MAXIMUM LIKELIHOOD APPROACH TO SPEECH ENHANCEMENT FOR NOISY REVERBERANT SIGNALS

Takuya Yoshioka, Tomohiro Nakatani, Takafumi Hikichi, and Masato Miyoshi

NTT Communication Science Laboratories, NTT Corporation
2-4, Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

ABSTRACT

This paper proposes a speech enhancement method for signals contaminated by room reverberation and additive background noise. The following conditions are assumed: (1) The spectral components of speech and noise are statistically independent Gaussian random variables. (2) The convolutive distortion channel is modeled as an auto-regressive system in each frequency bin. (3) The power spectral density of speech is modeled as an all-pole spectrum, while that of noise is assumed to be stationary and given in advance. Under these conditions, the proposed method estimates the parameters of the channel and those of the all-pole speech model based on the maximum likelihood estimation method. Experimental results showed that the proposed method successfully suppressed the reverberation and additive noise from three-second noisy reverberant signals when the reverberation time was 0.5 seconds and the reverberant signal to noise ratio was 10 dB.

Index Terms— Speech enhancement, Reverberation, Additive noise, Maximum likelihood estimation

1. INTRODUCTION

Speech signals captured by microphones in rooms are often corrupted by reverberation and background noise. The distortion system of interest is modeled as Fig. 1. Recovering the clean speech signals from the distorted signals observed at microphones is indispensable for many audio applications working in such realistic environments.

Thus far, it has been a difficult problem to recover clean speech signals from their noisy and reverberated version. Conventional speech dereverberation techniques based on a subspace method are quite sensitive to additive noise [1]. In recent years, we have developed dereverberation methods that are based on statistical models of speech [2, 3]. Those methods calculate maximum likelihood (ML) estimates of convolutive distortion channels, where the likelihood functions are defined based on the speech models. They exhibit robustness against noise with a relatively low noise level. However, recovering the clean speech signals in highly noisy reverberant environments remains a challenging problem [4].

In this paper, we propose a speech enhancement method for signals corrupted by both convolutive distortion channels and additive stationary noise. The proposed method is based on the ML estimation methodology. An observed signal is first analyzed with a short-time Fourier transform (STFT), and the resultant complex spectrogram is then processed. We assume that the spectral components of speech and noise signals are statistically independent Gaussian random variables. The convolutive distortion channel is modeled as an auto-regressive (AR) system in each frequency bin. The power spectral density (PSD) of speech is modeled by an all-pole spectrum. The

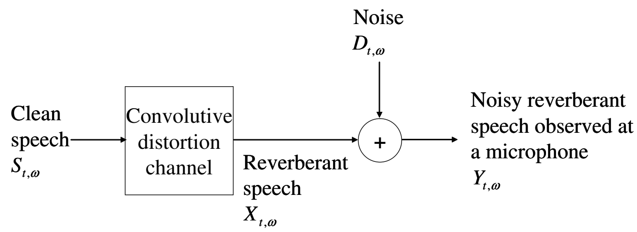

Fig. 1. Distortion system

PSD of noise is assumed to be known in advance. Under these conditions, we can define the likelihood function of the parameters of the channel and those of the all-pole speech model. By maximizing this likelihood function, the proposed method estimates the parameters. The proposed method was tested on signals corrupted by a channel with a reverberation time of 0.5 seconds and by additive noise with a reverberant signal to noise ratio (RSNR) of 10 dB. The proposed method successfully suppressed the reverberation and noise from the observed three-second signals.

