Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 34013, 12 pages
doi:10.1155/2007/34013

Research Article

Inverse Filtering for Speech Dereverberation Less Sensitive to Noise and Room Transfer Function Fluctuations

Takafumi Hikichi, Marc Delcroix, and Masato Miyoshi

Media Information Laboratory, NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

Received 16 November 2006; Accepted 2 February 2007

Recommended by Liang-Gee Chen

Inverse filtering of room transfer functions (RTFs) is considered an attractive approach for speech dereverberation given that the time invariance assumption of the used RTFs holds. However, in a realistic environment, this assumption is not necessarily guaranteed, and the performance is degraded because the RTFs fluctuate over time and the inverse filter fails to remove the effect of the RTFs. The inverse filter may amplify a small fluctuation in the RTFs and may cause large distortions in the filter's output. Moreover, when interference noise is present at the microphones, the filter may also amplify the noise. This paper proposes a design strategy for the inverse filter that is less sensitive to such disturbances. We consider that reducing the filter energy is the key to making the filter less sensitive to the disturbances. Using this idea as a basis, we focus on the influence of three design parameters on the filter energy and the performance, namely, the regularization parameter, modeling delay, and filter length. By adjusting these three design parameters, we confirm that the performance can be improved in the presence of RTF fluctuations and interference noise.

1. INTRODUCTION

Inverse filtering of room acoustics is useful in various applications such as sound reproduction, sound-field equalization, and speech dereverberation. Usually, room transfer functions (RTFs) are modeled as finite impulse response (FIR) filters, and inverse filters are designed to remove the effect of the RTFs. When the RTFs are known *a priori* or are capable of being accurately estimated, this approach has been shown to achieve high inverse filtering performance [1–4]. However, in actual acoustic environments, there are disturbances that affect the inverse filtering performance. One cause of these disturbances is the fluctuation in the RTFs resulting from changes in such factors as source position and temperature [5–9]. As a result, an inverse filter correctly designed for one condition may not work well for another condition, and compensation or adaptation processing may become necessary.

The sensitivity issue with inverse filtering in relation to the movement of a sound source or microphone has been addressed in several papers. In [8, 9], the sensitivity of inverse filters is quantified in terms of the mean-squared error (MSE), defined as the power of the deviation of the equalized impulse response from the ideal impulse. This MSE is theoretically derived based on statistical room acoustics. These studies claim that the region in which the MSE is below -10 dB is restricted to a few tenths of a wavelength of a target signal, revealing a high sensitivity to small positional changes. That is, when an inverse filter designed for a certain location is applied to recover signals observed at another location, the performance easily degrades and the MSE becomes high.

Inverse filters are usually obtained by inverting the autocorrelation matrix of the RTFs. Accordingly, in order to realize stable inverse filtering, either regularization [10] or the truncated singular value decomposition method [11–13] has been applied. With the latter method, the small singular values of the autocorrelation matrix of the RTFs are treated as zeros. Both methods have been applied to a sound reproduction system, and have been experimentally verified.

The purpose of this paper is to pursue ways of designing inverse filters that are less sensitive to RTF fluctuations and interference noise. When the RTFs fluctuate, the inverse filter may amplify the small fluctuation in the RTFs and may cause large distortions in the output signal of the inverse filter. Moreover, when the microphone signal contains noise,

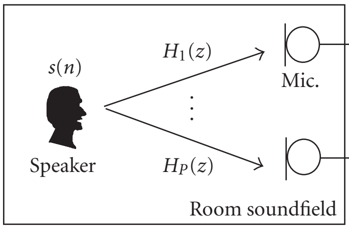


FIGURE 1: Single-source multimicrophone acoustic system. $H_i(z)$ represent room transfer functions.

the inverse filter may also amplify the noise. We expect the filtered signal to be less degraded when the filter energy is small. Hence, we believe that reducing the filter energy is the key to making the filters less sensitive. To confirm this belief, we focus on the influence of three parameters used in the design of inverse filters: the regularization parameter, filter length, and modeling delay. By selecting proper parameter values, we expect to reduce the filter energy, and hence make the filter more robust to RTF variations and noise.

The organization of this paper is as follows. The following section describes the acoustic system with a single source and multiple microphones considered in this paper. It then describes how inverse filters are calculated and analyzes the effect of the three design parameters on the filter energy. Section 3 reports experiments undertaken in the presence of noise. Section 4 describes experimental results for an inverse filter with RTF fluctuations caused by source position changes. Section 5 provides an analysis of the RTF fluctuations caused by source position changes. Section 6 concludes the paper.

2. PROBLEM FORMULATION

2.1. Acoustic system in consideration

We consider an acoustic system with a single sound source and multiple microphones as shown in Figure 1. The source signal is represented as $s(n)$, where n denotes a discrete time index, and the signals received by the microphones are $x_i(n)$, $i = 1, \dots, P$, where P is the number of microphones. Microphone signals $x_i(n)$ are given by

$$x_i(n) = h_i(n) * s(n) + w_i(n) \tag{1}$$

$$= \sum_{k=0}^{J} h_i(k)s(n-k) + w_i(n), \quad i = 1, \dots, P, \tag{2}$$

where $*$ denotes the convolution operation, $h_i(k)$, $k = 0, \dots, J$, denotes the room impulse response between the source and the ith microphone, and $w_i(n)$ denotes noise. The RTFs are expressed as

$$H_i(z) = \sum_{k=0}^{J} h_i(k)z^{-k}, \quad i = 1, \dots, P. \tag{3}$$

We assume hereafter that these RTFs have no common zeros among all the channels.

Equation (2) can be expressed in a matrix form as

$$\mathbf{x}(n) = \mathbf{H}^T\mathbf{s}(n) + \mathbf{w}(n), \tag{4}$$

where

$$\mathbf{x}(n) = \begin{bmatrix} \mathbf{x}_1(n) \\ \vdots \\ \mathbf{x}_P(n) \end{bmatrix}, \quad \mathbf{x}_i(n) = \begin{bmatrix} x_i(n) \\ x_i(n-1) \\ \vdots \\ x_i(n-M+1) \end{bmatrix}, \quad i = 1, \dots, P,$$

$$\mathbf{w}(n) = \begin{bmatrix} \mathbf{w}_1(n) \\ \vdots \\ \mathbf{w}_P(n) \end{bmatrix}, \quad \mathbf{w}_i(n) = \begin{bmatrix} w_i(n) \\ w_i(n-1) \\ \vdots \\ w_i(n-M+1) \end{bmatrix}, \quad i = 1, \dots, P,$$

$$\mathbf{s}(n) = \begin{bmatrix} s(n) \\ s(n-1) \\ \vdots \\ s(n-J-M+1) \end{bmatrix},$$

$$\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_P],$$

$$\mathbf{H}_i = \underbrace{\left.\begin{pmatrix} h_i(0) & 0 & \dots & 0 \\ h_i(1) & h_i(0) & \ddots & \vdots \\ \vdots & h_i(1) & \ddots & 0 \\ h_i(J) & \vdots & \ddots & h_i(0) \\ 0 & h_i(J) & & h_i(1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_i(J) \end{pmatrix}\right\}}_{M}(J+M), \tag{5}$$

and M is the block size of the microphone signals for each channel. The objective of dereverberation is to recover source signal $s(n)$ from the received signal $\mathbf{x}(n)$. This is achieved by filtering the received signal with the inverse filter of room acoustic system $\mathbf{H}$.

