical language modeling. He received the Yonezawa Prize from IEICE in 1975, the Signal Processing Society 1990 Senior Award from IEEE in 1991, the Technical Development Award from ASJ in 1994, IPSJ Yamashita SIG Research Award in 2000, and Paper Award from the Virtual Society of Japan in 2001. He is currently a professor of Nara Institute of Science and Technology (NAIST), where he is directing speech and acoustics laboratory. He is a member of Information Processing Society of Japan, the Acoustical Society of Japan (ASJ), Japan VR Society, the Institute of Electrical and Electronics, Engineers (IEEE), and International Speech Communication Society.