A Simple Mutual Deconvolution Algorithm for Acoustic Blind Dereverberation

Lei Liu and Wayne T. Padgett

***Abstract*—Reverberation is one of the main causes of speech degradation in audio applications. This paper proposes a novel multi-microphone approach that estimates impulse responses to dereverberate the speech without a priori information about the acoustic channel characteristics, statistical properties of speech and noise, or locations of source and microphones. The impulse responses approximated by a simple first-order method are utilized in a reverberant linearly constrained minimum variance (LCMV) model. Simulation results show an improvement in dereverberation compared to using the multi-channel least mean square (MCLMS) impulse response estimator.**

***Index Terms*—Blind channel identification, dereverberation, microphone array, noise reduction**

I. Introduction

A major cause of speech degradation in most practical audio recording applications is room reverberation. In daily talking conversations, although normal people are not affected by such degradation to some degree, audio quality suffers a lot in a machine recording system from the damaged speech intelligibility.

Currently, most research is focused on dereverberation with a microphone array. Two groups of methods exist based on whether or not they try to find coefficients in an equalizer to filter received array signals. Among the non-equalizer approaches, beamforming would be the most popular method. The original beamforming method is described in [1] and [2]. Although simple and robust, the performance is greatly limited by the shape and dimensions of the array system.Another class of dereverberation methods attempts to equalize the received signal to recover the original speech signal. Such methods are divided into two categories: one is based on the multiple input output inverse theorem (MINT) [3],to estimate the room impulse response first, then to find the inverse filter, and another one is to directly search for the inverse filter without making any identifications on the channel.

For those methods that require knowing channel characteristics, robustly estimating an acoustic impulse response in the presence of noise becomes a very important but difficult task. The cepstral mean method is based on the assumption that clean speech is uncorrelated with adjacent time windows, so averaging over several cepstrums results in an impulse response estimation [4]. Adaptive algorithms have been used due to their inherent ability to track the acoustic impulse response in a slowly varying context. The two most popular algorithms are the Multichannel Least-Mean-Square (MCLMS) [5] and the Normalized Multichannel Frequency-Domain LMS (NMCFLMS) [6].

In this paper, we proposed the mutual deconvolution method to estimate the acoustic impulse response, and combined it with the dereverberant linearly constrained minimum variance algorithm [7] to achieve blind dereverberation. Simulations showed our approach had efficiently removed echoes and reflections to a large degree.

This paper is organized as the following: Section II formulates the reverberation problem. Section III presents our proposed mutual deconvolution impulse response estimation method. Simulations taking the dereverberant LCMV algorithm with impulse response estimated by our method are presented in Section IV. It also includes performance comparison with the multi-channel least mean square (MCLMS) algorithm. Finally, conclusions and future work are discussed in Section V.

II. Problem Fomulation

Assuming reverberation is made up of a direct path signal and many delayed attenuated replicas, the impulse responses from the speech source to each microphone are

$$\begin{aligned} g_1(t) &= \alpha_1\delta(t-t_{11}) + \beta_1\delta(t-t_{12}) + \gamma_1\delta(t-t_{13}) + \cdots \\ g_2(t) &= \alpha_2\delta(t-t_{21}) + \beta_2\delta(t-t_{22}) + \gamma_2\delta(t-t_{23}) + \cdots \\ &\vdots \\ g_N(t) &= \alpha_N\delta(t-t_{N1}) + \beta_N\delta(t-t_{N2}) + \gamma_N\delta(t-t_{N3}) + \cdots \end{aligned} \quad (1)$$

where N is the number of sensors,

$\alpha_1, \beta_1, \gamma_1, \cdots, \alpha_2, \cdots, \alpha_N, \beta_N, \gamma_N, \cdots$ are attenuated weights in signal transmission,

and $t_{11}, t_{12}, t_{13}, \cdots, t_{21}, \cdots, t_{N1}, t_{N2}, t_{N3}, \cdots$ are the reflections' delay time after the direct path signal is received.