2.2. Inverse filter calculation

Generally, the inverse filter vector, denoted as $\mathbf{g}$, is calculated by minimizing the following cost function:

$$C = \|\mathbf{Hg} - \mathbf{v}\|^2, \tag{6}$$

where $\|\mathbf{a}\|$ denotes the l_2-norm of vector $\mathbf{a}$, where

$$\mathbf{g} = [\underbrace{g_1(1), \dots, g_1(M), \dots, g_P(1), \dots, g_P(M)}_{PM}]^T,$$
$$\mathbf{v} = [\underbrace{0, \dots, 0}_{d}, 1, 0, \dots, 0]^T, \tag{7}$$

M is the filter length for each channel, and d $(0 \le d \le PM)$ is the modeling delay [14]. Here, modeling delay can be selected arbitrarily. By applying this inverse filter $\mathbf{g}$ to the microphone signals, the filter's output signal is equivalent to the input signal delayed by d-taps. Hereafter, we consider that impulse responses $h_i(n)$ are normalized by their norm. When RTF matrix $\mathbf{H}$ is given, such inverse filter set can be calculated as

$$\mathbf{g} = \mathbf{H}^{+}\mathbf{v}, \tag{8}$$

where $\mathbf{A}^{+}$ is the Moore-Penrose pseudoinverse of matrix $\mathbf{A}$ [15]. The inverse filter set is calculated based on the multiple-input/output inverse theorem (MINT) [1]. The filter set with minimum length is obtained by setting M so that matrix $\mathbf{H}$ is square, which leads to $M = M_{\min} = J/(P-1)$. The filter length can be set at $M > J/(P-1)$ as well.

2.3. Inverse filters with disturbances

When noise is present at the microphones, distortion occurs in the output signal of the inverse filter. The larger the filter energy is, the larger the distortion can be. Thus, we introduce the filter energy into the cost function expressed in (6). By taking the filter energy into consideration, the cost function is modified as follows:

$$C = \|\mathbf{Hg} - \mathbf{v}\|^2 + \delta\|\mathbf{g}\|^2, \tag{9}$$

where $\delta(\geq 0)$ is a scalar variable. This parameter determines how much weight to assign to the energy term, and thus determines a tradeoff between the filter's accuracy and the amount of distortion. The same formulation is applied as the one used in multichannel active noise control systems [14, 16]. We would like to derive a solution that minimizes this cost function. Equation (9) can be rewritten as

$$\begin{aligned} C &= (\mathbf{Hg} - \mathbf{v})^T(\mathbf{Hg} - \mathbf{v}) + \delta\mathbf{g}^T\mathbf{g} \\ &= \mathbf{g}^T\mathbf{H}^T\mathbf{Hg} - \mathbf{g}^T\mathbf{H}^T\mathbf{v} - \mathbf{v}^T\mathbf{Hg} + \mathbf{v}^T\mathbf{v} + \delta\mathbf{g}^T\mathbf{g}. \end{aligned} \tag{10}$$

By taking derivatives with respect to $\mathbf{g}$ and setting them equal to zero, the following solution is derived:

$$\mathbf{g}_r = \left(\mathbf{H}^T\mathbf{H} + \delta\mathbf{I}\right)^{-1}\mathbf{H}^T\mathbf{v}, \tag{11}$$

where $\mathbf{I}$ is an identity matrix. This solution has a similar form to that of Tikhonov regularization for ill-posed problems [11–13, 17]. We hereafter refer to δ as a regularization parameter, and $\mathbf{g}_r$ as an inverse filter vector with regularization.

Equation (11) is an optimum solution when the interference noise is white noise with small variance δ, and the term $\delta\mathbf{I}$ corresponds to the correlation matrix of the noise. If the colored noise is considered as a more general case, its correlation matrix is replaced with term $\delta\mathbf{I}$ as

$$\mathbf{g}_r = \left(\mathbf{H}^T\mathbf{H} + \mathbf{R_n}\right)^{-1}\mathbf{H}^T\mathbf{v}, \tag{12}$$

where $\mathbf{R_n}$ is the noise correlation matrix.

Then, let us consider the situation where RTFs fluctuate. Suppose fluctuated RTFs denoted as $\overline{\mathbf{H}} + \widetilde{\mathbf{H}}$, where $\overline{\mathbf{H}}$ and $\widetilde{\mathbf{H}}$ represent the mean RTF and the fluctuation from the mean RTF, respectively. In this case, we consider the ensemble mean of the total squared error,

$$\begin{aligned} C &= E\langle \|(\overline{\mathbf{H}} + \widetilde{\mathbf{H}})\mathbf{g} - \mathbf{v}\|^2 \rangle \\ &= E\langle (\overline{\mathbf{H}}\mathbf{g} - \mathbf{v} + \widetilde{\mathbf{H}}\mathbf{g})^T(\overline{\mathbf{H}}\mathbf{g} - \mathbf{v} + \widetilde{\mathbf{H}}\mathbf{g})\rangle \\ &= (\overline{\mathbf{H}}\mathbf{g} - \mathbf{v})^T(\overline{\mathbf{H}}\mathbf{g} - \mathbf{v}) \\ &\quad + E\langle (\overline{\mathbf{H}}\mathbf{g} - \mathbf{v})^T\widetilde{\mathbf{H}}\mathbf{g} + (\widetilde{\mathbf{H}}\mathbf{g})^T(\overline{\mathbf{H}}\mathbf{g} - \mathbf{v}) + \mathbf{g}^T\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}}\mathbf{g}\rangle \\ &= \mathbf{g}^T\overline{\mathbf{H}}^T\overline{\mathbf{H}}\mathbf{g} - \mathbf{g}^T\overline{\mathbf{H}}^T\mathbf{v} - \mathbf{v}^T\overline{\mathbf{H}}\mathbf{g} + \mathbf{v}^T\mathbf{v} + \mathbf{g}^T E\langle \widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}}\rangle\mathbf{g}, \end{aligned} \tag{13}$$

where $E\langle\cdot\rangle$ represents the expectation operation. In this derivation, we assume $E\langle\widetilde{\mathbf{H}}\rangle$ is a zero matrix. Then, the following filter minimizes the cost function expressed in (13):

$$\mathbf{g}_r = \left(\overline{\mathbf{H}}^T\overline{\mathbf{H}} + \mathbf{R_H}\right)^{-1}\overline{\mathbf{H}}^T\mathbf{v}, \tag{14}$$

where $\mathbf{R_H} = E\langle\widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}}\rangle$. From discussions described above, we can treat the disturbances by using the filter expressed in the following form:

$$\mathbf{g}_r = \left(\mathcal{H}^T\mathcal{H} + \mathcal{R}\right)^{-1}\mathcal{H}^T\mathbf{v}, \tag{15}$$

where $\mathcal{H}$ is either $\mathbf{H}$ or the mean RTF $\overline{\mathbf{H}}$, and $\mathcal{R}$ is the correlation matrix of either the noise $\mathbf{R_n}$ or the fluctuation $\mathbf{R_H}$. If the fluctuation could be regarded as white noise, $\mathcal{R} = \delta\mathbf{I}$ could be applied to the inverse filter. In the following experiments, we investigate the performance of the inverse filter of the form

$$\mathbf{g}_r = \left(\mathcal{H}^T\mathcal{H} + \delta\mathbf{I}\right)^{-1}\mathcal{H}^T\mathbf{v}, \tag{16}$$

where

$$\mathcal{H} = \begin{cases} \mathbf{H} & \text{(noise case)}, \\ \overline{\mathbf{H}} & \text{(fluctuation case)}. \end{cases} \tag{17}$$

2.4. Influence of design parameters on filter energy

Regularization parameter δ increases the minimum eigenvalue of matrix $(\mathcal{H}^T\mathcal{H} + \delta\mathbf{I})$ in (16), and hence reduces the norm of the inverse filter. Increasing the regularization parameter is thus believed to reduce the sensitivity to RTF variations and noise. On the other hand, increasing this parameter reduces the accuracy of the inverse filter with respect to the true RTFs.

The effect of the filter length can be expected as follows. Equation (16) will give the minimum norm filter for a given length M. By increasing the filter length, we compare various filters with different lengths, and consequently expect that the filter with the smallest norm can be found.

A modeling delay d is also used to make the inverse filter stable. When a nonzero modeling delay d $(d \geq 1)$ is used, we also expect the filter norm to be reduced because the causality constraint is relaxed. The filter may correspond to the minimum-norm solution that could be obtained in the frequency domain [18].