2. TASK AND APPROACH

2.1. Task formulation

Let $X_{t,\omega}$ and $D_{t,\omega}$, respectively, denote the complex spectral components of a reverberant speech signal and a noise signal at the t-th time frame and a frequency bin with frequency ω. The spectral component $Y_{t,\omega}$ of an observed signal is given by

$$Y_{t,\omega} = X_{t,\omega} + D_{t,\omega}. \tag{1}$$

Let $S_{t,\omega}$ denote the complex spectral component of a clean speech signal. Provided that a convolutive distortion channel is modeled by an AR system in each frequency bin, the reverberant speech spectral component $X_{t,\omega}$ is then represented as

$$X_{t,\omega} = \sum_{k=1}^{K} g_{k,\omega}^{*} X_{t-k,\omega} + S_{t,\omega}, \tag{2}$$

where $g_{k,\omega}$ is the k-th regression coefficient of the distortion channel at frequency ω, K is the regression order of the distortion channel, and superscript $*$ denotes a complex conjugate operation. Note the following three points as regards the convolutive distortion model (2).

(p1) It is assumed that the convolutive distortion channel has no cross-band components [5].

(p2) The convolutive distortion channel is assumed to be invertible in each frequency bin. By working in the frequency domain, the noninvertibility problem of a non-minimum phase channel in the full band may be mitigated [6]. In fact, the proposed method was tested by using a non-minimum phase channel[1].

(p3) (2) ignores the part of the convolutive distortion that is shorter than the frame size.

The task to be solved is to estimate the clean speech spectrogram $\mathcal{S} = \{S_{t,\omega}\}_{t,\omega}$ from the observed spectrogram $\mathcal{Y} = \{Y_{t,\omega}\}_{t,\omega}$. $\{S_{t,\omega}\}_{t,\omega}$ means the set of $S_{t,\omega}$ over all the frame indices and discrete frequencies. Hereafter, we use the same notation. We also represent the spectrogram of the reverberant speech as $\mathcal{X} = \{X_{t,\omega}\}_{t,\omega}$.

To make the estimation task tractable, we introduce statistical models of the speech and noise spectra as constraints on the speech and noise. Under the abovementioned Gaussian assumptions regarding the speech and noise spectral components, the probability density functions (PDFs) of these components are given by

$$p(S_{t,\omega}|\,{}_s\lambda_t(\omega)) = \frac{1}{\pi\,{}_s\lambda_t(\omega)}\exp\Big\{-\frac{|S_{t,\omega}|^2}{{}_s\lambda_t(\omega)}\Big\} \quad (3)$$

$$p(D_{t,\omega}|\,{}_d\lambda_t(\omega)) = \frac{1}{\pi\,{}_d\lambda_t(\omega)}\exp\Big\{-\frac{|D_{t,\omega}|^2}{{}_d\lambda_t(\omega)}\Big\}, \quad (4)$$

where ${}_s\lambda_t(\omega)$ and ${}_d\lambda_t(\omega)$ are the speech and noise PSDs at the t-th frame, respectively. $S_{t,\omega}$ and $D_{t',\omega'}$ are considered to be statistically independent for any t, ω, t', and ω'. Furthermore, if $(t,\omega) \neq (t',\omega')$, $S_{t,\omega}$ and $S_{t',\omega'}$ (and $D_{t,\omega}$ and $D_{t',\omega'}$) are assumed to be statistically independent, although they are dependent in an exact sense owing to the window function and frame overlap. The speech PSD is, for simplicity, assumed to be of the all-pole form, though more complicated models may be incorporated. Thus, we have

$${}_s\lambda_t(\omega) = \frac{{}_s\sigma_t^2}{|A_t(e^{j\omega})|^2} \quad (5)$$

$$A_t(z) = 1 - a_{t,1}z^{-1} - \cdots - a_{t,P}z^{-P}, \quad (6)$$

where P is the number of poles and $\{a_{t,k},\ {}_s\sigma_t^2\}_k$ are the all-pole model parameters at the t-th frame. The noise PSD is assumed to be known in advance, and might be obtained from intervals without speech.