In the frequency domain, we have

$$\begin{aligned} Y_1 &= X * G_1 \\ &= X * (\alpha_1 e^{-j\omega t_{11}} + \beta_1 e^{-j\omega t_{12}} + \gamma_1 e^{-j\omega t_{13}} + \cdots) \\ Y_2 &= X * G_2 \\ &= X * (\alpha_2 e^{-j\omega t_{21}} + \beta_2 e^{-j\omega t_{22}} + \gamma_2 e^{-j\omega t_{23}} + \cdots) \\ &\vdots \\ Y_N &= X * G_N \\ &= X * (\alpha_N e^{-j\omega t_{N1}} + \beta_N e^{-j\omega t_{N2}} + \gamma_N e^{-j\omega t_{N3}} + \cdots) \end{aligned} \quad , (2)$$

Manuscript received July 23, 2012. This work was supported by the Rose-Hulman Institute of Technology.

Lei Liu is with the Electrical and Computer Engineering Department, Rose-Hulman Institute of Technology, Terre Haute, IN 47803 USA (phone: 812-241-8088; e-mail: liulei32@gmail.com).

Wayne T. Padgett is with the Electrical and Computer Engineering Department, Rose-Hulman Institute of Technology, Terre Haute, IN 47803 USA (e-mail: padgett@rose-hulman.edu).

where $Y_i, i=1,2,\cdots,N$ and X are Fourier transforms of the received signal at each microphone and the source signal, respectively.

Our goal of impulse response estimation becomes blindly finding the attenuating weights and delay times, only from the observations of y_i without any prior information on x.

Before introducing our proposed estimation method, we assume that the channel can be identified. According to [8], two inductive conditions must be satisfied to ensure system identifiability for all second order statistics-based blind channel identification methods:

(1) The polynomials formed from $g_i, i=1,2,\cdots N$ are co-prime, i.e. the channel transfer functions G_i do not share any common zeros;

(2) The autocorrelation matrix $R_{xx}=E[x(n)x^T(n)]$ of the source signal is of full rank, where $E[\bullet]$ denotes expectation.

III. Impulse Response Estimation

Different from any second-order statistical channel estimator, our proposed approach attempts to directly calculate impulse responses from the fact that the microphone array receives the direct-path signal first followed by the other attenuated and delayed copies from reflection paths. The delay-and-sum beamformer result is utilized to roughly estimate the source signal, then every channel's received signal is deconvolved with this estimate to acquire mutual information among channels, and finally room impulse responses from source to each microphone are reconstructed.

To take the delay-and-sum beamformer result as a rough source estimation, we must synchronize signals from the direct-path, based on knowledge of the direct sound wave's direction of arrival (DOA). A method for DOA estimation for broadband coherent waves is the frequency-smoothing MUSIC algorithm, presented in [9]. In the following context, we assume signals from direct-path have already been aligned, and we also assume signals are pre-amplified to compensate for the sampling system impulse response and for the signal energy loss in direct-path transmission. Thus,

$$t_{11}=t_{21}=\cdots=t_{N1}=T, \tag{3}$$

and $\alpha_1=\alpha_2=\cdots=\alpha_N=\alpha$. (4)

The transfer functions then become

$$\begin{aligned} G_1 &= \alpha e^{-j\omega T}+\beta_1 e^{-j\omega t_{12}}+\gamma_1 e^{-j\omega t_{13}}+\cdots \\ G_2 &= \alpha e^{-j\omega T}+\beta_2 e^{-j\omega t_{22}}+\gamma_2 e^{-j\omega t_{23}}+\cdots \\ &\vdots \\ G_N &= \alpha e^{-j\omega T}+\beta_N e^{-j\omega t_{N2}}+\gamma_N e^{-j\omega t_{N3}}+\cdots \end{aligned} , \tag{5}$$

which can be written as

$$\begin{aligned} G_i &= \alpha e^{-j\omega T}+\beta_i e^{-j\omega t_{i2}}+\gamma_i e^{-j\omega t_{i3}}+\cdots \\ &= \alpha e^{-j\omega T}\left(1+\frac{\beta_i}{\alpha}e^{-j\omega(t_{i2}-T)}+\frac{\gamma_i}{\alpha}e^{-j\omega(t_{i3}-T)}+\cdots\right). \\ i&=1,2,\cdots,N \end{aligned} \tag{6}$$

We remove the factor $\alpha e^{-j\omega T}$ from (6). However, for simplicity, we still use the symbol G_i as a transfer function but it now has a new definition,

$$G_i=1+a_i e^{-j\omega\tau_{i1}}+b_i e^{-j\omega\tau_{i2}}+\cdots,\ i=1,2,\cdots,N, \tag{7}$$

where

$$a_i=\frac{\beta_i}{\alpha},\ b_i=\frac{\gamma_i}{\alpha},\ \cdots,$$

and

$$\tau_{i1}=t_{i2}-T,\ \tau_{i2}=t_{i3}-T,\ \cdots,\ \tau_{i1}<\tau_{i2}<\cdots.$$

