BLIND DEREVERBERATION BASED ON ESTIMATES OF SIGNAL TRANSMISSION CHANNELS WITHOUT PRECISE INFORMATION ON CHANNEL ORDER

Takafumi Hikichi   Marc Delcroix   Masato Miyoshi

NTT Communication Science Laboratories, NTT Corporation
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237 Japan
{hikichi,marc.delcroix,miyo}@cslab.kecl.ntt.co.jp

ABSTRACT

This paper addresses the blind dereverberation problem of single-input multiple-output acoustic systems. Most approaches require an exact knowledge of the order of the room transfer functions. In this paper, we propose an equalization algorithm that is less sensitive to the order of the estimated transfer functions. First, the transfer functions are estimated using an overestimated order, and the inverse filter set for this estimated transfer functions is calculated. Since the estimated transfer functions have a common part, the signal processed by the inverse filter set contains distortion. Then, we compensate for this distortion using a common polynomial extraction technique. This algorithm enables a reverberated speech signal to be dereverberated as long as the channel is overestimated. Simulation results show that the proposed method is robust even when the order is highly overestimated.

1. INTRODUCTION

The distortion of speech signals by reverberation in a room is a crucial problem in many applications. For example, reverberation severely changes signal characteristics, thus degrading the recognition performance in current automatic speech recognition (ASR) systems. Considerable effort has been devoted to the blind dereverberation problem, but no adequate solution has yet been established.

A class of techniques has been proposed based on the assumption that the source signal has *i.i.d.* characteristics [1][2]. A major problem with these techniques is that the dereverberation process causes excessive whitening of the speech signals. One way to cope with this problem is to design a good inverse filter for the transfer function between the source and receivers directly. The harmonic structure of speech has been successfully used to obtain a precise estimate of the inverse filter from reverberant speech [3][4]. This technique works well even for long reverberation times. However, since this technique is highly dependent on source signal characteristics, the technique does not seem to be applicable to a wide variety of source signals.

Another way is to estimate transfer functions directly regardless of the kind of source signal, and design the inverse filter using these transfer functions. Many techniques for blind identification/equalization have already been developed, including the subspace (SS) method, the least squares (LS) subchannel matching technique, and the linear prediction (LP) method [5]. However, most of the techniques require an exact knowledge of the channel order. In practical situations, methods that are insensitive to the transfer function order are desirable.

In this paper, we propose a blind equalization algorithm that requires no precise information about the channel order. As long as the channel is overestimated, dereverberation is achieved. The following section reviews the conventional blind dereverberation method based on subchannel matching, and clarifies problems. Section 3 describes our proposed method. Section 4 describes a simulation to show the effectiveness of the proposed method. Section 5 summarizes the paper.

2. BLIND DEREVERBERATION

2.1. Conventional subchannel matching

We focus on the single-source two-microphone acoustic system shown in Fig. 1. We assumed that room transfer functions $h_1(z), h_2(z)$ have no common zeros, and are represented as j-th order polynomials:

$$h_i(z) = h_{i,0} + h_{i,1}z^{-1} + \ldots + h_{i,j}z^{-j}, \;\; i = 1, 2.$$

A source signal is represented as $s(n)$, and signals received by the two microphones are $x_1(n), x_2(n)$, respectively. The objective of blind deconvolution is to recover the source signal based only on the received signals.

As shown in Fig. 1, signals $x_1(n)$ and $x_2(n)$ pass through the filters $\hat{h}_1(z), \hat{h}_2(z)$ with order k, and one filtered signal is subtracted from the other to obtain error signal $e(n)$.

$$\begin{aligned} e(n) &= x_1(n)\hat{h}_2(z) - x_2(n)\hat{h}_1(z) \\ &= s(n)\{h_1(z)\hat{h}_2(z) - h_2(z)\hat{h}_1(z)\} \end{aligned} \qquad (1)$$

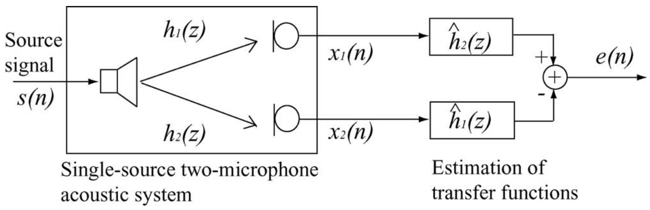

Fig. 1. Acoustic system and estimation of transfer functions in subchannel matching.