By putting (1) to (6) together, we have the PDF $p(\mathcal{Y},\mathcal{X}|\Theta)$ of the observed spectrogram $\mathcal{Y}$ and the reverberant speech spectrogram $\mathcal{X}$ given parameter set $\Theta = \{g_{k,\omega}, a_{t,l},\ {}_s\sigma_t^2\}_{k,\omega,l,t}$ as

$$p(\mathcal{Y},\mathcal{X}|\Theta) \propto \Big(\prod_{t,\omega}\frac{|A_t(e^{j\omega})|^2}{{}_d\lambda_t(\omega)\,{}_s\sigma_t^2}\Big)\exp\Big\{-\sum_{t,\omega}\Big(\frac{|Y_{t,\omega}-X_{t,\omega}|^2}{{}_d\lambda_t(\omega)} + \frac{|A_t(e^{j\omega})|^2|X_{t,\omega}-\sum_{k=1}^{K}g_{k,\omega}^*X_{t-k,\omega}|^2}{{}_s\sigma_t^2}\Big)\Big\}. \quad (7)$$

We can also obtain the PDF $p(\mathcal{Y}|\Theta)$ of the observed spectrogram $\mathcal{Y}$ given Θ by integrating out $\mathcal{X}$ from $p(\mathcal{Y},\mathcal{X}|\Theta)$.

2.2. Maximum likelihood estimation approach

As shown later, we can derive the posterior distribution $p(\mathcal{X}|\mathcal{Y},\Theta)$ of the reverberant speech spectrogram $\mathcal{X}$ conditioned on observed spectrogram $\mathcal{Y}$ and the parameter set Θ. Therefore, if the true values of Θ were known, clean speech spectrogram $\mathcal{S}$ could then be estimated by using this distribution based on (2). At this point, our interest is now directed at estimating the unknown parameter set Θ.

To estimate Θ, we employ the maximum likelihood (ML) estimation method. With this method, once the spectrogram $\mathcal{Y}$ of the noisy reverberant speech is observed, $\hat{\Theta}$ that maximizes the log likelihood function $\log p(\mathcal{Y}|\Theta)$ is calculated. In the next section, we specifically derive the ML estimator of Θ.

3. DERIVATION OF MAXIMUM LIKELIHOOD ESTIMATOR

3.1. Expectation maximization algorithm

To maximize the log likelihood function $\log p(\mathcal{Y}|\Theta)$, we employ the expectation maximization (EM) algorithm because the model (7) involves the latent variable $\mathcal{S}$. The EM algorithm iteratively maximizes the expected log likelihood function of complete data $\{\mathcal{Y},\mathcal{X}\}$ (often called an auxiliary function) as follows.

E-step Calculate $p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})$, i.e. the conditional posterior of the reverberant speech spectrogram, where $\Theta^{(i)}$ is the estimate of Θ in the i-th iteration.

M-step Update the estimate of Θ by $\Theta^{(i+1)} = \mathrm{argmax}_\Theta\, Q(\Theta|\Theta^{(i)})$,

where

$$Q(\Theta|\Theta^{(i)}) = \langle \log p(\mathcal{Y},\mathcal{X}|\Theta)\rangle_{p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})}. \quad (8)$$

Below we embody the E-step and M-step, and then summarize and discuss the algorithm of the proposed method.

3.2. Conditional posterior of reverberant speech spectrogram

In the E-step, we calculate the conditional posterior $p(\mathcal{X}|\mathcal{Y},\Theta)$ of the reverberant speech spectrogram. Here, we only show the derived $p(\mathcal{X}|\mathcal{Y},\Theta)$ owing to limited space. Let us define $\boldsymbol{X}_\omega$ and $\boldsymbol{Y}_\omega$ as $\boldsymbol{X}_\omega = [X_{T,\omega}^*,\cdots,X_{1,\omega}^*]^H$ and $\boldsymbol{Y}_\omega = [Y_{T,\omega}^*,\cdots,Y_{1,\omega}^*]^H$, respectively, where T is the number of time frames and superscript H denotes a conjugate transpose operation. Then, we have

$$p(\mathcal{X}|\mathcal{Y},\Theta) = \prod_\omega \mathcal{N}_{\mathbb{C}}\{\boldsymbol{X}_\omega;\boldsymbol{\mu}_\omega,\Sigma_\omega\} \quad (9)$$

