COMPUTATIONALLY EFFICIENT SINGLE CHANNEL DEREVERBERATION BASED ON COMPLEMENTARY WIENER FILTER

Kazunobu Kondo[†‡] *Yu Takahashi*[†] *Tatsuya Komatsu*[‡] *Takanori Nishino*[*] *Kazuya Takeda*[‡]

[†]Yamaha Corporation, Hamamatsu, Japan
[‡]Graduate School of Information Science, Nagoya University, Nagoya, Japan
[*]Graduate School of Engineering, Mie University, Tsu, Japan

ABSTRACT

A single-channel dereverberation method with low computational complexity is proposed. We introduce the complementary Wiener filter which can suppress a late reverberation during silence intervals via theoretical analysis and numerical calculation. An implementation represents reductions both of memory consumption and operative calculations compared to a conventional method; each reduction is almost by half. Dereverberation performance is evaluated by an experimental simulation using speech signals and measured room impulse responses. The performance under several hundreds of msec of the reverberation time is similar to the conventional method: 6 [dB] reverberation reduction and 3 [dB] improvement of target-to-interference ratio.

Index Terms— Speech enhancement, Reverberation, Wiener filter, Computational complexity

1. INTRODUCTION

Speech communication and recognition systems are generally used under a reverberant condition such as meeting rooms; generally, a reverberation time is under 1 sec. Speech quality and recognition performance are degraded in the reverberant condition. To cope with this problem, dereverberation techniques have been studied in recent years. Multi-microphone techniques are utilized to estimate a late reverberation based on a spatial correlation in [1–5], or to estimate a inverse filter by MINT theorem [6] in [7, 8]. A multi-microphone technique leads to a large-sized apparatus, therefore, single-channel methods are required; however, a serious drawback of a single-channel method is no spatial information. Some of single-channel speech enhancement techniques provide successful results, such as the spectral subtraction [9], MMSE-STSA [10], etc. These techniques have been applied to dereverberation methods, and they have succeeded such as in [8, 11–13].

In voice terminals, several functions to improve speech quality work concurrently, for example an echo canceler and a noise reduction. Each function should work with low computational complexity because of the concurrent execution. Therefore, more efficient dereverberation methods are always required by industries than previously proposed methods. The Wiener filer (WF) is commonly used to enhance the target signal such as in [14]. If the WF would be described as β and $0 \leq \beta \leq 1$, then $1-\beta$ could be so-called "complementary Wiener filter" (CWF) such as in [15]. We consider the CWF to suppress a late reverberation. It will be shown that the CWF can achieve not only lower computational complexity but also similar dereverberation performance in the following sections.

The work presented here is focused on the formulation to establish a low computational complexity dereverberation method, which takes advantage of performance similar to the conventional method in [11] with lower computational complexity. Approaches to improve dereverberation performance take a considerable complexity, because they utilize log-MMSE-STSA [16] in Chapter 3 of [8], long-term multi-step linear prediction in [12] or latent variable decomposition in [17], etc. To apply these methods to voice terminals, one issue is to achieve low complexity with similar performance, which was not considered in these earlier studies.

The rest of this paper is organized as follows. Section 2 shows a signal model and derivation of a power spectrum model. Section 3 introduces the proposed method with the CWF. Section 4 shows experimental results. Section 5 concludes this paper.

2. SIGNAL MODEL

2.1. Signal model in the time domain

$s(n)$ is a source signal at a sampling index n by a sampling frequency F_s. $x(n)$ is an observation which consists of a convolution between $s(n)$ and a room impulse response (RIR) $h(n)$; $x(n) = s(n) * h(n) = \sum_\mu s(n)h(\mu - n)$. In this paper, we consider a statistical model of the RIR proposed by Polack [18]. The model consists of the zero-mean Gaussian random process $b(n)$ and a reverberant time T_{60} [19]. T_{60} means a time length of an energy decay until -60 [dB]. The RIR $h(n)$ is formulated as $h(n) = b(n)e^{-\Delta n}(n \geq 0)$, and $\Delta = (3\ln 10)/(T_{60}F_s) = (3\ln 10)/N_{60}$, where N_{60} is the sample length which corresponds to T_{60}, and $h(n) = 0(n < 0)$. The spatial expectation $E_h[\cdot]$ of the statistical RIR model is described by the Dirac's delta function with a lag τ as:

$$E_h[h(n)h(n-\tau)] = \sigma_b^2 \delta(\tau) e^{-2\Delta n}, \tag{1}$$

where σ_b^2 is the variance of $b(n)$. Eq. (1) involves a spatial uncorrelation of the RIR model which is described as $E_h[h(n)h(n-\tau)] = 0(\tau \neq 0)$.

Linear prediction is widely used to model speech signals based on the quasi-stationary property, and this fact includes that speech is correlated during a phoneme; in contrast, phonemes are uncorrelated to each other. This is a successful assumption used in other research [11, 12]. Consider an autocorrelation function of the speech signal $R_s^{(m)}(\tau)$ with this assumption as follows:

$$\begin{aligned} R_s^{(m)}(\tau) &= E_n[s(n)s(n-\tau)] \quad |R_s^{(m)}(\tau)| \approx 0 \,(\text{iff } \tau \geq N_E) \\ &= E_n[s(n+mN_E)s(n+mN_E-\tau)], \end{aligned} \tag{2}$$

where $E_n[\cdot]$ means the temporal expectation and m means a frame index. N_E corresponds to the sample length of an early-reflection (ER). The length of the ER is commonly considered as a few

or several tens of msec. In addition, the RIR after the reverberation time T_{60} is assumed to be zero, because the energy is decayed enough without loss of generality. Consequently, the observed signal can be separated into $x(n) \equiv x_E(n) + x_R(n)$ as $x(n) = \sum_{i=0}^{N_E-1} h(i)s(n-i) + \sum_{i=N_E}^{N_{60}} h(i)s(n-i)$, where $x_E(n)$ and $x_R(n)$ are the ER and late reverberant (LR) components of the observed signal.

2.2. Autocorrelation function

In this section, we consider the autocorrelation function of the observed signal. The spatial and temporal expectation $E_{n,h}[\cdot]$ of the observed signal is considered.

2.2.1. Autocorrelation of the early-reflected signal

The statistical model proposed by Polack assumes spatial uncorrelation in Eq. (1). The RIR is a linear time-invariant system, therefore the temporal expectation should not be considered under a set of source and microphone positions. In contrast, the speech signals depend on only their phonemes, in other words speech is not spatially stochastic. The spatial and temporal expectation of the ER component $R_{E,x}^{(m)}(\tau) \equiv E_{n,h}[x_E(n)x_E(n-\tau)]$ can be considered as:

$$R_{E,x}^{(m)}(\tau) = \sum_{i=0}^{N_E-1}\sum_{l=0}^{N_E-1} E_h[h(i)h(l)]E_n[s(n-i)s(n-\tau-l)]. \quad (3)$$

The statistical RIR model is described by the i.i.d. Gaussian random process as mentioned in Section 2.1. Substituting this Eq. (1) into Eq. (3) and using Eq. (2), the autocorrelation of the ER signal is formulated as follows:

$$R_{E,x}^{(m)}(\tau) = \sigma_b^2 R_s^{(m)}(\tau)\frac{1-e^{-2\Delta N_E}}{1-e^{-2\Delta}} \equiv (\sigma_b')^2 R_s^{(m)}(\tau). \quad (4)$$

2.2.2. Autocorrelation of the late reverberant signal

Considering the same way in Section 2.2.1, the autocorrelation of the LR signal $R_{R,x}^{(m)}(\tau) \equiv E_{n,h}[x_R(n)x_R(n-\tau)]$ is formulated as follows:

$$R_{R,x}^{(m)}(\tau) = (\sigma_b')^2 \sum_{m'=1}^{M_{60}} e^{-2\Delta m' N_E} R_s^{(m-m')}(\tau), \quad (5)$$

where $M_{60} = N_{60}/N_E$ is the number of frames corresponding to T_{60}.

