INVERSE FILTERING FOR SPEECH DEREVERBERATION LESS SENSITIVE TO NOISE

[1]*Takafumi Hikichi*, [1,2]*Marc Delcroix* *and* [1,2]*Masato Miyoshi*

[1]hikichi@cslab.kecl.ntt.co.jp
[1]NTT Communication Science Laboratories, NTT Corporation
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237, Japan
[2] Graduate School of Information Science and Technology, Hokkaido University
Kita 14, Nishi 9, Kita-ku, Sapporo, 060-0814 Japan

ABSTRACT

Inverse filtering of room transfer functions (RTFs) is an attractive approach for speech dereverberation because high performance could be achieved. However, speech signals received at the microphones may suffer from disturbances such as RTF fluctuations and interfering noise. In such cases, the dereverberation performance may be severely degraded. In our previous report, we proposed reducing the energy of the inverse filter to make the filter less sensitive to RTF fluctuations. This paper evaluates this design method for the inverse filter in terms of the filter's sensitivity to additive noise. The experimental results show that the proposed method is effective in the presence of additive noise, as well as RTF fluctuations.

1. INTRODUCTION

Speech dereverberation is important for various speech applications such as hands-free telephony and automatic speech recognition with distant speakers. Of the existing dereverberation approaches [1, 2, 3, 4, 5, 6, 7], the techniques based on the inverse filtering of room transfer functions (RTFs) appear attractive since high performance could be achieved [4, 5, 6, 7]. However, these techniques are affected by disturbances on the received signals. One cause of the disturbances is the fluctuation in the RTFs resulting from changes in such factors as source position and temperature. Another cause of the disturbances is interfering noise.

In [8] we investigated the problem of RTF fluctuations caused by source position changes, and studied the effect of the inverse filter design parameters on the dereverberation performance. As described later, three design parameters were adjusted to reduce the filter energy. We showed that reducing the filter energy makes the inverse filter less sensitive to RTF fluctuations.

In this paper, we propose using the same strategy to handle disturbance caused by additive noise. Experiments are carried out to evaluate the effect of the design parameters on the dereverberation performance in the presence of noise. We also conducted an experiment in which we combined the disturbances coming from RTF fluctuations and noise.

2. DEREVERBERATION METHOD AND DESIGN PARAMETERS

2.1. Dereverberation algorithm

The dereverberation algorithm proposed in [9, 10] is used in this study. The algorithm is summarized briefly here. First, RTFs are estimated from the received signals. Then, the inverse filter is calculated using these RTF estimates based on the Multiple input/output INverse Theorem (MINT) [4]. A similar two-stage approach has also been used in [5, 11, 12].

It is difficult to estimate the RTFs from noisy and reverberant speech signals. Although this issue has been tackled recently [13], it is still an open problem. Moreover, even if we could remove the effect of noise, when the RTF order is overestimated, the estimates contain a common polynomial between the channels as well as the true RTFs. In [9, 10], we proposed the use of post-processing to remove the effect of this common polynomial. In this paper, however, as we are focusing on the design of inverse filters, we assume that the RTFs could be estimated with sufficient accuracy.

Using the RTF estimates, we can obtain an inverse filter by solving the following equation,

$$\mathbf{H}\mathbf{g} = \mathbf{v}, \qquad (1)$$

where

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

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

$$i = 1, \cdots, P,$$

$$\mathbf{g} = [\underbrace{g_1(1), \dots, g_1(M), \cdots, g_P(1), \dots, g_P(M)}_{PM}]^T,$$

$$\mathbf{v} = [\underbrace{0, \dots, 0}_{d}, 1, 0, \dots, 0]^T,$$

P is the number of channels, $h_i(n)$ is the impulse response estimate between the source and the i-th microphone, J is the number of taps of the impulse response estimates, $\mathbf{g}$ is the inverse filter vector, M is the filter length for each channel, and d is a modeling delay. An arbitrary delay can be inserted in the equalized response by setting d ($d \geq 0$). Matrix $\mathbf{H}$ is full row rank assuming the RTFs have no commmon zeros. Hereafter, we consider that the impulse response estimates are normalized by their norm. The inverse filter vector can be obtained by

$$\mathbf{g} = \mathbf{H}^{+}\mathbf{v}, \tag{2}$$

where $\mathbf{A}^{+}$ is the Moore-Penrose pseudo inverse of matrix $\mathbf{A}$. An inverse filter with the minimum length is calculated by setting M so that matrix $\mathbf{H}$ is square, i.e., $(J + M) = PM$, which leads to $M = J/(P-1)$. Note that the filter length can also be set at $M > J/(P-1)$.
By applying the inverse filter to the observed signals, a dereverberated signal is obtained.