As described above, we can expect the regularization parameter, filter length, and modeling delay to be effective in reducing the filter energy.

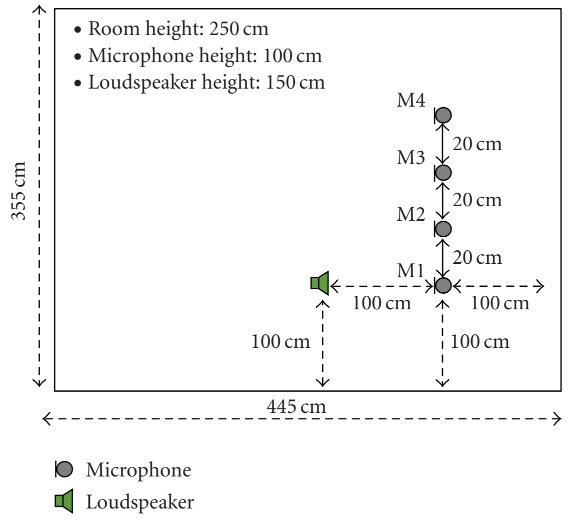


Figure 2: Source and microphone arrangement. M1, M2, M3, and M4 denote the microphones.

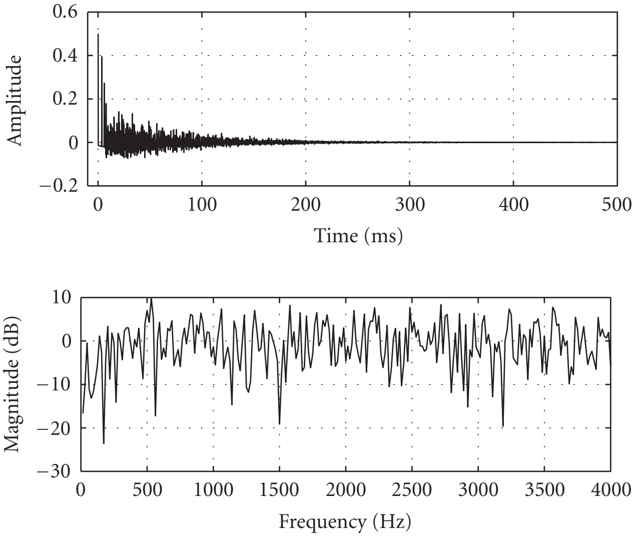


Figure 3: Waveform of a room impulse response $h_1(n)$ and its frequency characteristics.

3. EXPERIMENTS ON THE EFFECT OF NOISE

Experiments were performed to verify the effectiveness of our strategy in the presence of additive white noise.

3.1. Experimental setup

Figure 2 shows the arrangement of the source and the microphones used in the experiment. Four microphones are used ($P = 4$), and room impulse responses between the source and the microphones are simulated by using the image method [19]. The sampling frequency is set at 8 kHz. The impulse responses are truncated to 4000 samples ($J = 3999$), corresponding to −60 dB attenuation (the reverberation time of the room is 500 ms). Figure 3 shows an example of the impulse response and its frequency response.

We define the input and output SNRs as follows. For the ith microphone, the input SNR is defined as

$$\text{SNR}_{\text{in}} = 10\log_{10}\left(\frac{\sum_{n=0}^{N} y_i^2(n)}{\sum_{n=0}^{N} w_i^2(n)}\right), \tag{18}$$

where $y_i(n)$ is the reverberant signal without noise, and $w_i(n)$ is the noise. In the experiment, we adjust the input SNR by controlling the amplitude of the noise signal. The output SNR is defined as

$$\text{SNR}_{\text{out}} = 10\log_{10}\left(\frac{\sum_{n=0}^{N}\left(\mathbf{y}(n)^T\mathbf{g}_r\right)^2}{\sum_{n=0}^{N}\left(\mathbf{w}(n)^T\mathbf{g}_r\right)^2}\right), \tag{19}$$

where $\mathbf{y}(n) = \mathbf{H}^T\mathbf{s}(n)$ is the reverberant signal vector. This output SNR is obtained by filtering the reverberant and the noise signals separately and taking the power ratio of the output signals.

3.2. Evaluation criteria

In order to avoid any dependency of the results on the source signal, we used uncorrelated white signals with a duration of 3 seconds for both source signal and noise rather than speech.

The dereverberation performance is evaluated by using the signal-to-distortion ratio (SDR) defined as

$$\text{SDR} = 10\log_{10}\left(\frac{\sum_{n=0}^{N} s^2(n)}{\sum_{n=0}^{N}\left(s(n)-\hat{s}(n)\right)^2}\right), \tag{20}$$

where $s(n)$ is the original source signal and $\hat{s}(n)$ is the output signal of the inverse filter defined as $\hat{s}(n) = \mathbf{x}(n)^T\mathbf{g}_r$.

3.3. Results

Figure 4 shows the filter energy with various modeling delays and regularization parameters when the minimum filter length $M = M_{\min} = 1333$ is used, as described in Section 2.2. The energy decreases with increases in both the modeling delay and the regularization parameter, and shows the minimum value when $\delta = 10^{-1}$ and $d = 500$.

Figure 5 shows the inverse filter calculated with $\delta = 10^{-6}$ and $\delta = 10^{-1}$ when the modeling delay is fixed at $d = 500$. We clearly observed that the filter energy was reduced by increasing the regularization parameter.

Figure 6 shows the performance of the inverse filter with an input SNR of 20 dB. We observed that a proper regularization parameter value of $\delta = 10^{-2}$ gives the largest SDR for all the modeling delay values. This regularization parameter corresponds to the input SNR (20 dB). When the regularization parameter is smaller than 10^{-2}, the performance monotonically decreased as the regularization parameter decreased, according to the increase in the filter energy. Even though the filter norm decreases with $\delta = 10^{-1}$, the performance also deteriorated because the accuracy of the filter

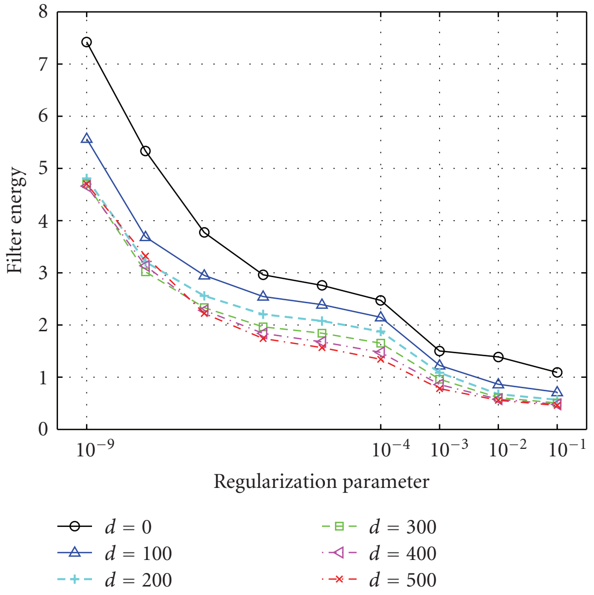
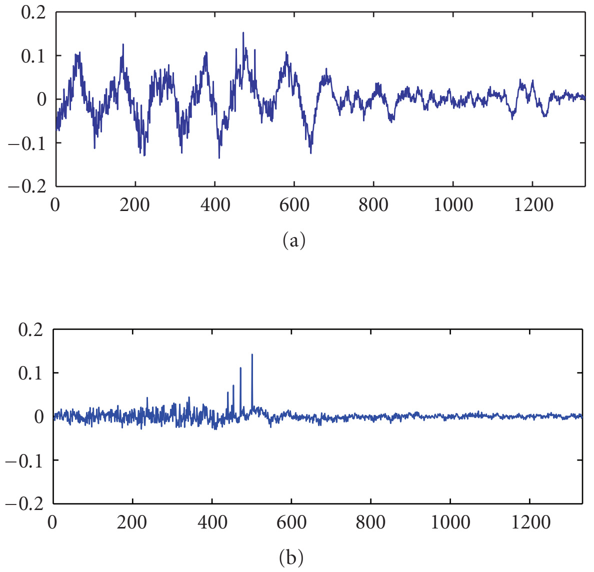


FIGURE 4: Filter energy as a function of regularization parameter and modeling delay (filter length is fixed at $M = 1333$).