$$\boldsymbol{\mu}_\omega = (B_\omega B_\omega^H + G_\omega A_\omega A_\omega^H G_\omega^H)^{-1}(B_\omega B_\omega^H)\boldsymbol{Y}_\omega \quad (10)$$

$$\Sigma_\omega = (B_\omega B_\omega^H + G_\omega A_\omega A_\omega^H G_\omega^H)^{-1}, \quad (11)$$

where $\mathcal{N}_{\mathbb{C}}\{\boldsymbol{x};\boldsymbol{\mu},\Sigma\}$ denotes the complex Gaussian PDF of random variable $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix Σ. In (11), G_ω is a T-dimensional Toeplitz matrix having $[1,-g_{1,\omega}^*,\cdots,-g_{K,\omega}^*,0,\cdots,0]^H$ as its first column and $[1,0,\cdots,0]$ as its first row. A_ω and B_ω are T-dimensional diagonal matrices whose t-th diagonals are $\sqrt{1/\,{}_s\lambda_{T-t+1}(\omega)}$ and $\sqrt{1/\,{}_d\lambda_{T-t+1}(\omega)}$, respectively.

(11) indicates that we can calculate $\boldsymbol{X}_\omega$ independently for each frequency and that the mean, or MMSE estimate, of $\boldsymbol{X}_\omega$ is the output of a Wiener filter applied to $\boldsymbol{Y}_\omega$. In particular, if G_ω is an identity matrix (i.e. a reverberation-free condition), (11) reduces to $X_{t,\omega} = \{{}_s\lambda_t(\omega)/({}_s\lambda_t(\omega) + {}_d\lambda_t(\omega))\}Y_{t,\omega}$, which corresponds to the well-known Wiener-filter-based noise suppression [8].

[1]For simplicity, we limit the derivation to a single-microphone case. However, the generalization of the proposed method to a multi-microphone case, where the channel is always invertible [7], is strainghtforward.

3.3. Maximization of auxiliary function

The M-step maximizes the auxiliary function $Q(\Theta|\Theta^{(i)})$ defined by (8). The auxiliary function is represented as

$$Q(\Theta|\Theta^{(i)}) = \sum_{t=1}^{T}\sum_{\omega}\Big\langle -\log({}_s\sigma_t^2) - \frac{|A_t(e^{j\omega})|^2|X_{t,\omega} - \sum_{k=1}^{K} g_{k,\omega}^* X_{t-k,\omega}|^2}{{}_s\sigma_t^2}\Big\rangle_{p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})}. \quad (12)$$

To derive a condition that $\boldsymbol{g}_\omega = [g_{1,\omega}^*, \cdots, g_{K,\omega}^*]^H$ maximizing (12) must satisfy, let us take the gradient of (12) with respect to $\boldsymbol{g}_\omega$ and set it at zero. See [9] for the gradient of a function of complex variables. Organizing the resultant equation leads to

$${}_xR_\omega\, \boldsymbol{g}_\omega = {}_x\boldsymbol{r}_\omega\,. \quad (13)$$

In (13), ${}_xR_\omega$ and ${}_x\boldsymbol{r}_\omega$ are defined as

$${}_xR_\omega = \sum_{t=1}^{T}\frac{|A_t(e^{j\omega})|^2}{{}_s\sigma_t^2}\langle \boldsymbol{X}_{t-1,\omega}\boldsymbol{X}_{t-1,\omega}^H\rangle_{p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})} \quad (14)$$

$${}_x\boldsymbol{r}_\omega = \sum_{t=1}^{T}\frac{|A_t(e^{j\omega})|^2}{{}_s\sigma_t^2}\langle \boldsymbol{X}_{t-1,\omega}X_{t,\omega}^*\rangle_{p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})}, \quad (15)$$

where $\boldsymbol{X}_{t,\omega} = [X_{t,\omega}^*, \cdots, X_{t-K+1,\omega}^*]^H$.