2.2.3. Cross-correlation between the early-reflected and late reverberant signal

The expectation of the RIR becomes the Dirac's delta function in Eq. (1), therefore, it only has a value at the same sampling index that is, $i = l$. The region of the ER and LR signals is never overlapped, and this fact leads to the condition that $E_h[h(i)h(l)] = 0$. Therefore, the cross-correlation term becomes zero.

2.2.4. Autocorrelation function of the observed signal

Using the results in Section 2.2.1, 2.2.2 and 2.2.3, the autocorrelation function of the observed signal in the m-th frame is obtained as follows:

$$R_x^{(m)}(\tau) = (\sigma_b')^2\left\{R_s^{(m)}(\tau) + \sum_{m'=1}^{M_{60}} e^{-2\Delta m' N_E} R_s^{(m-m')}(\tau)\right\}, \quad (6)$$

Note that $e^{-2\Delta m' N_E} = 1$ under the condition $m' = 0$.

2.3. Signal model in the frequency domain

In the previous sections, the observed signal is modeled in terms of the autocorrelation functions of the ER and LR components. By the Wiener-Khinchin theorem, the power spectra are obtained from the autocorrelation functions. Each signal is transformed into the frequency domain by the short time Fourier transform method (STFT). Describing STFT as $\mathcal{F}[\cdot]$, each signal is described as $S(k,m) = \mathcal{F}[s(n)]$, $X(k,m) = \mathcal{F}[x(n)]$. The power spectra of the source and observed signals, $P_S(k,m)$ and $P_X(k,m)$, are obtained by STFT from the autocorrelation functions $R_s^{(m)}(\tau)$ and $R_x^{(m)}(\tau)$. The Fourier transform is linear, therefore, the power spectrum of the observed signal is decomposed into the ER and LR components as well, as in the case of the time domain; $P_X(k,m) = P_{E,X}(k,m) + P_{R,X}(k,m)$, where $P_{(\cdot),X}(k,m)$ is the power spectrum for each component. $P_{E,X}(k,m)$ is obtained using the source power spectrum $P_S(k,m)$ as follows:

$$P_{E,X}(k,m) = \mathcal{F}[(\sigma_b')^2 R_s^{(m)}(\tau)] = (\sigma_b')^2 P_S(k,m), \quad (7)$$

and also $P_{R,X}(k,m)$ is obtained as follows:

$$P_{R,X}(k,m) = (\sigma_b')^2 \sum_{m'=1}^{M_{60}} e^{-2\Delta m' N_E} P_S(k, m-m'). \quad (8)$$

3. PROPOSED DEREVERBERATION METHOD

By applying a dereverberation spectral gain $G(k,m)$ to the observed signal $X(k,m)$, an output dereverberated signal $Y(k,m) = G(k,m)X(k,m)$ is obtained. The output signal is transformed into the time domain by inverse STFT, $y(n) = \mathcal{F}^{-1}[Y(k,m)]$.

3.1. Wiener filter for speech enhancement

First of all, we consider the WF for speech enhancement under the reverberant condition. The observed signal $X(k,m)$ can be decomposed into ER and LR components; $X(k,m) = X_E(k,m) + X_R(k,m)$. The averaged power spectrum is obtained by the expectation of the frame-by-frame power spectra, and it is also described by the inner product of each signal in the frequency domain as $P_{(\cdot)}(k) = E_m[P_{(\cdot)}(k,m)] = E_m[(\cdot)(k,m)(\cdot)^*(k,m)]$. $P_{(\cdot)}(k)$ is the averaged power spectrum and $E_m[\cdot]$ is the expectation over the frame index m.