2.2. Design parameters

Here, we briefly explain the three design parameters, namely, the regularization parameter, filter length, and modeling delay, and their influence on the inverse filter. We can expect these parameters to be effective in reducing the filter energy and hence increasing the robustness against disturbances in the room soundfield. A more detailed analysis of the effect of the design parameters on the filter energy can be found in [8].

2.2.1. Regularization parameter δ

The use of a regularization parameter in the design of the inverse filter leads to the following expression for the inverse filter vector,

$$\mathbf{g}_r = (\mathbf{H}^T\mathbf{H} + \delta\mathbf{I})^{-1}\mathbf{H}^T\mathbf{v}, \tag{3}$$

where $\delta(\geq 0)$ is the regularization parameter, and $\mathbf{I}$ is an identity matrix. The l_2-norm of this filter vector satisfies the following relation [8],

$$||\mathbf{g}_r||^2 \leq ||\mathbf{g}||^2. \tag{4}$$

That is, the regularization parameter δ has the effect of reducing the norm of the inverse filter, and this is believed to reduce the sensitivity to noise. On the other hand, the regularization parameter reduces the accuracy of the inverse filter, and a compromise should be adopted. It should be noted that the filter expressed as Eq. (3) gives the optimum solution when the RTFs show random fluctuations with variance δ.

2.2.2. Filter length M

Equation (3) will give the minimum norm filter for a given length M. Consequently, by increasing filter length M, we can expect to find a filter with the smallest norm among all possible filters.

2.2.3. Modeling delay d

Modeling delay d ($d > 0$) is inserted to compensate for the maximum phase component of the RTFs and to stabilize the inverse filter. Hence, we expect the filter norm to be reduced by choosing an appropriate delay.

3. EXPERIMENTS

Simulations were used to investigate the sensitivity of the inverse filter to the additive noise included in the observed signals. Figure 1 shows the arrangement of the source and microphones used in the experiment. Room impulse responses between the source and the microphones are simulated by using the image method [14]. The impulse responses are truncated to 1600 samples ($J = 1599$), corresponding to -60 dB attenuation. The sampling frequency is set at 8 kHz, then the duration of the impulse responses is 200 msec. The experimental conditions used in this study are the same as those used in our previous study [8].
Reverberant speech signals are simulated by convolving the original speech with the room impulse responses. Then, signals observed at the microphones are simulated by adding white noise with an SNR of 40 dB. These observed signals are filtered with the inverse filter calculated by Eq. (3) to obtain the dereverberated speech signal. Experiments were undertaken using several microphone pairs (P=2) made up from the four microphones, for example (M1, M2), (M2, M3). Due to the space limitation, we only show the results for the microphone pair (M3, M4). However, a similar tendency was observed for the other microphone pairs.

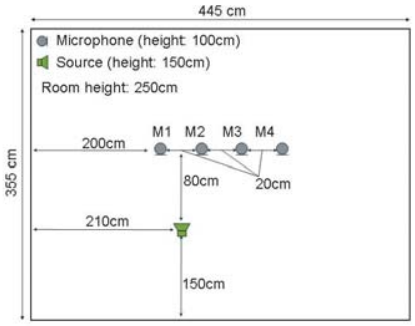

Figure 1: *Arrangement of the source and microphones. M1, M2, M3 and M4 denote the microphones.*

