VARIATIONAL BAYES STATE SPACE MODEL FOR ACOUSTIC ECHO REDUCTION AND DEREVERBERATION

Masahito Togami

Central Research Laboratory, Hitachi Ltd.
1-280, Higashi-koigakubo Kokubunji-shi, Tokyo 185-8601, Japan

ABSTRACT

In this paper, we propose a simultaneous optimization technique for speech dereverberation, acoustic echo reduction, and noise reduction, which can be utilized even when an analog-to-digital (A/D) converter and a digital-to-analog (D/A) converter are not synchronized. The proposed method utilizes a state-space model in which acoustic echo reduction filters are regarded as a time-varying state-vector due to asynchrony of the A/D converter and the D/A converter. In addition to the state-space model for acoustic echo reduction filters, the proposed method utilizes an additional state-space model in which noiseless multichannel speech signals are regarded as a state vector. By using the second state-space model, we can update the dereverberation filter under noisy environments. To optimize two types of state space models, the proposed method utilizes the variational Bayes framework. Two Kalman smoother based parameter optimization stages are performed alternatively. The proposed method is evaluated by using recorded data in a real teleconferencing room. The experimental results show that the proposed method can reduce acoustic echo signal, speech reverberation, and background noise more effectively than the conventional method by authors even when the A/D converter and the D/A converter are asynchronous.

Index Terms— Kalman smoother, dereverberation, echo reduction, variational Bayes, asynchronous

1. INTRODUCTION

Noise reduction, dereverberation, and acoustic echo cancellation are highly required for teleconferencing systems. To reduce reverberation, autoregressive model based dereverberation techniques [1][2] have been actively studied. To reduce background noise, single channel noise reduction techniques [3][4] and multichannel beamformer [5] have been studied. For acoustic echo cancellation, least square algorithms based on adaptive filters [7] have been commonly utilized. However, these functions focus on reduction of specific type of unwanted signals. When several unwanted signals are required to be reduced, cascade methods of several methods have been utilized. However, cascade methods suffer from mutual interference problems between the methods, which causes degradation of eventual speech enhancement performance. A joint optimization of non-linear echo reduction and non-linear late reverberation reduction is proposed [6]. However, linear filtering for dereverberation is not integrated. In the previously proposed method by the authors [8], we proposed a simultaneous optimization technique of acoustic echo cancellation, dereverberation, and noise reduction, in which the probability density function (PDF) of the acoustic echo path is assumed to be a time-invariant Gaussian distribution. The proposed method can reduce undesired signals effectively when an analog-to-digital (A/D) converter and a digital-to-analog (D/A) converter are synchronized. However, when the A/D converter and the D/A converter are unsynchronized, speech enhancement performance will be degraded. In this context, the authors proposed a multichannel acoustic echo reduction technique which tracks the sampling mismatch between the A/D converter and the D/A converter by using a state space model of the acoustic echo path [9]. In addition to the acoustic echo reduction, multichannel Wiener filtering is integrated in this method, and optimization of parameters are done simultaneously so as to increase the likelihood function by using the Kalman smoother based parameter optimization technique [10]. This method is shown to reduce acoustic echo and background noise effectively. However, speech dereverberation is not integrated into this method. On the other hand, authors also proposed a noise robust speech dereverberation technique which utilizes an autoregressive model of noiseless microphone input signals [11]. However, this method cannot reduce acoustic echo signal.

In this paper, we propose a simultaneous optimization technique which can reduce speech reverberation, acoustic echo signal, and background noise signal even when the A/D converter and the D/A converter are not synchronized. The proposed method can be interpreted as combination of the state-space model based acoustic echo reduction technique [9] and the state-space model based dereverberation technique [11]. The proposed method is not a cascade method which utilizes individual cost function for each function, but also one of simultaneous optimization techniques in which a unified likelihood function is utilized. The proposed method utilizes two types of state-space models. The first one is a state-space model for acoustic echo path. This model is close to the previous proposed Kalman smoother based acoustic echo reduction [9]. However, the previous model assumes instantaneous mixture for near-end speech sources. The proposed method utilizes a convolutive mixture model for near-end speech sources. The second state-space model is a state-space model for convolutive near-end speech sources. We utilize a noiseless autoregressive model for reverberant speech sources. To optimize these two state-space models, we utilize a variational-bayes state-space model framework. Experimental results show that the proposed method can reduce acoustic echo, reverberation, and background noise signal effectively even when the A/D converter and the D/A converter are unsynchronized.