By introducing $z=e^{j\omega}$, (7) can be further written as

$$G_i=1+a_i z^{-\tau_{i1}}+b_i z^{-\tau_{i2}}+\cdots,\ i=1,2,\cdots,N. \tag{8}$$

Signals received by an *N*-microphone array are

$$Y_i=XG_i, \tag{9}$$

and their average is

$$\overline{Y}=X\frac{1}{N}\sum_{i=1}^{N}G_i. \tag{10}$$

We seek relations among impulse responses by deconvolving each Y_i with $\overline{Y}$. Let us recall the channel identifiability conditions in section II if condition (2) about source signal X is satisfied, all channel characteristics G_i are contained in received signal Y_i. Thus we have

$$\begin{aligned} D_i &= \frac{Y_i}{\overline{Y}}=\frac{XG_i}{X\frac{1}{N}\sum_{i=1}^{N}G_i}=\frac{G_i}{\frac{1}{N}\sum_{i=1}^{N}G_i} \\ &= \frac{1+a_i z^{-\tau_{i1}}+b_i z^{-\tau_{i2}}+\cdots}{1+\frac{1}{N}\sum_{k=1}^{N}a_k z^{-\tau_{k1}}+\frac{1}{N}\sum_{k=1}^{N}b_k z^{-\tau_{k2}}+\cdots} \end{aligned}. \tag{11}$$

In order to simplify the process of finding $a_i,\ b_i,\ \cdots$ and $\tau_{i1},\ \tau_{i2},\ \cdots$ from D_i, we make a slightly stronger assumption than the channel identifiability requirement (1): no reflections are received by any microphone at the same time,

$$\tau_{ik}\neq\tau_{mn},\ \text{for any } (i,k)\neq(m,n). \tag{12}$$

Then (11) can be expanded to an infinite polynomial by the use of polynomial division deconvolution,

$$D_i=1+\frac{N-1}{N}a_i z^{-\tau_{i1}}-\frac{1}{N}\sum_{k\neq i}a_k z^{-\tau_{k1}}+\cdots. \tag{13}$$

In (13), because $\frac{N-1}{N}a_i z^{-\tau_{i1}}$ is dominant to any of $-\frac{1}{N}a_k z^{-\tau_{k1}}$ when *N* is large, it is possible to find $\frac{N-1}{N}a_i$ by comparing to a threshold θ, and locate the delay $z^{-\tau_{i1}}$. Based on this, we arrive at an impulse response reconstruction algorithm that is shown in TABLE I. We use $G_{rec,i}$ to stand for reconstructed transfer function of the *i*th channel, $\overline{G_{rec}}$ to be an average of $G_{rec,i}$, and R_i to represent residues of removing the deconvolution results based on knowledge of $G_{rec,i}$, from expected deconvolution results D_i.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$R_i = D_i - \frac{G_{rec,i}}{\overline{G_{rec}}}. \tag{14}$$

TABLE I
THE MUTUAL DECONVOLUTION ALGORITHM FOR BLIND IDENTIFICATION OF A REVERBERANT SYSTEM

Parameters:

$G_{rec,i}$, θ

Initialization:

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i,$$

$$D_i = \frac{Y_i}{\overline{Y}},\ i = 1, 2, \cdots, N$$

$$G_{rec,i} = 1,\ i = 1, 2, \cdots, N$$

$$\overline{G_{rec}} = \frac{1}{N}\sum_{i=1}^{N} G_{rec,i}$$

$$R_i = D_i - \frac{G_{rec,i}}{\overline{G_{rec}}}$$

Computation:

(a) If $\max(|R_i|) < \theta$, terminate computing. Otherwise,

(b) for $i = 1, 2, \cdots, N$, find the first $\varphi z^{-\lambda}$ that $|\varphi| > \theta$ from left side of polynomial R_i. If $\varphi z^{-\lambda}$ exists, update $G_{rec,i}$ with

$$G_{rec,i} = G_{rec,i} + \frac{N}{N-1}\varphi z^{-\lambda},$$

(c) $\overline{G_{rec}} = \frac{1}{N}\sum_{i=1}^{N} G_{rec,i}$,

(d) $R_i = D_i - \frac{G_{rec,i}}{\overline{G_{rec}}}$, go to (a).