On the other hand, by taking the gradient of (12) with respect to $\boldsymbol{a}_t = [a_{t,1}, \cdots, a_{t,P}]^H$ and setting it at zero, we obtain the condition for $\boldsymbol{a}_t$ as

$${}_sR_t\, \boldsymbol{a}_t = {}_s\boldsymbol{r}_t\,. \quad (16)$$

In (16), ${}_sR_t$ and ${}_s\boldsymbol{r}_t$ are defined as

$${}_sR_t = [{}_sr_t(l-m)]_{1\le l,m\le P} \quad (P\times P \text{ matrix}) \quad (17)$$

$${}_s\boldsymbol{r}_t = [{}_sr_t(1), \cdots, {}_sr_t(P)]^H \quad (18)$$

$${}_sr_t(k) = \frac{1}{N}\sum_{\omega}\langle |S_{t,\omega}|^2\rangle_{p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})}e^{j\omega k}, \quad (19)$$

where $S_{t,\omega}$ is given by (2) and N is the frame size. Also, the condition for ${}_s\sigma_t^2$ can be easily derived as

$${}_s\sigma_t^2 = \sum_{\omega}|A_t(e^{j\omega})|^2\langle |S_{t,\omega}|^2\rangle_{p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})}. \quad (20)$$

Note that $S_{t,\omega}$, and hence, $\boldsymbol{a}_t$ and ${}_s\sigma_t^2$ are dependent on $\boldsymbol{g}_\omega$.

We find that the solutions for (13) and for the simultaneous equation of (16) and (20) are mutually dependent. Hence, there is no closed-form solution for the simultaneous equation of (13), (16) and (20). We therefore propose updating the estimates of $\{\boldsymbol{g}_\omega\}_\omega$ and $\{\boldsymbol{a}_t, {}_s\sigma_t^2\}_t$ by using the following rules:

$$\boldsymbol{a}_t^{(i+1)} = {}_sR_t^{(i)\,-1}\,{}_s\boldsymbol{r}_t^{(i)} \quad (21)$$

$${}_s\sigma_t^{2\,(i+1)} = \sum_{\omega}|A_t(e^{j\omega})^{(i+1)}|^2\langle |S_{t,\omega}^{(i)}|^2\rangle_{p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})} \quad (22)$$

$$\boldsymbol{g}_\omega^{(i+1)} = {}_xR_\omega^{(i+1)\,-1}\,{}_x\boldsymbol{r}_\omega^{(i+1)}\,. \quad (23)$$

In (21) to (23), ${}_sR_t^{(i)}$, ${}_s\boldsymbol{r}_t^{(i)}$, $A_t(e^{j\omega})^{(i)}$, $S_{t,\omega}^{(i)}$, ${}_xR_\omega^{(i)}$, and ${}_x\boldsymbol{r}_\omega^{(i)}$ are calculated based on (17), (18), (6), (2), (14), and (15), respectively, where the values of Θ are set at $\Theta^{(i)}$. This variant of the EM algorithm, where the M-step is replaced by several conditional maximization steps, is called the expectation conditional maximization (ECM) algorithm [10]. The ECM algorithm is proven to converge to a local optimum.