2. PROBLEM STATEMENT

2.1. Input signal model

In this paper, we defines the speech enhancement problem at the time-frequency comain. The microphone input signal with multiple microphones, $\boldsymbol{x}_{l,k}$ (l is frame index, k is frequency index), is

modeled as $\boldsymbol{x}_{l,k} = [\ x_{l,k,1} \quad \dots \quad x_{l,k,N_m}\ ]^T$, where N_m is the number of the microphones and T is the transpose operator of a matrix/vector. Under the assumption that there are one speech source, far-end acoustic echo signal, and background noise signal, the microphone input signal is modeled as follows:

$$\boldsymbol{x}_{l,k} = \boldsymbol{G}_{l,k}\boldsymbol{d}_{l,k} + \boldsymbol{c}_{l,k} + \boldsymbol{w}_{l,k}, \tag{1}$$

where $\boldsymbol{w}_{l,k}$ is the multichannel noise signal. In the proposed method, each frequency bin is processed independently. Therefore, the frequency index k is omitted. The state-space model for the acoustic echo path can be modeled as follows:

$$\boldsymbol{g}_l = \boldsymbol{g}_{l-1} + \boldsymbol{r}_l, \tag{2}$$

where $\boldsymbol{r}_l$ is amount of change of the acoustic echo path. The far-end speech signal $\boldsymbol{d}_l$ is defined as $\boldsymbol{d}_l = [\ d_l \quad \dots \quad d_{l-L_d+1}\ ]^T$. $\boldsymbol{c}_l$ is a noiseless reverberant speech signal, which is defined as follows:

$$\boldsymbol{c}_l = \sum_{t=0}^{L_{\text{imp}}-1} \boldsymbol{h}_t s_{l-t}, \tag{3}$$

where L_{imp} is the length of the impulse response, $\boldsymbol{h}_t$ is a vector which is composed of multichannel impulse responses which is defined as $\boldsymbol{h}_t = [\ h_{1,t} \quad \dots \quad h_{N_m,t}\ ]^T$, where $h_{k,m,t}$ is the tth tap of the impulse response between the speech source and the mth microphone, and $s_{l,k}$ is the original signal.

The noiseless reverberant speech signal $\boldsymbol{c}_l$ can be converted into an auto-regressive model as follows:

$$\boldsymbol{c}_l = \sum_{l'=D}^{L_w-1} \boldsymbol{W}_{l'}\boldsymbol{c}_{l-l'} + \boldsymbol{e}_l, \tag{4}$$

where L_w is the length of the autoregressive coefficients, D is the tap-length of the early reflection, and $\boldsymbol{e}_l$ is the multichannel signal which is composed of direct-path and early reflection of the near-end speech signal. $\boldsymbol{e}_l$ is defined as $\boldsymbol{e}_l = \sum_{l'=0}^{D-1} s_{l-l'}\boldsymbol{h}_{l'}$. Furthermore, the noiseless reverberant speech signal $\boldsymbol{e}_l$ can be transformed into a 1st order Markov model as follows:

$$\boldsymbol{f}_l = \boldsymbol{A}\boldsymbol{f}_{l-1} + \boldsymbol{u}_l, \tag{5}$$

where $\boldsymbol{f}_l = [\ \boldsymbol{c}_l^H \quad \boldsymbol{c}_{l-1}^H \quad \boldsymbol{c}_{l-L_w+2}^H\ ]^H$, H is a Hermite transpose of a matrix/vector, $\boldsymbol{u}_l = [\ \boldsymbol{e}_l^T \quad \boldsymbol{0} \quad \boldsymbol{0}\ ]^T$,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{0}_{N_m\times N_m(D-1)} & \boldsymbol{W}_D & \dots & \boldsymbol{W}_{L_w-1} \\ \boldsymbol{I}_{N_m\times N_m} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0}_{N_m\times N_m} & \boldsymbol{I}_{N_m\times N_m} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0}_{N_m\times N_m} & \boldsymbol{0}_{N_m\times N_m} & \boldsymbol{I}_{N_m\times N_m} & \boldsymbol{0} \\ & \vdots & & \end{bmatrix}. \tag{6}$$