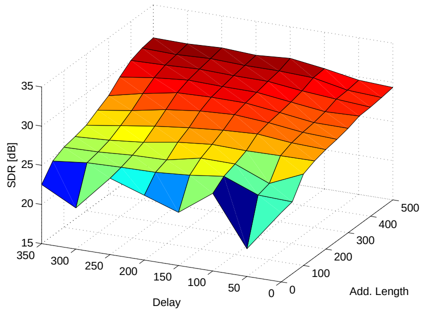

Figure 2: *Performance as a function of additional filter length M' and modeling delay d. Regularization parameter was set at δ = 0.*

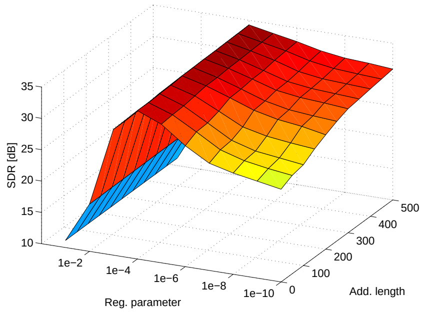

Figure 3: *Performance as a function of additional filter length M' and regularization parameter δ. Modeling delay was set at d=200.*

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

$$SDR = 10\log_{10}\left(\frac{\sum_{n=0}^{N} s^2(n)}{\sum_{n=0}^{N}(s(n)-\hat{s}(n))^2}\right), \qquad (5)$$

where $s(n)$ and $\hat{s}(n)$ are the original and the dereverberated speech signals, respectively.

Figure 2 shows the performance of the inverse filter designed with various filter lengths M $(= J + M'$, M': additional filter length) and modeling delay d. When the minimum filter length was used (M'=0), the performance was strongly dependent on the delay. By contrast, when a larger filter length was used (i.e. $M' = 500$), better and more stable performance was obtained. We investigated the filter energy and confirmed that it is inversely proportional to the performance.

In the second experiment, the modeling delay was fixed at $d = 200$, and the effects of filter length M and the regularization parameter δ were investigated. Figure 3 shows the performance in this case. The best performance was obtained with $\delta = 10^{-4}$. Hereafter, the regularization parameter value that provided the best performance is referred to as the best value. The best value corresponds with the SNR level (40 dB). The dependence on filter length becomes small when the best parameter value is used. It should be noted that, although the filter energy decreases with increases in δ, too large a δ value, such as $\delta \geq 10^{-2}$ also degrades the dereverberation performance.

In the third experiment, we evaluated the performance for several SNRs by using modeling delay $d = 200$ and minimum filter length $M = J$. Figure 4 shows the results for SNR = 10, 20, 30, and 40 dB. When the SNR is high, the best value of the regularization parameter corresponds well with the SNR level. In contrast, when the SNR is low, the performance curve becomes broad, and the correspondence becomes less obvious. Note that the performance is bounded by the SNR level, since no noise reduction is employed.

Figure 5 shows the performance as a function of the regularization parameter when there were RTF fluctuations resulting from source position changes. Here, we consider new source positions and apply the inverse filter to each of the corresponding reverberant speech signals. SDR values calculated by Eq. (5) are averaged over the new positions to obtain the overall performance. We evaluated the performance for position changes of 2, 4, 6 and 8 cm. Figures 4 and 5 exhibit similar trends. However, the performance is less sensitive to the best value of the regularization parameter than when noise is present.

Figure 6 shows the performance when both the RTF fluctuation and the noise are present simultaneously. The results for ‘noise only’ and ‘fluctuation only’ cases are also plotted in the same figure. In terms of the performance at $\delta = 10^{-3}$, the difference between ‘noise only’ and ‘fluctuation only’ is about 2 dB. The performance worsens by about 4 dB from ‘noise only’ to ‘noise+fluctuation’. These results are plausible if the noise induced distortion has no correlation with the distortion caused by the RTF fluctuation.