If $k = j$ and $e(n) = 0$ for all n, $\hat{h}_1(z)$ and $\hat{h}_2(z)$ are identified up to an arbitrary scalar, i.e.,

$$\hat{h}_1(z) = \alpha h_1(z), \;\; \hat{h}_2(z) = \alpha h_2(z). \tag{2}$$

In fact, the estimates are obtained by minimizing the mean square value of $e(n)$. The solution is obtained through the eigenvalue decomposition of the autocorrelation matrix of the observed signals. The estimate of the transfer functions is the eigenvector that corresponds to the minimum eigenvalue of the autocorrelation matrix [6][7].

When the transfer function estimates are obtained, an exact inverse filter set can be calculated using multi-channel inverse filter theory [8]. When inverse filters are expressed as $w_1(z), w_2(z)$, these filters satisfy

$$\begin{aligned} \hat{h}_1(z)w_1(z) + \hat{h}_2(z)w_2(z) &= 1, \\ h_1(z)w_1(z) + h_2(z)w_2(z) &= \frac{1}{\alpha}. \end{aligned} \tag{3}$$

This relation demonstrates that a perfect dereverberation is achieved.

2.2. Problems

If the order of the transfer function is overestimated, then $\hat{h}_1(z)$ and $\hat{h}_2(z)$ become

$$\begin{gathered} \hat{h}_1(z) = c(z)h_1(z), \;\; \hat{h}_2(z) = c(z)h_2(z) \\ c(z) = c_0 + c_1 z^{-1} + \ldots + c_m z^{-m}. \end{gathered} \tag{4}$$

where polynomial $c(z)$ is common to the two transfer function estimates. Let us consider inverse filtering using these transfer function estimates.

$$\begin{gathered} \hat{h}_1(z)w_1(z) + \hat{h}_2(z)w_2(z) = 1, \\ c(z)\{h_1(z)w_1(z) + h_2(z)w_2(z)\} = 1. \end{gathered} \tag{5}$$

Here, filters $w_1(z)$ and $w_2(z)$ no longer achieve inverse filtering the actual transfer functions, and naturally, the dereverberation fails. The transfer function order could be obtained if the dimension of the nullspace in the autocorrelation matrix of the observed signals is precisely calculated [7][9], i.e., by counting the number of the smallest eigenvalues. Another way to find the optimum order is to use a proper cost function [10][11]. Different from these methods, we propose a method where exact order estimation is not required as long as the order is overestimated.

3. PROPOSED ALGORITHM

3.1. Effect of overestimation

Inverse filters are calculated using overestimated transfer functions, as shown in Eq. (5). These inverse filters satisfy the following equation.

$$h_1(z)w_1(z) + h_2(z)w_2(z) \propto \frac{1}{c_{min}(z)}, \tag{6}$$

where $c_{min}(z)$ is a minimum phase polynomial where non-minimum phase zeros of $c(z)$ are reflected inside the unit circle on the z-plane (Proof is shown in Appendix). Here, $1/c_{min}(z)$ can be expressed as $1/c_{min}(z) = 1 - (d_1 z^{-1} + \ldots + d_{m'} z^{-m'})$. Equation (6) means that the dereverberated signal obtained through $w_1(z)$ and $w_2(z)$ sufferred from extra degradation, in other words, it was colored by $c_{min}(z)$. To remove this distortion, $c_{min}(z)$ is extracted from the transfer function estimates based on multi-channel linear prediction.

3.2. Common polynomial extraction

Applying the two-channel linear prediction scheme to the transfer function estimates $\hat{h}_1(z)$ and $\hat{h}_2(z)$, $1/c_{min}(z)$ can be extracted [12] as a characteristic polynomial of matrix $\mathbf{Q}$ as shown below.

$$\begin{aligned} \mathbf{Q} &= \lim_{\delta \to 0} (\hat{\mathbf{H}}^T \hat{\mathbf{H}} + \delta^2 \mathbf{I})^{-1} \hat{\mathbf{H}}^T \mathbf{S} \hat{\mathbf{H}} \\ &= \lim_{\delta \to 0} (\mathbf{H}^T \mathbf{C}^T \mathbf{C} \mathbf{H} + \delta^2 \mathbf{I})^{-1} \mathbf{H}^T \mathbf{C}^T \mathbf{S} \mathbf{C} \mathbf{H} \end{aligned} \tag{7}$$

where $\hat{\mathbf{H}}$ is a convolution matrix similarly defined as $\mathbf{H}$,

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2], \mathbf{H}_i = \begin{pmatrix} h_{i,0} & 0 & \ldots & 0 \\ h_{i,1} & h_{i,0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ h_{i,j} & & & h_{i,0} \\ 0 & h_{i,j} & & h_{i,1} \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & h_{i,j} \end{pmatrix}, \; i = 1, 2,$$

$$\mathbf{C} = \begin{pmatrix} c_0 & 0 & \ldots & 0 \\ c_1 & c_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ c_m & & & c_0 \\ 0 & c_m & & c_1 \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & c_m \end{pmatrix}, \mathbf{S} = \begin{pmatrix} 0 & 1 & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ & & \ddots & 1 \\ 0 & \ldots & \ldots & 0 \end{pmatrix}.$$