In the standard state-space model, the state vectors converted into one state vector $[\ \boldsymbol{g}_l^T \quad \boldsymbol{f}_l^T\ ]^T$. However, it is infeasible because the computational cost of the Kalman smoother is proportional to square of the number of the state vector. Therefore, the state-space models are summarized as follows:

State-transition equations

$$\boldsymbol{g}_l = \boldsymbol{g}_{l-1} + \boldsymbol{r}_l, \tag{7}$$
$$\boldsymbol{f}_l = \boldsymbol{A}\boldsymbol{f}_{l-1} + \boldsymbol{u}_l. \tag{8}$$

The observation equation is defined as follows:

$$\boldsymbol{x}_l = \boldsymbol{D}_l\boldsymbol{g}_l + \boldsymbol{J}\boldsymbol{f}_l + \boldsymbol{w}_l, \tag{9}$$

where $\boldsymbol{J} = \begin{bmatrix} \boldsymbol{I}_{N_m\times N_m} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$.

$$\boldsymbol{D}_l = \Big[\ d_l\boldsymbol{I}_{N_m\times N_m} \quad d_{l-1}\boldsymbol{I}_{N_m\times N_m} \quad \dots \quad d_{l-L_d+1}\boldsymbol{I}_{N_m\times N_m}\ \Big], \tag{10}$$

where $\boldsymbol{I}$ is an identity matrix.

The output signal after dereverberation, acoustic echo reduction, and background noise reduction is a part of the MMSE estimate of the second state transition equation:$\hat{\boldsymbol{e}}_{\text{MMSE},l} = E[\boldsymbol{e}_l|\mathcal{X},\mathcal{D}]$, where $\mathcal{X} = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_{L_T}\}$, $\mathcal{D} = \{\boldsymbol{d}_1,\dots,\boldsymbol{d}_{L_T}\}$, and L_T is the number of the frames.

3. PROPOSED METHOD

3.1. Probabilistic models

In the proposed method, we utilize a local Gaussian model [12] in which the probability density function (PDF) of the near-end speech source is set to the time-varying Gaussian distribution as $p(\boldsymbol{e}_l) = \mathcal{N}(\boldsymbol{0}, v_l\boldsymbol{R}_e)$, where v_l is the time-varying variance of the speech source signal and $\boldsymbol{R}_e$ is the covariance matrix of the steering vector of the near-end speech source. The state-transition noise $\boldsymbol{r}_l$ is defined as stationary Gaussian distribution as $\mathcal{N}(\boldsymbol{0}, \sigma_r\boldsymbol{I})$. The PDF of the observation noise is $p(\boldsymbol{w}_l) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_w)$.

3.2. Summary of proposed method

In the proposed method, parameters with related to the proposed state space model are estimated so as to maximize the likelihood function. However, EM (Expectation-Maximization) algorithm for the standard state-space model [10] cannot be utilized for the state-space model with two state transition equations. Instead of the EM based optimization method, we utilize a variational bayes approximiation. The Q function can be obtained as follows:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \int_{\mathcal{G},\mathcal{F}} q(\mathcal{G}|\boldsymbol{\theta}^{(t)})q(\mathcal{F}|\boldsymbol{\theta}^{(t)}) \log \frac{p(\mathcal{X},\mathcal{G},\mathcal{F}|\boldsymbol{\theta})}{q(\mathcal{G}|\boldsymbol{\theta}^{(t)})q(\mathcal{F}|\boldsymbol{\theta}^{(t)})} d\mathcal{G}d\mathcal{F}, \tag{11}$$

