# A COMPUTATIONALLY RESTRAINED AND SINGLE-CHANNEL BLIND DEREVERBERATION METHOD UTILIZING ITERATIVE SPECTRAL MODIFICATIONS

*Kazunobu Kondo*[†]

[†]Yamaha Corporation, Hamamatsu, Japan

## ABSTRACT

A computationally restrained, single-channel, blind dereverberation method is proposed. The proposed method consists of two iterative spectral modifications, which employs spectral subtraction for noise reduction, and a complementary Wiener filter for dereverberation. Modulation transfer function is employed to calculate the dereverberation parameters. Late reverberation is estimated without any delaying operations, in contrast to other commonly used dereverberation methods. The proposed method achieves very balanced dereverberation and distortion reduction performance, in spite of the proposed 'rough' $T_{60}$ estimation technique. Some signal delay occurs as the result of Short Time Fourier Transform, but this delay is equivalent to the delay caused by conventional noise reduction methods such as spectral subtraction. Computational cost is sufficiently restrained, despite of the use of iterative spectral processing.

***Index Terms***— Dereverberation, Noise reduction, Wiener filter, Modulation transfer function, Computational cost

## 1. INTRODUCTION

Speech communication and recognition systems are generally used in noisy and reverberant environments, such as meeting rooms, and reverberation time is usually under 1 second. However, speech quality and recognition performance are degraded under these conditions. To counter this degradation, various dereverberation techniques have been developed. Multi-microphone techniques estimate late reverberation using spatial correlation [1–5], or estimate an inverse filter using the MINT theorem [6] in [7, 8]. Although multi-microphone techniques can be applied to devices such as smart-phones and other portable equipment, single-channel methods are still useful for some speech enhancement applications. One serious challenge when performing single-channel dereverberation is that there is no spatial information which can be utilized. Various single-channel speech enhancement techniques have been used for dereverberation, such as spectral subtraction [9] and MMSE-STSA [10], and successful results have been reported [8, 11–13].

Several functions are used in voice terminals to improve speech quality, such as echo cancellation and noise reduction. Since all of these functions work concurrently, each function needs to be computationally efficient, because overall computational cost should usually be kept low. A Wiener filter (WF) is often used to enhance the target signal. If the range of WF $\beta$ is $0 \le \beta \le 1$, then $1 - \beta$ could be referred to as a "complementary Wiener filter" (CWF) [14]. In a previous study, the author of the current study utilized computationally efficient CWFs for the purpose of dereverberation [15], but their performance proved to be insufficient for longer reverberation times, in addition the parameters were heuristically determined by simply using a grid search.

In this paper, an improved dereverberation method based on the use of a CWF is proposed, in which the power spectrum is iteratively modified. The parameters are estimated using modulation transfer function (MTF) related to the reverberation. The proposed method blindly estimates reverberation time, and CWF parameters are then calculated using this estimate. The CWF is then used to estimated late reverberation for dereverberation processing, during which the estimated late reverberations are iteratively subtracted from the observed signal.

The rest of this paper is organized as follows: Section 2 briefly describes the signal model and dereverberation method previously proposed by the author. Section 3 describes the proposed method. Section 4 describes the dereverberation experiment and its results. Section 5 concludes this paper.

## 2. DEREVERBERATION USING A CWF

This section describes the signal model and CWF-based dereverberation method previously proposed by the author [15]. Observed signal $X(k, m)$ can also be described as $S(k, m)H(k, m)$, where $S(k, m)$ and $H(k, m)$ represent the source signal and the room impulse response (RIR), respectively, and where $k$ and $m$ are frequency bin and frame indexes, respectively. The RIR is considered to conform with Polack's statistical model [16]. The observed power spectrum $P_X(k, m)$ is formulated using source power spectrum $P_S(k, m)$ and Polack's statistical RIR model:

$$P_X(k, m) = C \sum_{m'=0}^{M_{60}} e^{-2\Delta m' N_E / T_{60}} P_S(k, m - m'), \qquad (1)$$
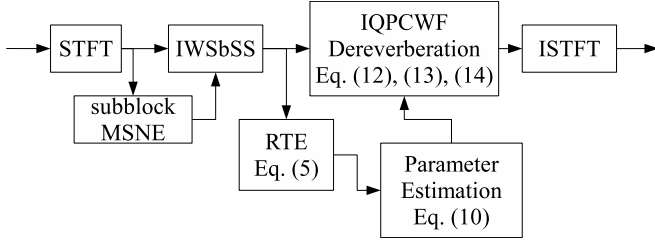
**Fig. 1**. Diagram of proposed dereverberation method.