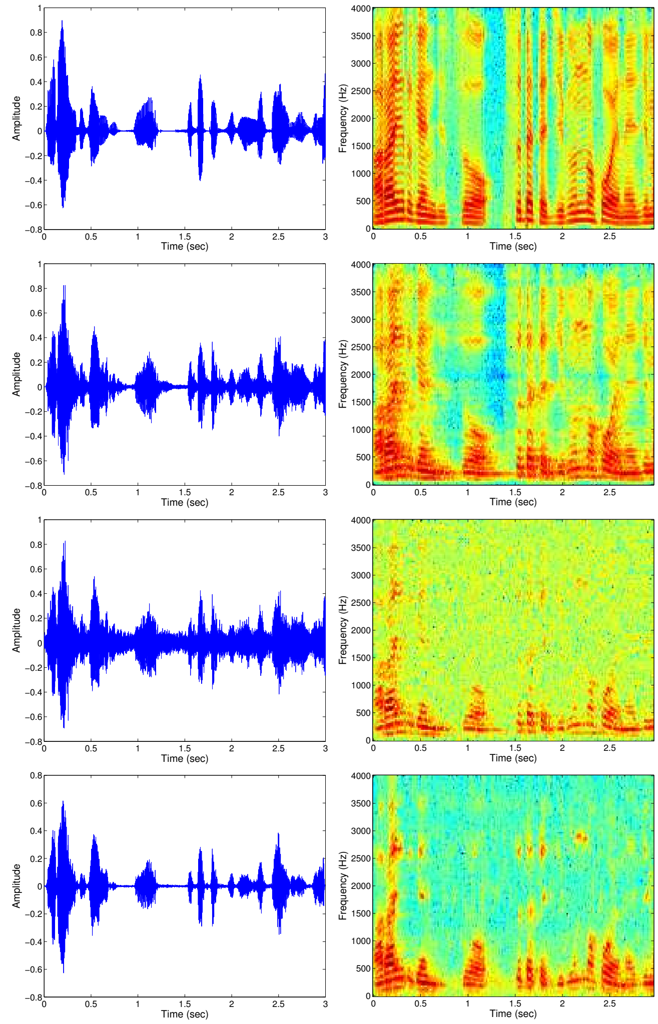

Fig. 2. Waveforms (left panels) and spectrograms (right panels) of, from top to bottom, clean speech, reverberant speech, noisy reverberant speech, and recovered speech.

3.4. Summary and discussion

To summarize Sections 3.1 to 3.3, the algorithm for estimating the parameter set Θ is drawn as follows.

(s1) Set initial parameter estimate $\Theta^{(0)}$ and iteration index i at 0.

(s2) Calculate the means and covariances that specify $p(\mathcal{X}|\mathcal{Y},\Theta^{(i)})$ of (11).

(s3) Calculate $\{\boldsymbol{a}_t^{(i+1)}, {}_s\sigma_t^{2\,(i+1)}\}_t$ and $\{\boldsymbol{g}_\omega^{(i+1)}\}_\omega$ by using (21) to (23).

(s4) Increment i and return to (s2) unless convergence is reached.

It is noteworthy that the proposed method is an extension of the noise suppression method based on the all-pole speech model [11] as well as that of the dereverberation method proposed in [2]. In fact, if we ignore the convolutive distortion channel by setting its order K at 0, the proposed method becomes identical to the noise suppression method described in [11]. On the other hand, if we ignore the

Table 1. SASNR improvement in dB.

	Condition (○: enabled, ×: disabled)		
Noise suppression	○	×	○
Dereverberation	×	○	○
Male	4.25	1.80	7.77
Female	4.67	1.17	7.67

additive noise by setting the noise variance ${}_d\lambda_t(\omega)$ at 0 for any ω, the proposed method is reduced to the frequency domain version of the dereverberation method described in [2].

4. EXPERIMENTAL RESULTS

We conducted experiments to evaluate the performance of the proposed method. Japanese utterances of 10 males and 10 females were taken from the JNAS database. All the utterances were three seconds long. The signals were sampled at 8 kHz and quantized at 16-bit resolution. The speech signals of these utterances were convolved with a non-minimum phase room impulse response measured in a room with a reverberation time of 0.5 seconds to synthesize reverberant signals. The reverberant signals were then contaminated by additive white Gaussian noises with a reverberant signal to noise ratio (RSNR) of 10 dB. The system paramters were set as follows: the STFT frame size was 256 samples, the STFT frame shift was 128 samples, the window function was a Hanning window, the order of the convolutive distortion channel K was 30 for all frequency bins, the assumed number of poles of the speech signals P was 12, and the number of iteratons for the EM algorithm was 5. The proposed method was implemented as a Matlab program. Each iteration was performed with a real time factor of 0.67 on a Linux PC equipped with a 3.6 GHz Pentium 4 processor.