where $\boldsymbol{\theta}$ is the parameter, which includes $\{v_l\}_{1\le l\le L_T}$, $\boldsymbol{R}_e$, $\boldsymbol{R}_w$, σ_r, $\{\boldsymbol{W}_{l'}\}_{D\le l'\le L_w-1}$, $\mathcal{G}$ is $\{\boldsymbol{g}_l\}_{1\le l\le L_T}$, and $\mathcal{F}$ is $\{\boldsymbol{f}_l\}_{1\le l\le L_T}$. $\boldsymbol{\theta}^{(t)}$ is estimation of the parameter after the tth iteration. In the variational Bayes framework, the joint probability of $\mathcal{G}$ and $\mathcal{F}$, $p(\mathcal{G},\mathcal{F})$ is approximated as $q(\mathcal{G}|\boldsymbol{\theta}^{(t)})q(\mathcal{F}|\boldsymbol{\theta}^{(t)})$.

3.3. E step

In the E step, approximated probability density functions $q(\mathcal{G}|\boldsymbol{\theta}^{(t)})$, $q(\mathcal{F}|\boldsymbol{\theta}^{(t)})$ are optimized so as to increase the Q function alternately.

3.3.1. Update of $q(\mathcal{G})$

$q(\mathcal{G})$ are updated under the assumption that $q(\mathcal{F})$ is fixed as follows:

$$\log q(\mathcal{G}) = const. + \log p(\mathcal{G}) - \frac{1}{2}\sum_{l=1}^{L_T} \text{tr}\left\{\boldsymbol{R}_w^{-1}(\boldsymbol{x}_l - \boldsymbol{J}\tilde{\boldsymbol{f}}_l - \boldsymbol{D}_l\boldsymbol{g}_l)(\boldsymbol{x}_l - \boldsymbol{J}\tilde{\boldsymbol{f}}_l - \boldsymbol{D}_l\boldsymbol{g}_l)^H\right\}, \tag{12}$$

where H is the Hermite transpose of a matrix/vector and $\tilde{\boldsymbol{f}}_l$ is estimated mean vector in $q(\mathcal{F})$. The PDF of $\mathcal{G}$ can be regarded as a multichannel Gaussian distribution with a state-space model as follows:

State transition equation

$$\boldsymbol{g}_l = \boldsymbol{g}_{l-1} + \boldsymbol{r}_l, \tag{13}$$

Modified observation equation

$$\boldsymbol{x}_l - \boldsymbol{J}\tilde{\boldsymbol{f}}_l = \boldsymbol{D}_l\boldsymbol{g}_l + \boldsymbol{w}_l, \tag{14}$$

where $\boldsymbol{x}_l - \boldsymbol{J}\tilde{\boldsymbol{f}}_l$ is a virtual observed signal. Therefore, $q(\mathcal{G})$ can be calculated by performing the Kalman smoother [13].

3.3.2. Update of $q(\mathcal{F})$

$q(\mathcal{F})$ are updated under the assumption that $q(\mathcal{G})$ is fixed as follows:

$$\begin{aligned}\log q(\mathcal{F}) &= const. + \log p(\mathcal{F}) + -\frac{1}{2}\sum_{l=1}^{L_T} \\ &\times \mathrm{tr}\left\{\boldsymbol{R}_w^{-1}(\boldsymbol{x}_l - \boldsymbol{D}_l\tilde{\boldsymbol{g}}_l - \boldsymbol{J}\boldsymbol{f}_l)(\boldsymbol{x}_l - \boldsymbol{D}_l\tilde{\boldsymbol{g}}_l - \boldsymbol{J}\boldsymbol{f}_l)^H\right\},\end{aligned} \tag{15}$$

where $\tilde{\boldsymbol{g}}_l$ is estimated mean vector in $q(\mathcal{G})$. The PDF of $\mathcal{F}$ can be also regarded as a multichannel Gaussian distribution with a state-space model as follows:

State transition equation

$$\boldsymbol{f}_l = \boldsymbol{A}\boldsymbol{f}_{l-1} + \boldsymbol{u}_l, \tag{16}$$

Modified observation equation

$$\boldsymbol{x}_l - \boldsymbol{D}_l\tilde{\boldsymbol{g}}_l = \boldsymbol{J}\boldsymbol{f}_l + \boldsymbol{w}_l, \tag{17}$$

where $\boldsymbol{x}_l - \boldsymbol{D}_l\tilde{\boldsymbol{g}}_l$ is a virtual observed signal. Therefore, $q(\mathcal{F})$ can be calculated by performing the Kalman smoother.