where $T_{60}$, $F_s$ and $N_E$ represent reverberation time, sampling frequency, and frame size of the Short Time Fourier Transform, respectively. $\Delta$ is the energy decay rate of the reverberation, which can be represented as:

$$\Delta = \{3 \log 10\} / F_s. \tag{2}$$

$M_{60}$ is the number of frames corresponding to $T_{60}$; $M_{60} = F_s T_{60} / N_E$. $C$ is a constant representing RIR energy. The CWF is obtained from the ratio between two exponentially moving averages (EMAs) of the observed power spectra. $G(k, m)$ is the spectral gain function for dereverberation, which can be represented as:

$$G(k, m) = \min \left\{ 1, \ R_X^{(S)}(k, m) / R_X^{(L)}(k, m) \right\}, \tag{3}$$

where

$$R_X^{(\cdot)}(k, m) = \alpha^{(\cdot)} P_X(k, m) + (1 - \alpha^{(\cdot)}) R_X(k, m - 1). \tag{4}$$

$\alpha^{(\cdot)}$ is an EMA coefficient. $(S)$ and $(L)$ represent shorter and longer time constants, respectively, which are needed to express the condition $\alpha^{(S)} > \alpha^{(L)}$. Finally, the dereverberated signal $Y(k, m)$ is obtained by: $Y(k, m) = G(k, m) X(k, m)$.

## 3. PROPOSED DEREVERBERATION METHOD

A block diagram of the proposed dereverberation method is shown in Fig. 1. Stationary noise power is calculated on a basis of minimum statistics noise estimation (MSNE). Stationary noise is reduced by iterative, weak, sub-block-wise spectral subtraction (IWSbSS). Before dereverberation, $T_{60}$ is roughly estimated (RTE), and dereverberation parameters are estimated by using MTF. Iterative, quasi-parametric, CWF (IQPCWF) reduces the reverberation.

### 3.1. Iterative weak sub-block-wise spectral subtraction

MSNE is one of the most common methods used to estimate the power of stationary noises. In order to reduce computational cost, the minimum power spectrum is evaluated in sub-blocks with noise at the minimum power level [17]. The iterative, weak, spectral subtraction (IWSS) method has also been proposed for improving speech quality [18–20]. IWSS is

used to restrain 'musical noise' artifacts which are generated during noise reduction processing, and must estimate different noise prototypes at different iterative stages. When using MSNE, each noise candidate in a sub-block occurs at a different time interval, so the candidates are expected to differ from one another. Therefore, sub-block-wise minimum noise spectrum estimation can be considered as a sub-block noise prototype, and can be subtracted at one stage of IWSS.

### 3.2. $T_{60}$ estimation using an adaptive threshold operation involving median filtering

In the acoustics research field, $T_{60}$, the time it takes for a sound to decay to 60 dB below its final power level, is traditionally used to represent reverberation time. It is usually measured in the divided frequency bands, such as the octave bands, for example. A basic frequency of 500 Hz is usually used to determine $T_{60}$ in the architectural acoustics field [21], so frequency bins around 500 Hz are important when estimating $T_{60}$. In the proposed $T_{60}$ estimation method, reverberation of the observed signal is tentatively separated only around 500 Hz, by using the quasi-complementary Wiener filter (QCWF) described in Section 3.3.2 of this paper. The QCWF parameters are fixed and should be 'strong', since tentative separation is only performed to estimate $T_{60}$.