IV. SIMULATION

We implement this algorithm in a discrete-time system with a sample rate of 11025 Hz, and use the Fast Fourier Transform (FFT) and the Inverse Fast Fourier Transform (IFFT) to complete deconvolution. We simulate our mutual deconvolution algorithm for a simple reverberation model considering only the strongest 4 reflections. An 8-microphone array receives the speech signal from the far end.

The reconstructed impulse response from our algorithm, and the results from the MCLMS algorithm are illustrated together in Fig. 1.

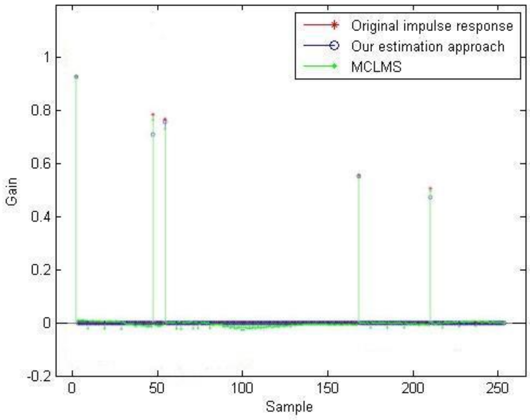
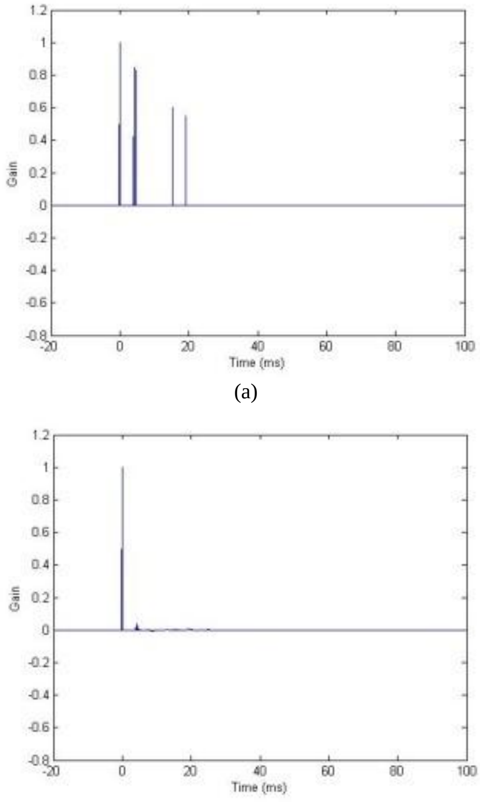


Fig. 1 Reconstructed impulse response

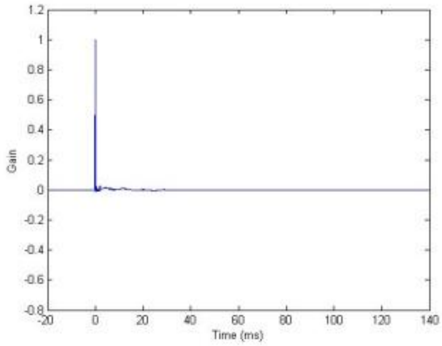
Next, dereverberation using the dereverberant LCMV algorithm is implemented with both impulse responses estimated by our approach and the MCLMS method for comparison. After the signals are recovered, they are deconvolved with the original source signal to show the dereverberation performance. In this evaluation method, a complete dereverberation has a result like a single pulse, representing the direct-path signal. Any other non-zero part in the result indicates the partially dereverberated signal remaining in the output. The less the residual part is, the better the algorithm performs. Fig. 2 shows the evaluation results. We find for simplified acoustic impulse responses, both our proposed mutual deconvolution method and the MCLMS algorithm achieved high-quality dereverberation.



(a)

(b)

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



(c)

Fig. 2 Impulse responses: (a) at array input, (b) at the output of dereverberant LCMV with impulse response from our proposed method, (c) at the output of dereverberant LCMV using the impulse response from the MCLMS filter

In a simulation where the impulse response is closer to the actual situation, the mutual deconvolution algorithm also achieves a reasonably good result. Fig. 3 illustrates a more reverberant impulse response and its estimate from our proposed method with a threshold of 0.2. Delay times of most early reflections are obtained accurately, although errors in their values exist.