3.4. M step

Parameters $\boldsymbol{\theta}$ are updated so as to increase the Q function. The time-varying variance of the near-end speech signal, v_l, the covariance matrix of the steering vector $\boldsymbol{R}_e$, and the auto-regressive coefficient $\boldsymbol{W}_{l'}$ are updated via the sufficient statistics of $q(\mathcal{F})$ in a similar way with the previously proposed Kalman smoother based dereverberation technique [11]. σ_r is updated via the sufficient statistics of $q(\mathcal{G})$ in a similar way with the acoustic echo reduction technique [9]. Finally, the covariance matrix of the multichannel observed noise is estimated as follows:

$$\begin{aligned}\boldsymbol{R}_w &= \frac{1}{L_T}\sum_{l=1}^{L_T}\{\boldsymbol{x}_l\boldsymbol{x}_l^H - \boldsymbol{J}\tilde{\boldsymbol{f}}_l\boldsymbol{x}_l^H - \boldsymbol{x}_l\tilde{\boldsymbol{f}}_l^H\boldsymbol{J}^H \\ &- \boldsymbol{D}_l\tilde{\boldsymbol{g}}_l\boldsymbol{x}_l^H - \boldsymbol{x}_l\tilde{\boldsymbol{g}}_l^H\boldsymbol{D}_l^H \\ &+ \boldsymbol{J}\tilde{\boldsymbol{f}}_l\tilde{\boldsymbol{g}}_l^H\boldsymbol{D}_l^H + \boldsymbol{D}_l\tilde{\boldsymbol{g}}_l\tilde{\boldsymbol{f}}_l^H\boldsymbol{J}^H \\ &+ \boldsymbol{J}\boldsymbol{R}_{f,l}\boldsymbol{J}^H + \boldsymbol{D}_l\boldsymbol{R}_{g,l}\boldsymbol{D}_l^H\},\end{aligned} \tag{18}$$

where $\boldsymbol{R}_{f,l}$ is the covariance matrix of $\boldsymbol{f}_l$ which is estimated in the Kalman smoother and $\boldsymbol{R}_{g,l}$ is the covariance matrix of $\boldsymbol{g}_l$. The output signal in which reverberation and acoustic echo signal are reduced is the first N_m elements of the mean vector of $\boldsymbol{u}_l$ that is estimated by $\tilde{\boldsymbol{f}}_l - \boldsymbol{A}\tilde{\boldsymbol{f}}_{l-1}$.

4. EXPERIMENT

4.1. Experimental conditions

Experimental environment and microphone array alignment are shown in Fig. 1. The impulse responses were recorded at Location 1, 2, 3 by using TSP (Time Stretched Pulse) method [14]. We evaluate the acoustic echo reduction and dereverberation performance when an A/D converter and a D/A converter are not synchronized. The far-end speech is played by using the D/A converter attached with the personal computer. Recording of the microphone input signal is done by using the A/D converter which is not synchronized with the D/A converter. Therefore, the acoustic echo path is time-varying. The original source signals of the near-end speech signals and the far-end speech signals are extracted from the TIMIT database [15]. The number of the near-end speech signals and the number of the far-end speech signals are 34 each. The other experimental conditions are shown in Table. 1.