The speech enhancement performance was evaluated by using a segmental amplitude signal to noise ratio (SASNR) defined as

$$\text{SASNR} = \frac{1}{T}\sum_{t=1}^{T} 10\log_{10}\frac{\sum_\omega |S_{t,\omega}|^2}{\sum_\omega ||\hat{S}_{t,\omega}| - |S_{t,\omega}||^2} \quad \text{(dB)}. \qquad (24)$$

We first provide an example of the experimental results. Fig. 2 shows the waveforms and spectrograms of clean speech, reverberant speech, noisy reverberant speech, and recovered speech. It can be seen that the proposed method successfully suppressed both the reverberant components and the noise, although the speech spectral components were underestimated in the higher frequency band, where the signal to noise ratios were very low.

Table 1 summarizes the improvement in SASNR for each gender. The proposed method improved the SASNRs by an average of 7.72 dB. When the dereverberation function is disabled by setting the assumed convolutive distortion channel order K at 0, the average improvement in the SASNRs was 4.46 dB. On the other hand, when the noise suppression function is disabled by setting ${}_d\lambda_t(\omega)$ at 0, the average improvement of the SASNRs fell down to 1.49 dB. Therefore, it can be concluded that the proposed method could enhance the quality of the noisy reverberant speech signals by integrating the dereverberation and noise suppression.

The residual distortion is attributed partly to the part of the convolutive distortion shorter than the frame size, which the proposed method ignores. The inaccuracy of the convolutive distortion model (2) may be another reason for the residual distortion.

5. CONCLUSION

In this paper, we described an ML estimation method for the regression coefficients of a convolutive distortion channel in each frequency bin and the parameters of an all-pole speech model. An EM algorithm was employed, in which the estimates of all the parameters and those of the reverberant speech spectrograms are updated alternately. The derived algorithm is an extension of the noise suppression method described in [8] and the speech dereverberation method reported in [2], both of which are based on the all-pole speech model. Experimental results showed that the proposed method could suppress the reverberation and additive noise from three-second noisy reverberant signals.

Future work includes adaptive estimation of the parameters of the channel and the speech model in order to cope with a time-varying convolutive distortion channel. Adaptive estimation of the noise PSD should also be addressed.

6. REFERENCES

[1] T. Hikichi, M. Delcroix, and M. Miyoshi, "Blind dereverberation based on estimates of signal transmission channels without precise information on channel order," in *Proc. Int'l Conf. Acoust., Speech, Signal Process.*, 2005, pp. 1069–1072.

[2] T. Yoshioka, T. Hikichi, and M. Miyoshi, "Dereverberation by using time-variant nature of speech production system," *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007, Article ID 65698, 15 pages, doi:10.1155/2007/65698.

[3] T. Nakatani, B. H. Juang, T. Yoshioka, K. Kinoshita, and M. Miyoshi, "Importance of energy and spectral features in Gaussian source model for speech dereverberation," in *Proc. IEEE Worksh. Appl. Signal Process. Audio, Acoust.*, 2007, pp. 299–302.

[4] K. Kinoshita, T. Nakatani, M. Delcroix, and M. Miyoshi, "Multi-step linear prediction based speech dereverberation in noisy reverberant environment," in *Proc. Interspeech*, 2007, pp. 854–857.

[5] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, 2007.

[6] H. Wang and F. Itakura, "Dereverberation of speech signals based on sub-band envelope estimation," *IEICE Trans. Fund.*, vol. E74-A, no. 11, pp. 3576–3583, 1991.

[7] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, 1988.

[8] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979.

[9] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, fourth edition, 2001.

[10] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ECM algorithm: a general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

[11] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 26, no. 3, pp. 197–210, 1978.