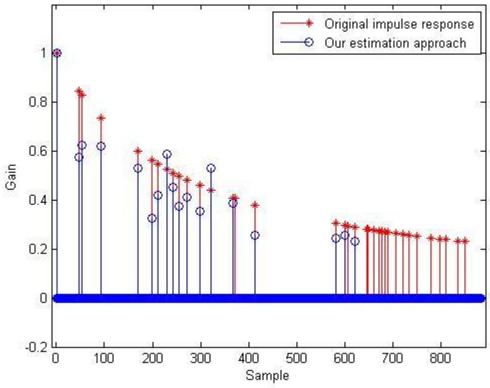
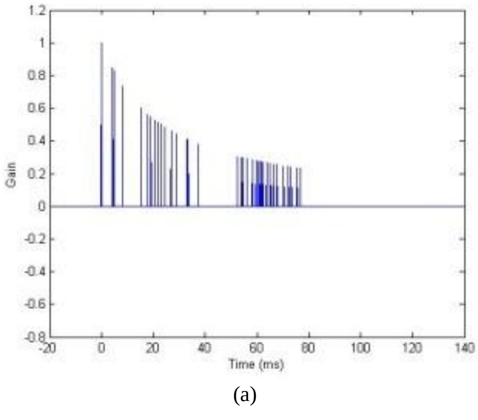
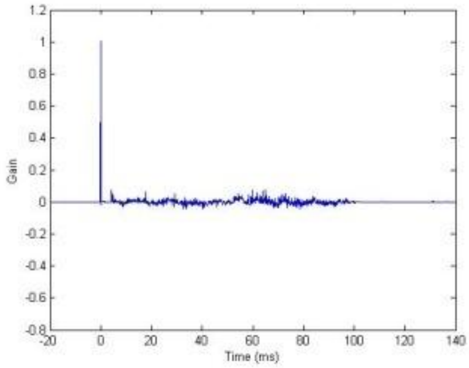


Fig. 3 A more reverberant impulse response and its estimate from mutual deconvolution method

If we take the estimate in Fig. 3 into the dereverberant LCMV algorithm, and deconvolve the output with the original source signal, we can get a graphic view for the output shown in Fig. 4. Although noise exists, it is still a huge improvement from the input. In this case, however, the MCLMS algorithm consumes impractical computing resources because of its long filter settings. The result is far from convergence after 6 hours running of the MCLMS algorithm on a four-core computer.



(a)



(b)

Fig. 4 Impulse responses: (a) at array input, (b) at the output of dereverberant LCMV with impulse response from our proposed method

V. Conclusion

From simulations above, our proposed mutual deconvolution method achieved reasonably high-quality acoustic impulse response estimation. It also shows great advantage in computing efficiency compared to second-order statistical channel identification algorithms. By combining the mutual deconvolution method with dereverberant LCMV algorithm, blind dereverberation is accomplished.

Future work includes eliminating errors from FFT/IFFT deconvolution which generates a finite length series that differs from an infinite length polynomial deconvolution result.

References

[1] D. H. Johnson and D. E. Dudgeon, Array signal processing concepts and techniques, Prentice-Hall, 1993.

[2] J. L. Flanagan, J. D. Johnston, R. Zahn and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1508-1518, 1985.

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

[4] J. Schroder, T. Rohdenburg, V. Hohmann and S. D. Ewert, "Classification of reverberant acoustic situations".

[5] Y. Huang and J. Benesty, "Adaptive blind channel indentifaction: Multichannel and mean square and newton algorithms," in *Proceedings of the IEEE International Conference Acoustics, Speech, Signal Process*, 2002.

[6] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel indentification," *IEEE Trans. on Signal Processing*, vol. 51, pp. 11-24, 2003.

[7] J. Benesty, J. Chen and Y. Huang, Microphone array signal processing, Springer, 2008.

[8] Y. Huang, J. Benesty and G. W. Elko, "Adaptive eigenvalue decomposition algorithm for real-time acoustic source localization system," in *Proceedings of the IEEE International Conference Acoustics, Speech, Signal Process*, 1999.

[9] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," *Proceedings of WASPAA*, pp. 221-224, Oct. 2009.

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)