We consider the target signal as $X_E(k,m)$, because several tens of msec is a meaningful length of RIR for measures such as 'D_{50}' (Definition) or 'C_{80}' (Clarity) in the room acoustics research field [20]. 'D_{50}' is the early to total sound energy ratio. 'C_{80}' is the early to late arriving sound energy ratio. Each number means the time length of the early sound interval of msec. The objective function to minimize the mean square error is formulated as follows:

$$\mathcal{J}(\beta(k)) = E_m\big[\{X_E(k,m) - \beta(k)X(k,m)\} \{X_E(k,m) - \beta(k)X(k,m)\}^*\big], \quad (9)$$

where $\beta(k)$ is the frequency dependent variable. Differentiating $\mathcal{J}(\beta(k))$ with respect to $\beta(k)$, considering that $E_m[X_E(k,m)X^*(k,m)]$ and $E_m[X_E^*(k,m)X(k,m)]$ become $P_{E,X}(k)$, and the WF is obtained as follows:

$$\beta(k) = \frac{P_{E,X}(k)}{P_X(k)} = \frac{P_{E,X}(k)}{P_{E,X}(k) + P_{R,X}(k)}. \quad (10)$$

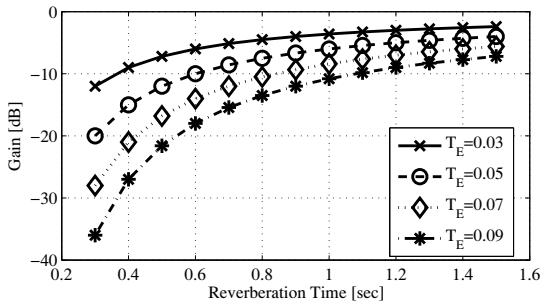

Fig. 1. LR suppression of the complementary Wiener filter.

$\beta(k)$ enhances the target signal under the reverberant condition. However, a critical issue is how to estimate the ER component $P_{E,X}(k)$ in Eq. (10).

3.2. Complementary Wiener filter

The CWF, $1-\beta(k)$ as mentioned in Section 1, can suppress the LR component without an estimation of the ER component, and in the following sections, we will introduce a theoretical analysis. Substituting Eq.(7) and Eq.(8) into Eq. (10), the CWF is obtained as follows:

$$1-\beta(k) = \frac{E_m[P_{R,X}(k,m)]}{P_X(k)} = e^{-\frac{6\log 10}{T_{60}}T_E}. \tag{11}$$

The right side of Eq. (11) is obtained from the RIR model , and Fig. 1 shows characteristics of the CWF. When T_{60} is under 1 sec, a remarkable fact is that the CWF can take small values under -10 [dB] based on the RIR model. When T_{60} is over 1 sec, the performance degrades according to a longer reverberation. This result shows the performance limitation of the CWF for dereverberation.

3.2.1. Hypothesis of speech presence

Hereafter, we focus on the hypothesis of the speech presence which is usually considered in the voice activity detection (VAD) research field. Note that considering the hypothesis of the speech presence for the dereverberation method is also adopted by Habets [8]. The speech presence of the source signal is considered as the hypothesis H_1, in contrast, speech absence (silence) is hypothesized as H_0. Using these hypotheses, the source power spectrum is represented as:

$$P_S(k,m) = \begin{cases} P_S(k,m)|_{m\in H_1} & \text{for target interval} \\ P_S(k,m)|_{m\in H_0} & \text{for silence interval} \end{cases}, \tag{12}$$

where $P_S(k,m)|_{m\in H_{(\cdot)}}$ means the hypothetical power spectrum. In addition, we consider probability of each hypothesis; ρ is the target presence probability and $1-\rho$ is the silence probability.