According to the definition of the Moore-Penrose generalized inverse, matrix $\mathbf{Q}$ can be further rewritten as,

$$\begin{aligned}\mathbf{Q} &= (\mathbf{CH})^{+}\mathbf{SCH} \\ &= \mathbf{H}^T(\mathbf{HH}^T)^{-1}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{SCH},\end{aligned} \quad (8)$$

where $+$ denotes the Moore-Penrose generalized inverse [13]. The characteristic polynomial of matrix $\mathbf{Q}$, $f_c(\mathbf{Q})$, is expressed as,

$$\begin{aligned}f_c(\mathbf{Q}) &= f_c(\mathbf{H}^T(\mathbf{HH}^T)^{-1}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{SCH}) \\ &= f_c(\mathbf{HH}^T(\mathbf{HH}^T)^{-1}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{SC}) \\ &= f_c((\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{SC}) \\ &= f_c(\begin{pmatrix} d_1 & 1 & 0 & \ldots & 0 \\ d_2 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ d_{m'-1} & \vdots & & 0 & 1 \\ d_{m'} & 0 & \ldots & & 0 \end{pmatrix}) \\ &= 1 - (d_1 z^{-1} + \ldots + d_{m'} z^{-m'}).\end{aligned} \quad (9)$$

That is, $1/c_{min}(z)$ can be calculated as the characteristic polynomial of matrix $\mathbf{Q}$.

By applying the filter $c_{min}(z)$ obtained by the above procedure to the deconvolved signal, we can compensate for the distortion described in section 3.1. The configuration of the proposed method is shown in Fig. 2.

3.3. Algorithm

We can summarize the proposed algorithm as follows:

1. Estimate the transfer functions based on the eigenvalue decomposition of the autocorrelation matrix.
2. Calculate the inverse filters using the transfer function estimates.
3. Filter the observed signals with the inverse filters.
4. Estimate the common polynomial shown in Eq. (9) from the transfer function estimates using two-channel linear prediction.
5. Compensate for the effect of this common polynomial and recover the input signal.

4. SIMULATIONS

We conducted simulations to test the described method in the ideal case of a noise free environment. The input signals were Japanese sentences, taken from ATR's speech database [14]. Room transfer functions were simulated using the image method [15]. The simulated room was designed as shown in Fig. 3. The room impulse responses were truncated to 300 taps corresponding to a duration of 18.75 ms. The truncated transfer functions were confirmed to be non-minimum phase.

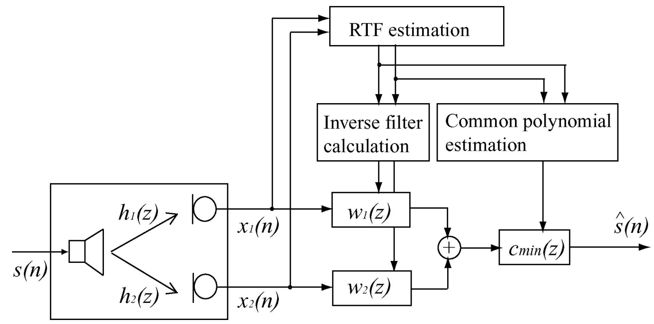

Fig. 2. Configuration of proposed method.

The simulation conditions are summarized in Table 1.

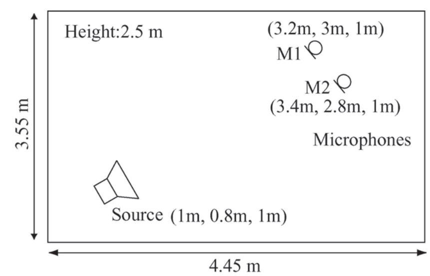

Fig. 3. Simulated soundfield.

Table 1. Simulation conditions.

Length of $h_1(z)$ and $h_2(z)$	300 taps
Length of $\hat{h}_1(z)$ and $\hat{h}_2(z)$	$300 \sim 600$ taps†
Reflection coefficient	0.8
Duration of speech signals	≈ 5 sec
Sampling frequency	16 kHz

†(Length of $c(z)$ $0 \sim 300$ taps)

We evaluate the performance using the signal to distortion ratio (SDR) as defined below:

$$SDR = 10\log_{10}\left(\frac{\sum|s(n)|^2}{\sum|s(n)-\hat{s}(n)|^2}\right), \quad (10)$$

where $s(n)$ is the input speech signal and $\hat{s}(n)$ is the estimated speech signal.