(a)

(b)

FIGURE 5: An example of inverse filter $g_1(n)$ calculated with $\delta = 10^{-6}$ (a) and $\delta = 10^{-1}$ (b) (modeling delay is fixed at $d = 500$).

decreased and the deviation of the equalized response from the ideal one became large.

In the second experiment, the modeling delay was fixed at $d = 500$, and the effect of filter length M was investigated with various regularization parameters δ. Figures 7 and 8 show the filter energy and corresponding performance in this case. In Figure 7, the energy decreases with increases in both the filter length and the regularization parameter, although the effect of the filter length is less significant when a large

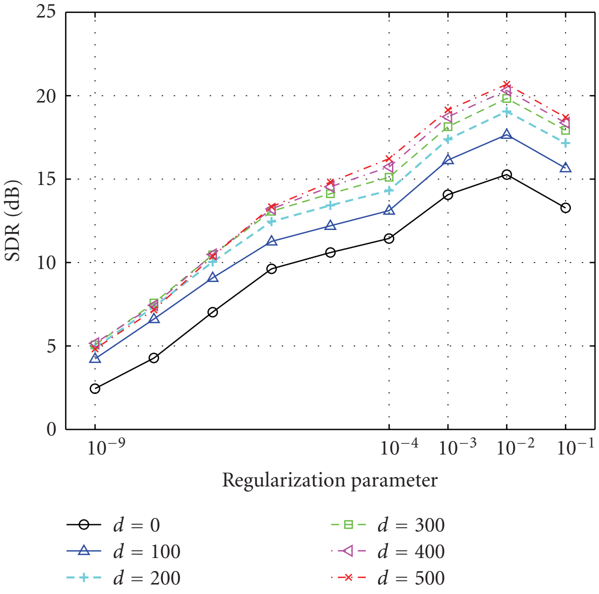


FIGURE 6: Performance as a function of regularization parameter and modeling delay with an SNR of 20 dB (filter length is fixed at $M = 1333$).

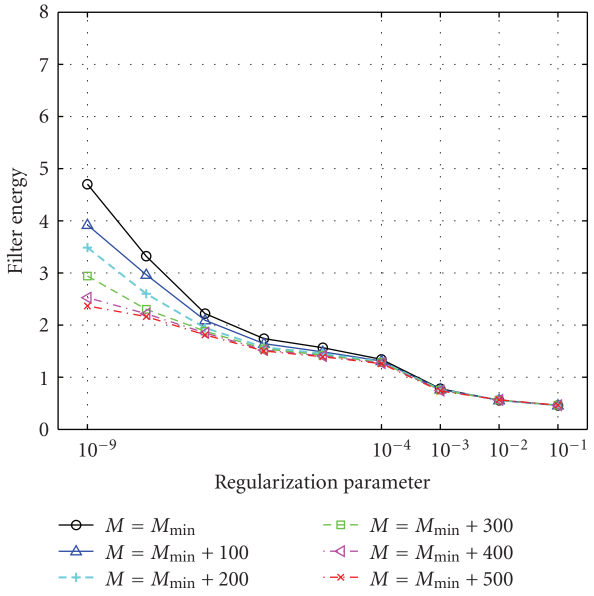


FIGURE 7: Filter energy as a function of regularization parameter and filter length (modeling delay is fixed at $d = 500$).

regularization parameter such as $\delta = 10^{-1}$ to $\delta = 10^{-2}$ is used. In Figure 8, the best performance was obtained with $\delta = 10^{-2}$ for all the filter lengths used in this experiment, which corresponds to the input SNR level. The performance was also improved by using the larger filter length.

In the third experiment, we evaluated the performance for several SNR values by using modeling delay $d = 500$ and filter length $M = 1333$ (minimum case), or $M = 1333 + 500$ (lengthened case). Figure 9 shows the results

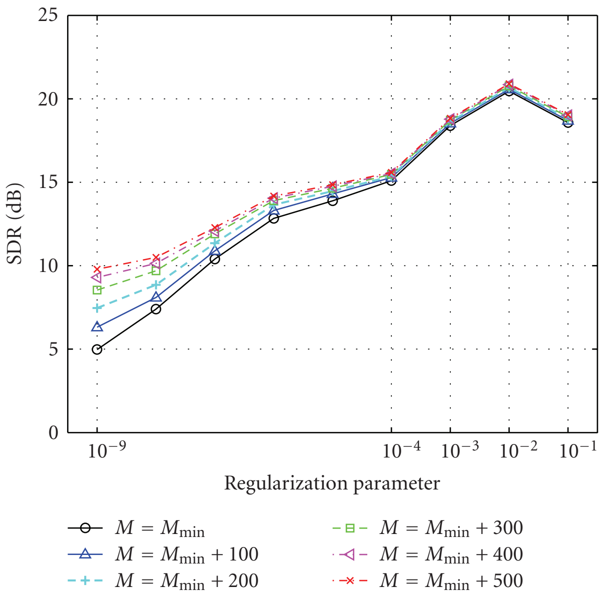


Figure 8: Performance as a function of regularization parameter and filter length with an SNR of 20 dB (modeling delay is fixed at $d = 500$).

obtained with input SNRs of 10, 20, 30, and 40 dB. As the input SNR increases, the regularization parameter that provides the best performance decreases. We observe that the best regularization parameter corresponds to the input SNR. We also observe that the performance evaluated with SDR is bounded by the input SNR level. In addition, when the input SNR is 20 dB, the output SNR defined in (19) is about 20 dB, indicating that the input noise is not amplified.

By using a proper delay and a larger filter length, the inverse filter's energy and equalization error can be reduced. Furthermore, appropriate choice of the regularization parameter is effective for reducing the equalization error. In the next section, we investigate the applicability of this strategy to the RTF fluctuations.

4. EXPERIMENTS FOR RTF FLUCTUATIONS

Simulations are undertaken to investigate the effect of the RTF fluctuations on the inverse filter. Here, we consider the fluctuations caused by source position fluctuations in the horizontal plane for the sake of simplicity. The more general case of three-dimensional fluctuations is not investigated in this paper.

4.1. Experimental setup

We consider the same room as in the previous experiment shown in Figure 2. As for the source positions, we simulate the fluctuations in source position as follows. As shown in Figure 10, we consider N equally spaced new positions placed on a circle of radius r centered at the original position. As a model of fluctuation, we assume that the source is located at each of these N positions with equal probability, and that the averaged RTF over these positions is obtained through either measurement or estimation. This averaged RTF is referred to as "reference RTF," and is used to calculate inverse filters according to (16). In the following simulation, the number of source positons is fixed to $N = 8$.

4.2. Evaluation procedure

The performance of the inverse filter for fluctuations in the source position is evaluated as follows.

(1) An inverse filter set is calculated based on the reference RTFs according to (16).

(2) For each new source position j ($j = 1, \ldots, 8$), equalization is achieved by filtering reverberant signals with the inverse filter set calculated in (1).