Considering the hypothetical power spectra, $P_S(k,m)|_{m\in H_0}$ can be ignored without a loss of generality, because the power in the silence interval is negligible. The observed power spectrum is transformed with the relationship $E_m[P_S(k,m)|_{m\in H_1}] = \rho E_m[P_S(k,m)]$ as follows:

$$\begin{aligned} &P_X(k)|_{m\in H_1\cup H_0} \\ &= (\sigma_b')^2 E_m\left[\sum_{m'=0}^{M_{60}} e^{-2\Delta m' N_E} P_S(k,m-m')|_{m-m'\in H_1}\right] \\ &= (\sigma_b')^2 \sum_{m'=0}^{M_{60}} e^{-2\Delta m' N_E} E_m[P_S(k,m-m')|_{m-m'\in H_1}] \\ &= \rho P_X[k] = \rho\{P_{E,X}[k] + P_{R,X}[k]\}. \end{aligned} \tag{13}$$

3.2.2. Observation in the silence interval

When the observed signal is in the silence interval, the summation of $P_X(k)|_{m\in H_1\cup H_0}$ in Eq. (13) is always considered under $m'\neq 0$ as follows:

$$\begin{aligned} &E_m[P_X(k,m)|_{m\in H_0}] \\ &= (\sigma_b')^2 \sum_{m'=m_1'}^{M_{60}} e^{-2\Delta m' N_E} E_m[P_S(k,m-m')|_{m\in H_0\cap m-m'\in H_1}], \\ &\equiv E_m\left[\mathrm{Par}_{(m_1')}[P_{R,X}(k,m)|_{m\in H_0\cap m-m'\in H_1}]\right], \end{aligned} \tag{14}$$

m_1' is the frame index from the beginning of the silence interval. $\mathrm{Par}_{(m_1')}[\cdot]$ denotes a partial sum operator, and $P_{R,X}(k)|_{m\in H_0\cap m-m'\in H_1}$ means a LR power spectrum which is observed in the silence interval of the source signal. The expectation of the partial sum satisfies inequalities as follows:

$$\begin{aligned} &E_m\left[\mathrm{Par}_{(m_1')}[P_{R,X}(k,m)|_{m\in H_0\cap m-m'\in H_1}]\right] \\ &\quad\le E_m[P_{R,X}(k,m)|_{m\in H_0\cap m-m'\in H_1}], \\ &\quad< E_m[P_{R,X}(k,m)|_{(m\in H_0\cup H_1)\cap(m-m'\in H_1)}] \\ &\quad= \rho E_m[P_{R,X}(k,m)]. \end{aligned} \tag{15}$$

Using Eq.(15), the ratio of two power spectra in Eq. (13) and Eq. (14) is less than the CWF as:

$$\begin{aligned} \frac{E_m[P_X(k,m)|_{m\in H_0}]}{P_X(k)|_{m\in H_1\cup H_0}} &< \frac{P_{R,X}(k)}{P_{E,X}(k)+P_{R,X}(k)} \\ &= 1-\beta(k). \end{aligned} \tag{16}$$

3.2.3. Observation in the target interval

When the observed signal is in the target interval, the ratio of the two power spectra of the observed signal, in the current frame m and the averaged one, can be considered by the same as way in Section 3.2.2.

$$\begin{aligned} \frac{E_m[P_X(k,m)|_{m\in H_1}]}{P_X(k)|_{m\in H_1\cup H_0}} &> \frac{\rho P_{E,X}(k)}{\rho\{P_{E,X}(k)+P_{R,X}(k)\}} \\ &= \beta(k). \end{aligned} \tag{17}$$

This ratio is greater than the WF.

3.3. Discussion and implementation

As shown in Section 3.2.2 and 3.2.3, when the observation is in the target interval, the statistical characteristic of the ratio between current and averaged power spectra preserves the target signal; the lower limit of the ratio is equal to the WF. In contrast, when the observation is in the silence interval, the ratio reduces reverberation by Eq. (16); the upper limit of the ratio is equal to the CWF. Note that the proposed method doesn't use a VAD method and only uses the current and averaged power spectra.