Figure 4 shows the performance when the estimated length of the transfer functions varied from 300 to 600 taps. The performance was very good (over 60 dB) when we used the correct length. Furthermore, the SDR values remained good even when the length was overestimated by up to 600.

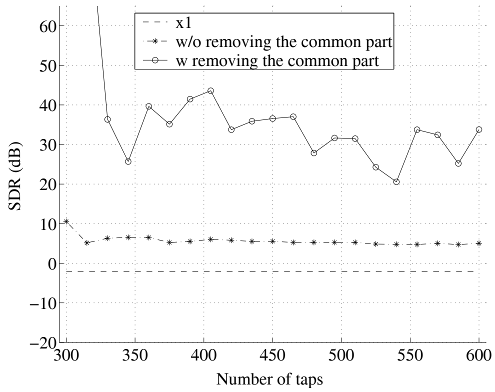

Fig. 4. Performance as a function of the estimated length (Solid line: proposed method, dashed line: received signal $x_1(n)$, dash-dotted line: dereverberated signal without cancelling $1/c_{min}(z)$).

The linear prediction based method successfully compensated for the distortion caused by order mismatch and provided an average SDR of 30 dB. Note that the proposed method never use the correct order of $h_1(z)$ and $h_2(z)$. The method only uses the order of $\hat{h}_1(z)$ and $\hat{h}_2(z)$.

5. SUMMARY

We proposed a speech dereverberation algorithm robust to the order overestimation of the room transfer functions. First, the inverse filters are calculated using the estimates of the possibly overestimated transfer functions, then the distortion introduced by overestimation is compensated for. Simulation results show that the proposed method is robust even when the order is highly overestimated.

As future work, we must consider robustness to noise. The computer accuracy problem also has to be investigated. Since increasing the length of the transfer functions leads to the calculation of larger matrix $\mathbf{Q}$, we have to study accurate calculation methods of $f_c(\mathbf{Q})$ shown in Eq. (9) for large matrix $\mathbf{Q}$.

6. APPENDIX

Using a matrix form, Eq. (5) can be expressed as,

$$\mathbf{CHw} = [1, 0, \ldots, 0]^T = \mathbf{J}.$$

Using the Moore-Penrose generalized inverse [13], $\mathbf{w}$ can be calculated as

$$\mathbf{w} = \mathbf{H}^T(\mathbf{HH}^T)^{-1}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{J}.$$

Using this $\mathbf{w}$, the left hand side of Eq. (6) becomes,

$$\begin{aligned}\mathbf{Hw} &= \mathbf{HH}^T(\mathbf{HH}^T)^{-1}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{J} \\ &\propto (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{J}\end{aligned}$$

Since $(\mathbf{C}^T\mathbf{C})^{-1}$ calculated when $c(z)$ is a minimum phase polynomial is equivalent to the one when $c(z)$ is a nonminimum phase polynomial. This corresponds to $1/c_{min}(z)$.

7. REFERENCES

[1] X. Sun and S. C. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," *Proceedings of the ICA*, pp. 59–64, 2001.

[2] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," *Proc. IEEE Workshop on Signal Processing in Advances in Wireless Communications*, pp. 101–104, 1997.

[3] T. Nakatani, M. Miyoshi, and K. Kinoshita, "One microphone blind dereverberation based on quasi-periodicity of speech signals," *Advances in Neural Information Processing Systems (NIPS) 16 (to appear)*, 2004.

[4] T. Nakatani and M. Miyoshi, "Blind dereverberation of single channel speech signal based on harmonic structure," *Proceedings of the ICASSP IEEE*, vol. 1, pp. 92–95, 2003.

[5] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal processing advances in wireless and mobile communications*, Prentice Hall PTR, 2001.

[6] 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.

[7] 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.

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

[9] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP J. Applied Signal Processing*, vol. 2003, no. 11, pp. 1074–1090, 2003.

[10] W. Bobillet, E. Grivel, R. Guidorzi, and M. Najim, "Cancelling convolutive and additive coloured noises for speech enhancement," *Proceedings of the ICASSP IEEE*, vol. 2, pp. 777–780, 2004.

[11] I. Santamaria, J. Via, and C. C. Gaudes, "Robust blind identification of SIMO channels: a support vector regression approach," *Proceedings of the ICASSP IEEE*, vol. 5, pp. 673–676, 2004.

[12] T. Hikichi and M. Miyoshi, "Blind algorithm for calculating the common poles based on linear prediction," *Proceedings of the ICASSP IEEE*, vol. 4, pp. 89–92, 2004.

[13] D. A. Harville, *Matrix algebra from a statistician's perspective*, pp. 493–495, Springer-Verlag, 1997.

[14] ATR International, *Speech database*, http://www.red.atr.co.jp/database_page/digdb.html, (in Japanese).

[15] 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.