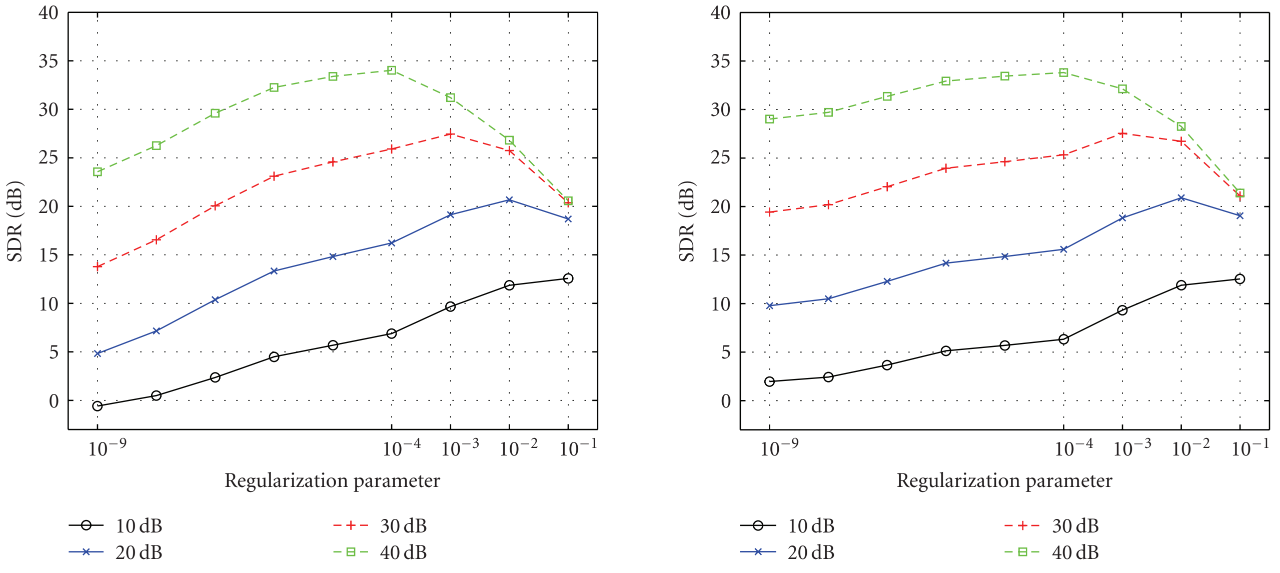
(3) SDR values are calculated for all of the dereverberated signals obtained in (2), and the SDR values are averaged over the 8 positions to obtain the overall performance measure.

4.3. Results

The influence of the design parameters on performance is evaluated in the same manner as in the previous experiment. Figure 11 shows the performance of an inverse filter designed with various modeling delays d and regularization parameters δ with radius $r = 1$ cm. This radius corresponds to one eighth of a wavelength of the center frequency of signals in consideration. Conventional studies have shown considerable degradation in the performance for this displacement. In general, the performance shows a similar tendency to that obtained in the previous experiment. That is, the performance is inversely proportional to the filter energy, and improved with increases in the regularization parameter and modeling delay. We observed that the best performance was obtained at $\delta = 10^{-2}$ and $d = 500$. However, the performance is rather flat compared with that in Figure 6. For a change of source position of $r = 1$ cm, the best performance was 12 dB.

In the second experiment, the modeling delay was fixed at $d = 500$, and the effects of filter length M and regularization parameter δ were investigated. Figure 12 shows the performance in this case. Here also, we observed that the performance is inversely proportional to the filter energy. Furthermore, the performance depends on the regularization parameter less than in the case of additive noise. In the case of additive noise, the noise correlation matrix $\mathbf{R_n}$ in (12) could be well approximated to $\delta\mathbf{I}$. On the contrary, the correlation matrix of the fluctuation $\mathbf{R_H}$ in (14) could not be correctly approximated to $\delta\mathbf{I}$.

Figure 13 shows the performance for position variations of $r = 1, 2, 3$, and 4 cm. The modeling delay was set at $d = 500$, and the filter length was set at $M = 1333$ (minimum case) and $M = 1333 + 500$ (lengthened case). In both cases, when $r = 1$ cm, $\delta = 10^{-2}$ shows the maximum SDR value of around 12 dB. For $r = 2, 3$, and 4 cm, the best regularization parameter was $\delta = 10^{-1}$.



(a)

(b)

Figure 9: Performance as a function of regularization parameter for SNR values of 10, 20, 30, and 40 dB ($d = 500$). Filter length was set at $M = 1333$ (a), and $M = 1333 + 500$ (b), respectively.

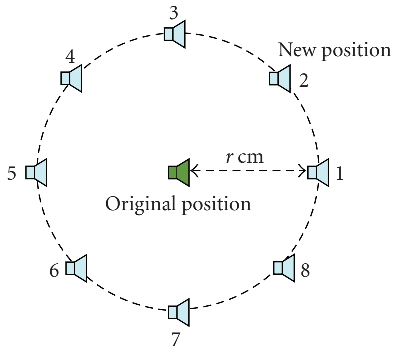


Figure 10: Source positions considered in the experiment.

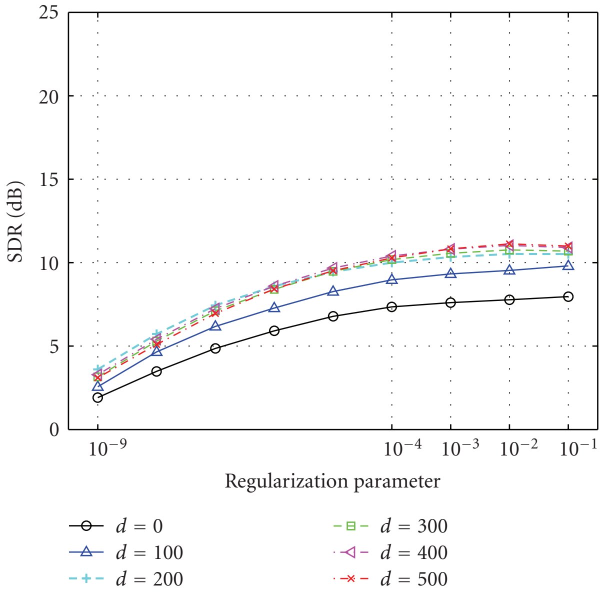


Figure 11: Performance as a function of the regularization parameter and modeling delay (filter length is fixed at $M = 1333$).

Again, by using an appropriate delay and filter length, the inverse filter's energy could be reduced, and accordingly the inverse filtering performance could be improved. Furthermore, an appropriate choice of regularization parameter was effective. However, the effect of adjusting this regularization parameter is less obvious than with additive noise.

In the next section, we analyze the RTF fluctuations caused by position changes, and discuss the differences between the results for RTF fluctuations and additive noise.

5. DISCUSSION

5.1. Comparison between RTF fluctuations and noise

We compare the results for RTF fluctuations shown in Figure 9 and the results for noise shown in Figure 13. As shown in Figure 9, the dereverberation performance has a maximum point for a certain regularization parameter value, and this best value corresponds to the SNR value of the observed signals. For example, with SNR $= 20$ dB, the best value is $\delta = 10^{-2}$ and this gives a maximum SDR of 20 dB, that is, we obtained almost the same SDR level as the input SNR. When a smaller δ is used such as 10^{-9}, the filter energy becomes large, and hence this results in a small SDR of 5 (minimum-length case) to 10 dB (lengthened filter case). By contrast, for RTF fluctuations of $r = 1$ cm (corresponding to one eighth of a wavelength of the center frequency of signals

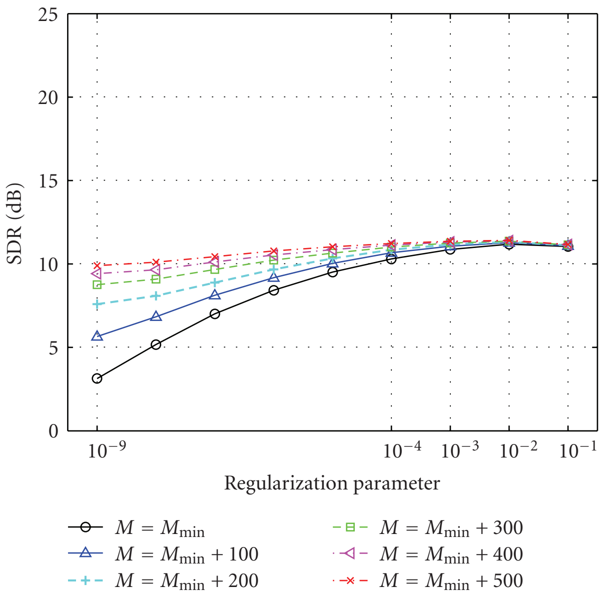


FIGURE 12: Performance as a function of the regularization parameter and additional filter length (modeling delay is fixed at $d = 500$).

in consideration) as shown in Figure 13, although the best value for the regularization parameter is almost the same, that is, $\delta = 10^{-2}$, the corresponding SDR was around 12 dB, and the curve was much broader than in Figure 9. That is, the performance does not depend greatly on δ.