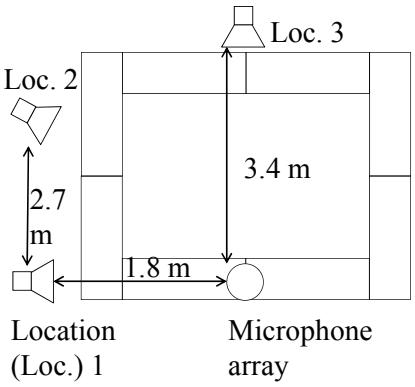

Fig. 1. Experimental environment and microphone array alignment

Table 1. Experimental conditions

Sampling rate	16000 [Hz]
Frame size	1024 [pt]
Frame shift	256 [pt]
Number of microphones N_m	3
L_d	8 [tap]
D	2
L_w	10
Number of EM iterations	10

Signal to Noise ratio (SNR) between the near-end speech and summation of the recorded far-end speech and the background noise signal was set to 0 dB. The evaluation measures are Signal To Distortion Ratio (SDR)[16], Cepstrum Distance (CD) [17], and PESQ [18]. The proposed method (PROPOSED) was compared with previously proposed method which combines speech dereverberation, acoustic echo reduction, and noise reduction by multichannel Wiener filtering with stationary probability density function of the acoustic echo path (TOGAMI_2014) [8] and modification of the proposed method with time-invariant assumption of the variance of the

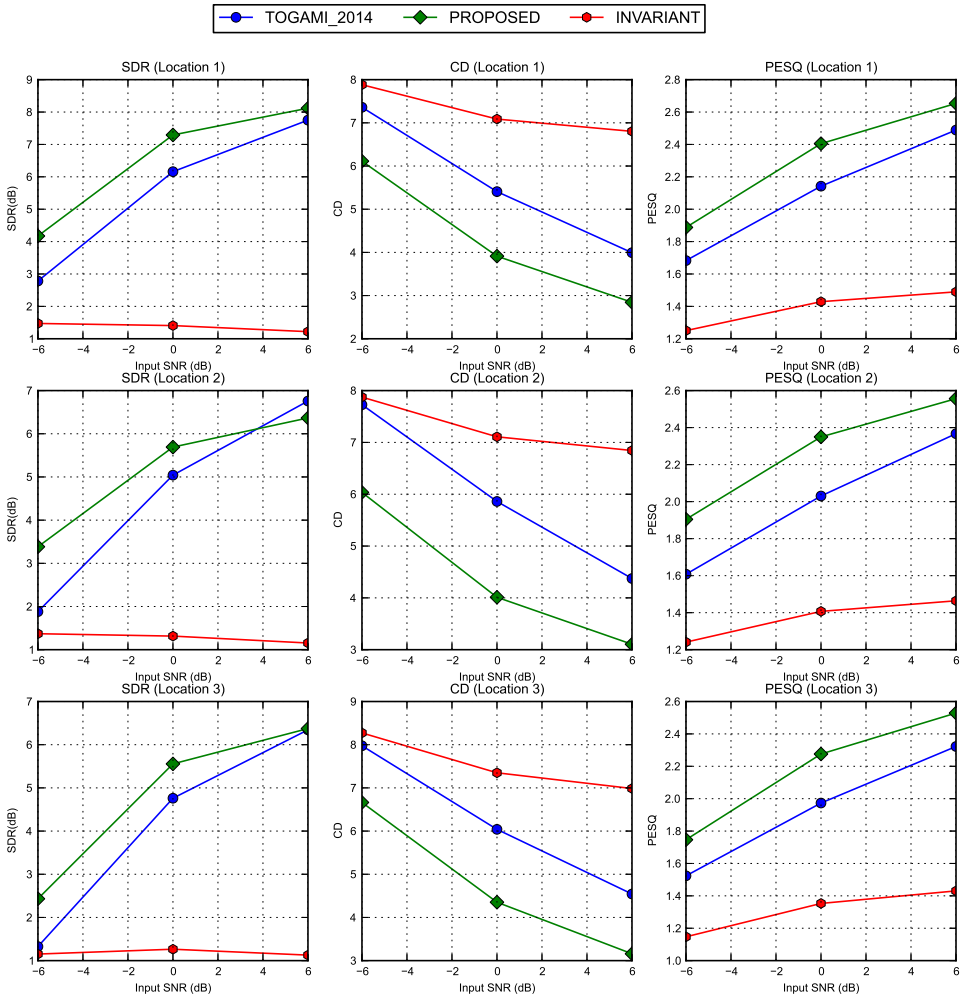

Fig. 2. Experimental results when SNR is 0 dB

near-end speech source (INVARIANT). The experimental results are shown in Fig. 2. In each location, the proposed method is shown to be superior to the conventional methods. Therefore, the proposed combination of acoustic echo reduction and speech dereverberation is effective.