Table 1. Experimental conditions

Room impulse response	RWCP database (measured) [22]
	T_{60}: 0.3, 0.47, 0.6, 0.78, 1.3 sec
Source signal	JNAS speech database [23]
(Number of sources)	(8; 4 males, 4 females)
Sampling frequency	16 kHz
FFT and window size	found by the grid search
Shift size	1/4 of window size
Analysis window	Hann
Synthesis window	Tukey

(Window size: 128, 256, 512, 1024, 2048)
(FFT size: same as or 2 times of a window size)

For an implementation, the dereverberation spectral gain is considered as:

$$G(k,m) = \begin{cases} 1 & \frac{P_X(k,m)}{R_X(k,m)} > 1 \\ \dfrac{P_X(k,m)}{R_X(k,m)} & \text{otherwise} \end{cases}, \quad (18)$$

where $R_X(k,m)$ is an approximation of the averaged power spectra and it is formulated by the exponentially moving average (EMA) of the power spectrum $P_X(k,m)$ as:

$$R_X(k,m) = \alpha P_X(k,m) + (1-\alpha) R_X(k,m-1), \quad (19)$$

where α is a weighting factor of the EMA.

4. EVALUATIONS

4.1. Experimental dereverberation simulation

The proposed method is evaluated by reverberation reduction [21] (RR) and the improvement of target-to-interference ratio (TIR). RR describes an amount of a reverberation reduction, meaning the energy change under the hypothesis H_0. TIR is formulated as $10\log_{10}\frac{\sum_n x_E^2(n)}{\sum_n \{x_E(n)-z(n)\}^2}$, where $z(n)$ is the signal that should be evaluated. We consider the improvement from input to output TIR; in the case of input TIR, $z(n)$ is the observed signal $x(n)$, and in the case of output TIR, $z(n)$ is the processed signal $y(n)$. Speech signals are convolved by measured RIR in Table 1, and results of each objective criteria are averaged over eight speech signals in Table 1 for each T_{60}. The proposed method is compared to the conventional method by Lebart [11] (LebSS), because this is a representative single-channel dereverberation; however, considerable computational costs are required. For LebSS, the T_{60} has to be informed, and in this evaluation we obtain T_{60} from the RWCP database [22] description. Optimal smoothing parameters of LebSS for an estimation of the reverberant spectrum and a priori SNR are found by a grid search for the maximum TIR improvement. For the proposed method, the time lengths of phonemes are different from each speaker, therefore we also apply EMA to the denominator of the spectral gain and weighting factors of EMA are found by grid search. For each method, the STFT parameters are also found by grid search.

4.2. Discussion with computational complexity

For T_{60} under 0.5 sec, both the TIR improvement and RR of the proposed method are almost same as LebSS in Fig. 2 and Fig. 3.

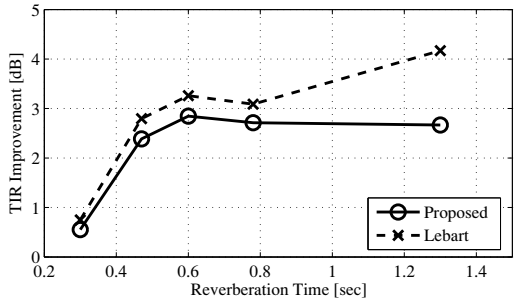

Fig. 2. Target-to-interference ratio improvement

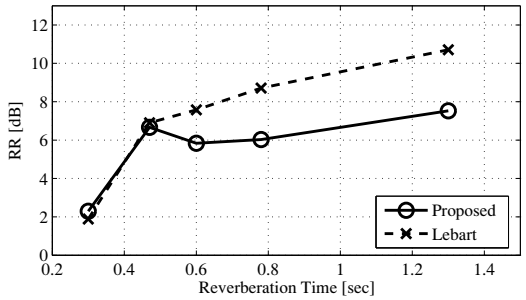

Fig. 3. Reverberation reduction

For T_{60} above 0.5 sec, the RR is degraded. On the other hand, the degradation of the TIR improvement is still small for $0.5 < T_{60} < 0.8$ sec. For the condition as $T_{60} > 1$ sec, the proposed method is worse than LebSS on both of the objective measures.