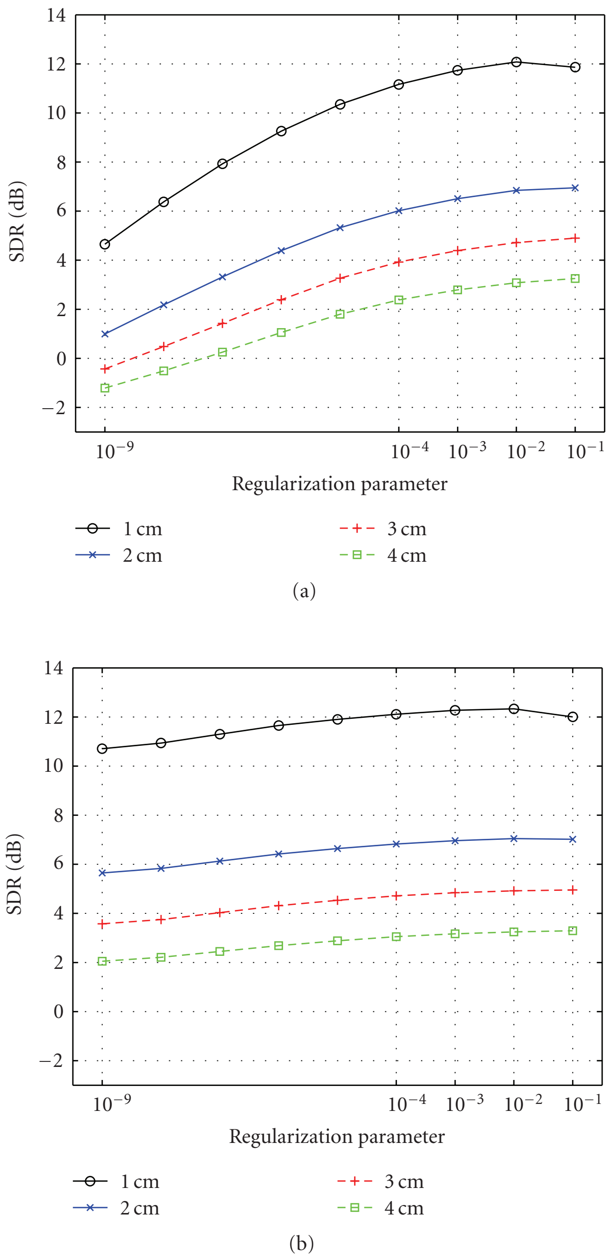
The cause of the difference between these two results is discussed here. We analyze the effect of using this filter in the fluctuation case on the performance using the fluctuation model described in Section 5.1. Let us denote the RTF matrix corresponding to each source position as $\mathbf{H}_j = \overline{\mathbf{H}} + \widetilde{\mathbf{H}}_j$, where $\overline{\mathbf{H}}$ represents the reference RTF matrix averaged over the positions, and $\widetilde{\mathbf{H}}_j$ represents the fluctuation between the reference RTF and the RTF for the jth new postion. If the source switches back and forth among all the possible positions with equal probability, we can consider that the periods in which the source locates at each position are rearranged and put together. Then, the total error may be calculated as the sum of errors for all the positions as

$$C = \frac{1}{N}\sum_{j=1}^{N} ||\mathbf{H}_j\mathbf{g} - \mathbf{v}||^2 = \frac{1}{N}\sum_{j=1}^{N} ||(\overline{\mathbf{H}} + \widetilde{\mathbf{H}}_j)\mathbf{g} - \mathbf{v}||^2. \qquad (21)$$

By considering sufficiency large number of N, we replace spatial averaging with an expectation,

$$\begin{aligned} C &= E\langle ||(\overline{\mathbf{H}} + \widetilde{\mathbf{H}})\mathbf{g} - \mathbf{v}||^2 \rangle \\ &= E\langle (\overline{\mathbf{H}}\mathbf{g} - \mathbf{v} + \widetilde{\mathbf{H}}\mathbf{g})^T (\overline{\mathbf{H}}\mathbf{g} - \mathbf{v} + \widetilde{\mathbf{H}}\mathbf{g}) \rangle. \end{aligned} \qquad (22)$$

This turns out to be (13).



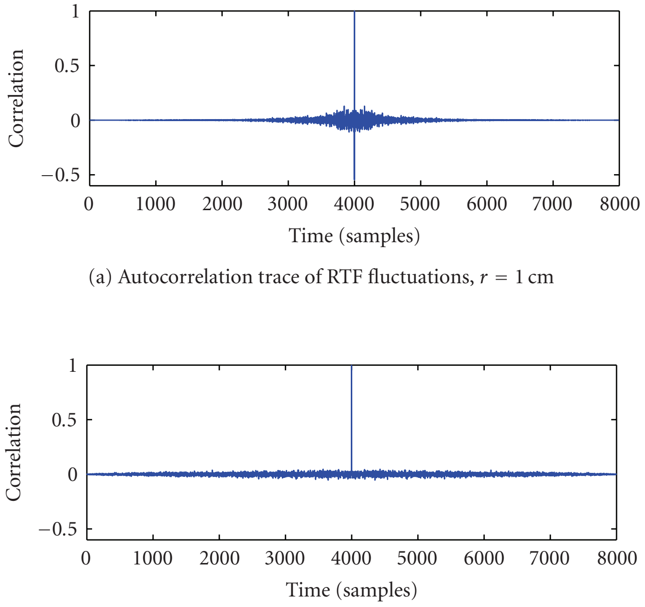
(a)

(b)

FIGURE 13: Performance as a function of the regularization parameter for position variations of $r = 1, 2, 3$ and 4 cm ($d = 500$). Filter length was set at $M = 1333$ (a), and $M = 1333 + 500$ (b), respectively.

Let us evaluate the difference in performance between $E\langle \widetilde{\mathbf{H}}^T \widetilde{\mathbf{H}} \rangle$ and $\delta\mathbf{I}$. First, we compare autocorrelation traces of an example RTF fluctuation and of a random signal used in the experiment. Figure 14 shows these autocorrelations. There is a discrepancy between these two correlations. This may explain why the adjustment of the regularization parameter is of limited efficiency in the presence of RTF fluctuations.

Then, the inverse filter in (15) is used to compare the performance with $\mathcal{H} = \overline{\mathbf{H}}$ and regularization matrices $\mathcal{R}$



(a) Autocorrelation trace of RTF fluctuations, r = 1 cm

(b) Autocorrelation trace of a random signal

Figure 14: Autocorrelation coefficients.

Table 1: Regularization performance.