5. RELATION TO PRIOR WORK

The proposed method is an extension of the Kalman smoother based acoustic echo reduction technique [19]. However, this technique is for only acoustic echo reduction. Additionally, the time-varying assumption of speech sources are not fully-utilized.

6. CONCLUSION

In this paper, we propose a speech enhancement technique for acoustic echo reduction, speech dereverberation, and noise reduction. The proposed method utilizes two types of state-space models. Optimization of the parameters based on two state-space models is performed by using variational Bayes framework. The experimental results showed that the proposed method is effective.

7. REFERENCES

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

[2] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-step Linear Prediction," *IEEE Trans. Audio, Speech, and Language Process.,* vol. 17, no. 4, pp. 534–545, May 2009.

[3] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustic., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Feb. 1979.

[4] Y. Ephraim and D. Malah, "Speech enhancement using minimum mean-square error short-time spectral amplitude estimator", *IEEE Trans. Acoustic., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[5] D.H. Johnson and D.E. Dudgeon, "Array signal processing-concepts and Techniques," PTR Prentice Hall, New Jersey, USA, 1993.

[6] E.A.P. Habets, S. Gannot, I. Cohen, and P. Sommen, "Joint Dereverberation and Residual Echo Suppression of Speech Signals in Noisy Environments," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 16, no. 8, pp. 1433–1451, 2008/11.

[7] E. Hänsler, G. Schmidt, "Acoustic Echo and Noise Control: A Practical Approach,", John Wiley & Sons, 2004.

[8] M. Togami, Y. Kawaguchi, "Simultaneous Optimization of Acoustic Echo Reduction, Speech Dereverberation, and Noise Reduction against Mutual Interference," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 22, no. 11, pp. 1612–1623, 2014/11.

[9] M. Togami, Y. Kawaguchi, R. Takashima, " Frequency Domain Acoustic Echo Reduction based on Kalman Smoother with Time-Varying Noise Covariance Matrix," *IEEE ICASSP2014*, pp. 5909–5913, 2014/5.

[10] Z. Ghahramani and G.E. Hinton, "Parameter estimation for linear dynamical systems," Technical Report CRG-TR-96-2, Department of Computer Science, University of Toronto, 1996.

[11] M. Togami and Y. Kawaguchi, "Noise Robust Speech Dereverberation with Kalman Smoother," *Proc. ICASSP2013*, pp. 7447–7451, 2013/5.

[12] N.Q.K. Duong, E. Vincent, R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. Speech Audio Process.*, vol. 18, no. 7, pp. 1830–1840, 2010/9.

[13] A.H. Jazwinski, *Stochastic Processes and Filtering Theory.* Academic Press, 1970.

[14] Y. Suzuki, F. Asano, H.Y. Kim, and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Amer.* vol. 97, no. 2, pp. 1119–1123, Feb. 1995.

[15] TIMIT corpus [Online]. Available: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1.

[16] M. Togami, Y. Kawaguchi, R. Takeda, Y. Obuchi, and N. Nukaga, "Optimized Speech Dereverberation From Probabilistic Perspective for Time Varying Acoustic Transfer Function," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 21, no. 7, pp. 1369–1380, 2013/7.

[17] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. Speech Audio Process.*, vol. 17, no. 7, pp. 1717–1731, Sep. 2010.

[18] ITU-T P.862, "Perceptual evaluation of speech quality: An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech coders," ITU-T, 2001.

[19] C. Paleologu, J. Benesty, S. Ciochina, "Study of the General Kalman Filter for Echo Cancellation," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 21, no. 8, pp. 1539–1549, 2013/8.