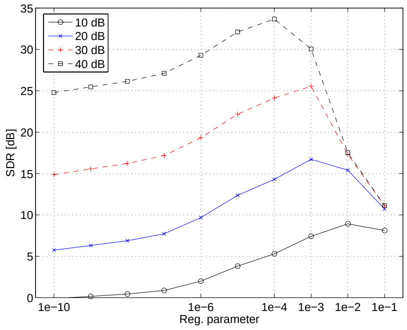

Figure 4: *Performance as a function of regularization parameter for SNR values of 10, 20, 30, and 40 dB.*

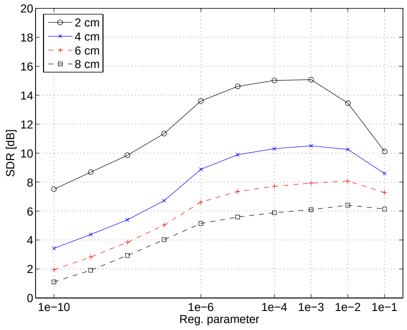

Figure 5: *Performance as a function of regularization parameter for position variation of 2, 4, 6, and 8 cm.*

4. SUMMARY

To achieve dereverberation in a noisy environment, we evaluated the inverse filter design method in terms of the filter's sensitivity to additive noise. The results showed that the dereverberation performance could be improved by properly adjusting the filter design parameters, which led to a reduction of the filter energy. Consequently, this approach was shown to be effective for additive noise, as well as for RTF fluctuation, as reported in [8].

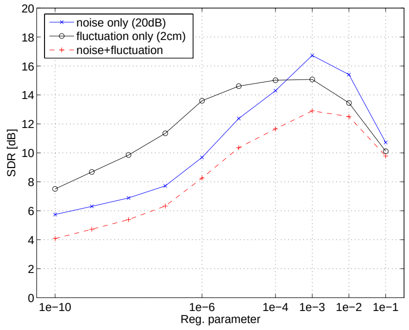

Figure 6: *Performance comparison for distortions caused by noise, RTF fluctuation, and noise + RTF fluctuation.*

5. REFERENCES

[1] B. Yegnanarayana and P. S. Murthy, "Enhancement of reverberant speech using LP residual signal," *IEEE Trans. SAP*, vol. 8, no. 3, pp. 267–281, 2000.

[2] N. Gaubitch, P. Naylor, and D. Ward, "Multi-microphone speech dereverberation using spatio-temporal averaging," *Proceedings of EUSIPCO EURASIP*, pp. 809–812, 2004.

[3] E. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," *Proceedings of ICASSP IEEE*, vol. 4, pp. 173–176, 2005.

[4] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. ASSP*, vol. 36, no. 2, pp. 145–152, 1988.

[5] 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 Trans. SAP*, vol. 13, no. 5, pp. 882–895, 2005.

[6] T. Yoshioka, T. Hikichi, M. Miyoshi, and H. G. Okuno, "Robust decomposition of inverse filter of channel and prediction error filter of speech signal for dereverberation," *Proceedings of EUSIPCO EURASIP (in press)*.

[7] M. Delcroix, T. Hikichi, and M. Miyoshi, "Precise dereverberation using multi-channel linear prediction," *IEEE Trans. ASLP (in press)*.

[8] T. Hikichi, M. Delcroix, and M. Miyoshi, "On robust inverse filter design for room transfer function fluctuations," *Proceedings of EUSIPCO EURASIP (in press)*.

[9] T. Hikichi, M. Delcroix, and M. Miyoshi, "Blind dereverberation based on estimates of signal transmission channels without precise information on channel order," *Proceedings of ICASSP IEEE*, vol. 1, pp. 1069–72, 2005.

[10] T. Hikichi, M. Delcroix, and M. Miyoshi, "Speech dereverberation algorithm using transfer function estimates with overestimated order," *Acoust. Sci. and Tech.*, vol. 27, no. 1, pp. 28–35, 2006.

[11] M. I. Gurelli and C. L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. SP*, vol. 43, no. 1, pp. 134–149, 1995.

[12] K. Furuya and Y. Kaneda, "Two-channel blind deconvolution of nonminimum phase FIR systems," *IEICE Trans. Fundamentals*, vol. E80-A, no. 5, pp. 804–808, 1997.

[13] N. Gaubitch, M. K. Hasan, and P. Naylor, "Noise robust adaptive blind channel identification using spectral constraints," *Proceedings of ICASSP IEEE*, vol. 5, pp. 93–96, 2006.

[14] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.