Regularization matrix $\mathcal{R}$	(1) $\delta\mathbf{I}$, $\delta = 10^{-2}$	(2) $E\langle \widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}}\rangle \approx (1/8)\sum_{j=1}^{8}\widetilde{\mathbf{H}}_j^T\widetilde{\mathbf{H}}_j$
Average SDR (dB)	12.0	15.7

defined as

(1) $\mathcal{R} = \delta\mathbf{I}$, $\delta = 10^{-2}$,

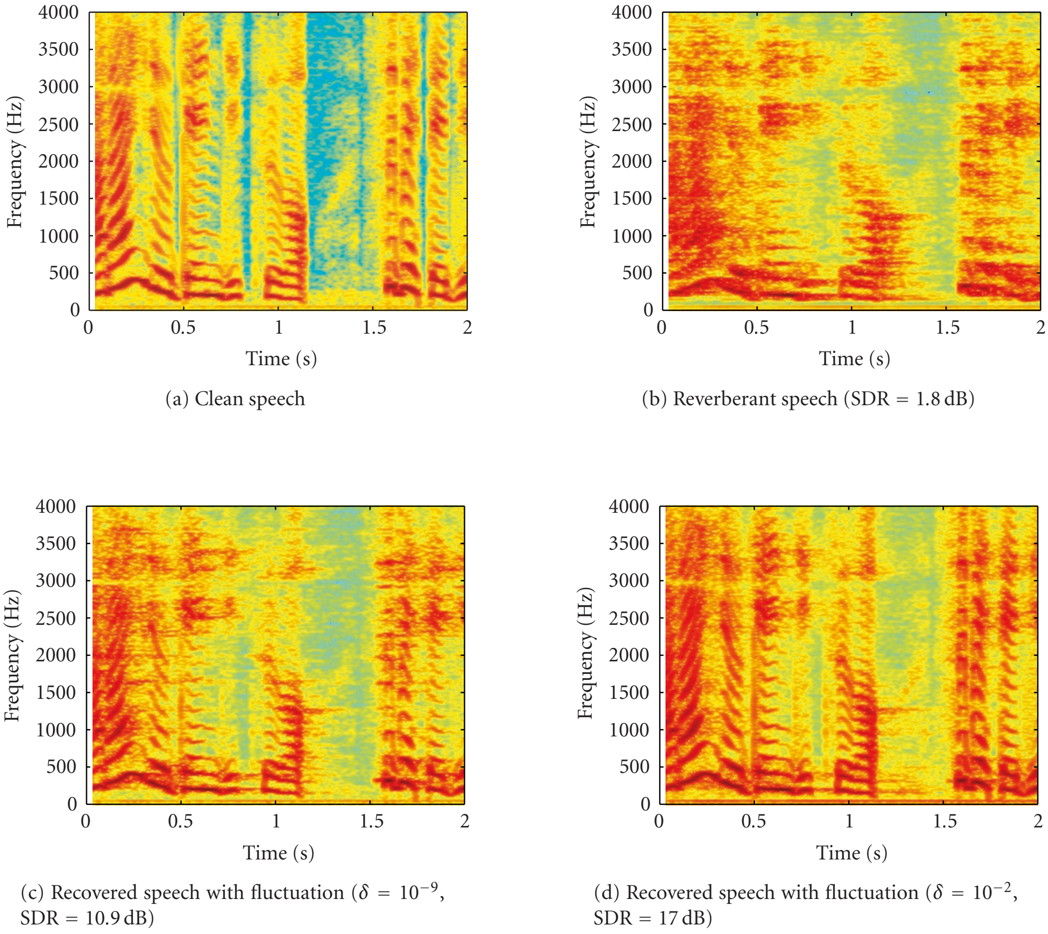
(2) $\mathcal{R} = E\langle \widetilde{\mathbf{H}}^T\widetilde{\mathbf{H}}\rangle \approx (1/8)\sum_{j=1}^{8}\widetilde{\mathbf{H}}_j^T\widetilde{\mathbf{H}}_j$, $\widetilde{\mathbf{H}}_j = \mathbf{H}_j - \overline{\mathbf{H}}$.

The performance of the inverse filter calculated with (15) is shown in Table 1. The performance with the correlation matrix in (2) is improved by 3.7 dB compared with the matrix in (1). This result shows the effect of incorporating the autocorrelation of the RTF fluctuations. If the time structure of the fluctuations could be obtained, for example by estimating the averaged autocorrelation of the fluctuation, more robust inverse filters could be obtained. Future work should include finding ways to estimate such fluctuation's time structure.

5.2. Results of speech dereverberation

Finally, the dereverberation performance is shown using speech signals. Figure 15 shows spectrograms of the (a) original, (b) reverberant, and (c), (d) dereverberated speech signals. The reference RTFs were used to calculate the inverse filter, and the RTFs corresponding to the 5th new position in Figure 10 were used to calculate the reverberant speech and for dereverberation. The source position change is 1 cm. The filter length was set at $M = 1333$, and the modeling delay was $d = 500$. The SDR of the reverberant speech is 1.8 dB. Figure 15(c) shows a spectrogram of the dereverberated speech signal filtered by the inverse filter with the regularization parameter $\delta = 10^{-9}$. Although the figure appears less reverberant than Figure 15(b), there is some degradation and an SDR of 10.9 dB was obtained. Figure 15(d) shows a spectrogram of the dereverberated speech filtered by the inverse filter with $\delta = 10^{-2}$. When the proper regularization parameter was used, the SDR improved by up to 17 dB. This SDR value is 5 dB higher than that obtained using a white signal as shown in Figure 13. This difference comes from the fact that the distortion mainly occurs in the higher frequency range, where speech has low energy.

Figure 16(a) shows a spectrogram of noisy and reverberant speech. The SNR level at the microphone is 20 dB, and the SDR with respect to the source speech signal is 0.5 dB. Figure 16(b) shows a spectrogram of the dereverberated signal when $\delta = 10^{-9}$ is used. The SDR of the dereverberated speech signal is 5.1 dB. Although it appears less reverberant, the frequency components of the speech are buried in those of the noise. This is because the incoming noise was amplified by the filter. Figure 16(c) shows a spectrogram of the dereverberated signal when $\delta = 10^{-2}$ is used. When the proper regularization parameter was used, the noise became less noticeable, because the filter energy was small. As a result, an SDR of 15.9 dB was achieved while the output SNR was kept over 20 dB.



(a) Clean speech

(b) Reverberant speech (SDR = 1.8 dB)

(c) Recovered speech with fluctuation ($\delta = 10^{-9}$, SDR = 10.9 dB)

(d) Recovered speech with fluctuation ($\delta = 10^{-2}$, SDR = 17 dB)

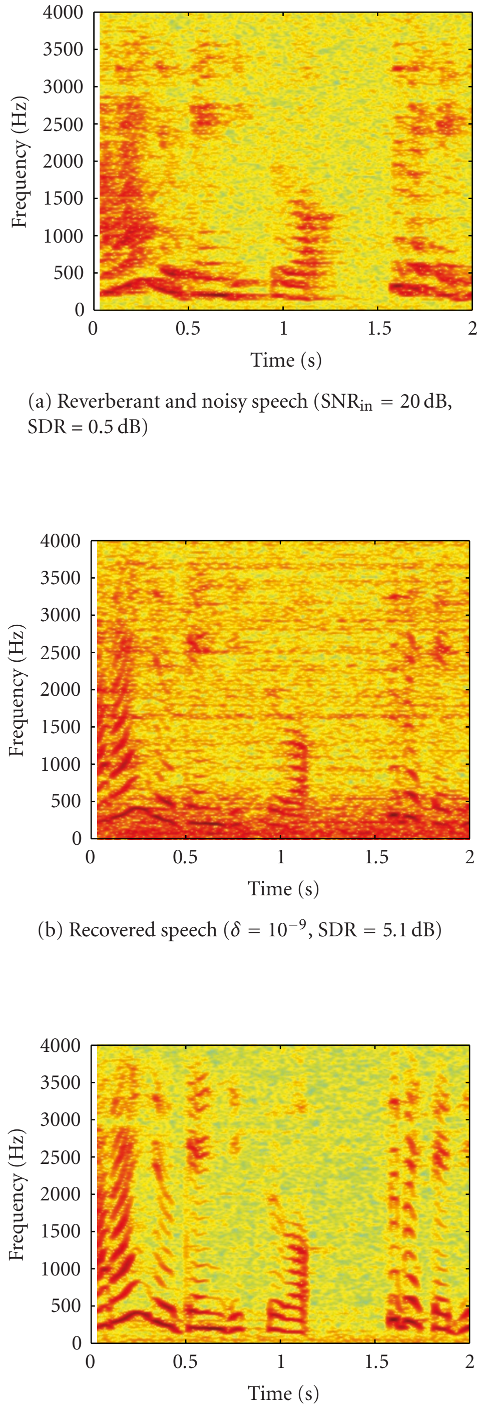
Figure 15: Spectrograms of speech signals.