The computational complexity of the proposed method in one frequency bin is less than the conventional method; it is reduced by 50%. Memory consumption of LebSS must store a few frames of power spectra and a priori SNR; in contrast, each EMA stores one power spectrum for the proposed method. Divisions on embedded processors are usually implemented by an iterative method, therefore the number of divisions is one of the dominant computational costs. Two divisions are needed for LebSS to obtain a reverberant spectrum and a priori SNR, on the other hand the proposed method uses only one division to obtain a spectral gain.

Therefore, the proposed method is computationally efficient and shows similar performance to the conventional method in practical situations where T_{60} is under several hundreds of msec such as in a meeting room.

5. CONCLUSION

A computationally efficient single-channel dereverberation method is proposed. The proposed method is based on the complementary Wiener filter, and the memory consumption and the number of operative calculations are by half reduced compared to the conventional method. The reductions are achieved for the frequency domain operation, with similar dereverberation performance compared to the conventional method for practical situations. Future work includes analyzing under not only reverberant conditions but also noisy conditions, and evaluating the subjective sound quality.

6. REFERENCES

[1] J.B. Allen, D.A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Am.*, vol. 62, no. 4, pp. 912–915, Oct. 1977.

[2] E.A.P. Habets, "Towards multi-microphone speech dereverberation using spectral enhancement and statistical reverberation models," in *Proc. of Asilomar Conference 2008*, Oct. 2008, pp. 806–810.

[3] M. Jueb and P. Vary, "Binaural dereverberation based on a dual-channel wiener filter with optimized noise field coherence," in *Proc. of ICASSP 2010*, Mar. 2010, pp. 4710–4713.

[4] T. Gerkmann, "Cepstral weighting for speech dereverberation without musical noise," in *Proc. of EUSIPCO 2011*, Sep. 2011, pp. 2309–2313.

[5] A. Schwarz and K. Reindl an W. Kellermann, "On blocking matrix-based dereverberation for automatic speech recognition," in *Proc. of IWAENC 2012*, Sep. 2012, pp. 1–4.

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

[7] M. Miyoshi, M. Delcroix, and K. Kinoshita, "Calculating inverse filters for speech dereverberation," *IEICE Trans. Fundamentals*, vol. E91-A, no. 6, pp. 1303–1309, Jun. 2008.

[8] P. A. Naylor and N. D. Gaubitch, *Speech Dereverberation*, Springer-Verlag, London, UK, 2010.

[9] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. ASSP*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

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

[11] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica*, vol. 87, no. 3, pp. 359–366, May 2001.

[12] 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. ASLP*, vol. 17, no. 4, pp. 534–545, May 2009.

[13] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–773, Sep. 2009.

[14] T. Inoue, H. Saruwatari, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise in wiener filter via higher-order statistics," in *Proc. of APSIPA 2010*, Dec. 2010, pp. 121–124.

[15] F. Migliaccio, M. Reguzzoni, F. Sansò, and C. C. Tscherning, "An enhanced space-wise simulation for goce," in *Proc. of 2nd International GOCE User Workshop*, Mar. 2004.

[16] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. ASSP*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[17] R. Singh, B. Raj, and P. Smaragdis, "Latent-variable decomposition based dereverberation of monaural and multi-channel signals," in *Proc. of ICASSP 2010*, Mar. 2010, pp. 1914–1917.

[18] J-D. Polack, *La transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Dissertation. Université du Maine, 1988.

[19] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, no. 6, pp. 409, Mar. 1965.

[20] H. Kuttruff, *Room Acoustics*, Spon Press, London, UK, 4th edition, 2000.

[21] E.A.P. Habets, "Single-channel speech dereverberation based on spectral subtraction," in *Proc. of ProRISC 2004*, Nov. 2004, pp. 250–254.

[22] "RWCP sound scene database," http://research.nii.ac.jp/src/en/RWCP-SSD.html.

[23] "JNAS speech database," http://research.nii.ac.jp/src/en/JNAS.html.