6. CONCLUSION

With a view of extending the applicability of inverse-filter-based dereverberation, this paper examined a design method for an inverse filter, in which the filter design parameters were adjusted to reduce the filter energy. The regularization parameter, modeling delay, and filter length were selected to improve the performance when the RTFs fluctuated and when slight interference noise was present at the microphone signals. Simulation results showed that the inverse filtering performance could be improved by properly adjusting the design parameters, which led to a reduction in the filter energy. Consequently, this approach was shown to be effective for both RTF fluctuation and interference noise.

We discussed the differences between the results we obtained for RTF fluctuations and white noise. We observed that the performance with the regularization parameter did not improve greatly with regard to the RTF fluctuations, while the performance for the white noise showed a clear peak corresponding to the input SNR level. This is because RTF fluctuations are not random, and the regularized inverse filter implicitly assumes that the fluctuation is random. To demonstrate this, we used the autocorrelation of the fluctuation to calculate the inverse filter. The simulation result revealed that the RTF fluctuation had time structures. Future work thus includes finding ways to incorporate such fluctuation's time structures into the filter design process.

Systematic determination of the design parameters also remains as future work. Among the design parameters, a proper choice of the regularization parameter was important for the improvement in the performance, and the choice of the filter length and the modeling delay was less crucial than the regularization parameter. In the noisy case, the optimum regularization parameter that provides the best performance corresponds to the input SNR level, as shown in Figure 9. Thus, one way to determine the parameter is through the estimation of the input SNR [20]. For the RTF fluctuations, on the other hands, automatic determination of the parameter may not be simple. However, we observed from the results shown in Figure 13 that a relatively large value such as $\delta = 10^{-1}$ was effective in avoiding the degradation for small positional changes. Thus, using such a large value may be one solution for the RTF fluctuations.



(a) Reverberant and noisy speech ($\mathrm{SNR}_{\mathrm{in}} = 20\,\mathrm{dB}$, $\mathrm{SDR} = 0.5\,\mathrm{dB}$)

(b) Recovered speech ($\delta = 10^{-9}$, $\mathrm{SDR} = 5.1\,\mathrm{dB}$)

(c) Recovered speech ($\delta = 10^{-2}$, $\mathrm{SDR} = 15.9\,\mathrm{dB}$, $\mathrm{SNR}_{\mathrm{out}} = 20\,\mathrm{dB}$)

Figure 16: Spectrograms of speech signals.

ACKNOWLEDGMENT

The authors thank Mr. Takeaki Kubota of Nagoya University for arranging the experimental data and conducting the simulation described in the discussion (Section 5).

REFERENCES

[1] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, 1988.

[2] K. Furuya and Y. Kaneda, "Two-channel blind deconvolution of nonminimum phase FIR systems," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E80-A, no. 5, pp. 804–808, 1997.

[3] T. Hikichi, M. Delcroix, and M. Miyoshi, "Blind dereverberation based on estimates of signal transmission channels without precise information on channel order," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 1, pp. 1069–1072, Philadelphia, Pa, USA, March 2005.

[4] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 882–895, 2005.

[5] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *Journal of Sound and Vibration*, vol. 102, no. 2, pp. 217–228, 1985.

[6] T. Hikichi and F. Itakura, "Time variation of room acoustic transfer functions and its effects on a multi-microphone dereverberation approach," in *Proceedings of the Workshop on Microphone Arrays: Theory, Design and Application*, Piscataway, NJ, USA, October 1994.

[7] M. Omura, M. Yada, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Compensating of room acoustic transfer functions affected by change of room temperature," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, vol. 2, pp. 941–944, Phoenix, Ariz, USA, March 1999.

[8] B. D. Radlović, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, 2000.

[9] F. Talantzis and D. B. Ward, "Robustness of multichannel equalization in an acoustic reverberant environment," *The Journal of the Acoustical Society of America*, vol. 114, no. 2, pp. 833–841, 2003.

[10] H. Tokuno, O. Kirkeby, P. A. Nelson, and H. Hamada, "Inverse filter of sound reproduction systems using regularization," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E80-A, no. 5, pp. 809–820, 1997.

[11] P. C. Hansen, "The truncated SVD as a method for regularization," *BIT Numerical Mathematics*, vol. 27, no. 4, pp. 534–553, 1987.

[12] Y. Tatekura, Y. Nagata, H. Saruwatari, and K. Shikano, "Adaptive algorithm of iterative inverse filter relaxation to acoustic fluctuation in sound reproduction system," in *Proceedings of the 18th International Congress on Acoustics (ICA '04)*, vol. 4, pp. 3163–3166, Kyoto, Japan, April 2004.

[13] Y. Tatekura, S. Urata, H. Saruwatari, and K. Shikano, "On-line relaxation algorithm applicable to acoustic fluctuation for inverse filter in multichannel sound reproduction system," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 7, pp. 1747–1756, 2005.

[14] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, 1998.

[15] D. A. Harville, *Matrix Algebra from a Statistician's Perspective*, Springer, New York, NY, USA, 1997.

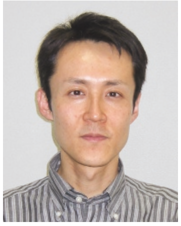
[16] S. J. Elliott, C. C. Boucher, and P. A. Nelson, "The behavior of a multiple channel active control system," *IEEE Transactions on Signal Processing*, vol. 40, no. 5, pp. 1041–1052, 1992.

[17] J. W. Hilgers, "On the equivalence of regularization and certain reproducing kernel Hilbert space approaches for solving first kind problems," *SIAM Journal on Numerical Analysis*, vol. 13, no. 2, pp. 172–184, 1976.

[18] A. Kaminuma, S. Ise, and K. Shikano, "A method of designing inverse system multi-channel sound reproduction system using least-norm-solution," in *Proceedings of the International Symposium on Active Control of Sound and Vibration (Active '99)*, vol. 2, pp. 863–874, Fort Lauderdale, Fla, USA, December 1999.

[19] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[20] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, 2001.



Takafumi Hikichi was born in Nagoya, in 1970. He received his Bachelor and Master of Electrical Engineering degrees from Nagoya University in 1993 and 1995, respectively. In 1995, he joined the Basic Research Laboratories of NTT. He is currently working at the Signal Processing Research Group of the Communication Science Laboratories, NTT. He is a Visiting Associate Professor of the Graduate School of Information Science, Nagoya University. His research interests include physical modeling of musical instruments, room acoustic modeling, and signal processing for speech enhancement and dereverberation. He received the 2000 Kiyoshi-Awaya Incentive Awards, and the 2006 Satoh Paper Awards from the ASJ. He is a Member of IEEE, ASA, ASJ, and IEICE.

Marc Delcroix was born in Brussels in 1980. He received the Master of Engineering from the Free University of Brussels and Ecole Centrale Paris in 2003. He is currently doing his Ph.D. at the Graduate School of Information Science and Technology of Hokkaido University. He is doing his research on speech dereverberation in collaboration with NTT Communication Science Laboratories. He received the 2006 Satoh Paper Awards from the ASJ. He is a Member of IEEE and ISCA.

Masato Miyoshi received the M.E. degree from Doshisha University in Kyoto in 1983. Since joining NTT as a Researcher that year, he has been engaged in the research and development of acoustic signal processing technologies. Currently, he is a Group Leader of the Media Information Laboratory of NTT Communication Science Laboratories in Kyoto. He is also a Visiting Associate Professor of the Graduate School of Information Science and Technology, Hokkaido University. He received the 1988 IEEE ASSP Senior Awards, the 1989 ASJ Kiyoshi-Awaya Incentive Awards, and the 1990 and 2006 ASJ Satoh Paper Awards. He also received the Ph.D. degree from Doshisha University in 1991. He is a Member of IEEE, AES, ASJ, and IEICE.