Observed power spectra $P_X(k, m)$ can be separated into early reflection (ER) and late reverberation (LR) components using the statistical RIR model. The bin-wise ER spectra are averaged over the frequency bins in the designated region around 500 Hz, and the same process is also performed for the LR spectra. The power envelope is calculated for both ER and LR using the moving average from each averaged spectrum. Median filtering is then applied to these envelopes to estimate thresholds which can be used to identify activity. The total power of the ER and LR components, $\tilde{P}_{E,X}$ and $\tilde{P}_{R,X}$, respectively, are calculated for the active intervals. Thus, $T_{60}$ can be calculated as follows:

$$\hat{T}_{60} \approx \{N_E \cdot 2\Delta\} / \left\{ \log \left( 1 + E_m \left[ \tilde{P}_{E,X} / \tilde{P}_{R,X} \right] \right) \right\}, \tag{5}$$

where $E_m[\cdot]$ represents an expectation regarding frame index $m$, and $\hat{T}_{60}$ is the estimated reverberation time.

### 3.3. Dereverberation using an iterative complementary Wiener filter

#### 3.3.1. Parameter estimation using MTF

EMA coefficients $\alpha_{(\cdot)}$ in Eq. (4) can be converted into forgetting factor $\zeta_{(\cdot)} = 1 - \alpha_{(\cdot)}$, which specifies how quickly the filter "forgets" past sample information. When the z-transform is applied to the EMA, dereverberation gain Eq. (3) is recalculated as:

$$G(k, e^{j\omega_m T_H}) = \frac{1 - \zeta_{(S)}}{1 - \zeta_{(L)}} \frac{1 - \zeta_{(L)} e^{-j\omega_m T_H}}{1 - \zeta_{(S)} e^{-j\omega_m T_H}}, \tag{6}$$

where $\omega_m$ is the modulation angular frequency, and $T_H$ is the time length of a frame shift. Eq. (6) illustrates how dereverberation gain Eq. (3) corresponds to the first order auto-regressive moving-average (ARMA) filter. In addition, Eq. (6) can be divided into two filters,

$$G(k, e^{j\omega_m T_H}) = H_L(\zeta_{(L)}, e^{j\omega_m T_H}) H_H(\zeta_{(S)}, e^{j\omega_m T_H}), \quad (7)$$

where $H_L$ and $H_H$ represent low-pass and high-pass filters, respectively.

MTF represents the loss of modulation as a result of reverberation [22]. MTF can also be formulated using the statistical RIR model [16]. Unoki [23] formulated MTF $m(\omega_m)$ as follows:

$$m(\omega_m) = \sqrt{1 + \{\omega_m T_{60}/(2F_s\Delta)\}^2}. \quad (8)$$

When the two coefficients of the ARMA filter, $\zeta_{(S)}$ and $\zeta_{(L)}$, were estimated using the modified Yule-Walker method and MTF in our preliminary experiments, dereverberation performance was low. Therefore, a two-step optimization method is proposed. First, MTF is only used to optimize the coefficient of the high-pass filter $H_H$. Then the Yule-Walker method is applied to estimate the low-pass filter $H_L$. For $H_H$, the cosine term can be expanded using the Taylor series in the amplitude response. Neglecting the higher-order terms of the Taylor series, and comparing the coefficients of $m(\omega_m)$ and $H_H$, the following relationship is obtained:

$$\left\{\zeta_{(L)}/(1 - \zeta_{(L)})^2\right\} N_H^2 = \{1/(2\Delta)\}^2, \quad (9)$$

where $N_H = T_H F_s$. This is a quadratic equation, so it has two solutions. Considering the value range of the solutions, $\zeta_{(L)}$ can be obtained as follows:

$$\zeta_L = \left\{2 + (2N_H\Delta)^2 - \sqrt{\left\{2 + (2N_H\Delta)^2\right\}^2 - 4}\right\}/2 \quad (10)$$

The combined amplitude response, $|m(\omega_m)||H_H|$, tends towards a high-pass response due to omitting the higher-order terms of the Taylor series, which results in over compensation. For the low-pass filter $H_L$, the Yule-Walker method is used to estimate coefficient $\zeta_{(S)}$ of the first order AR filter. $H_L$ compensates for high-pass amplitude response $|m(\omega_m)||H_H|$. Finally, $G(k, z)$ in Eq. (6) is determined by 2-step optimization, and dereverberation spectral gain in Eq. (3) is obtained by $\hat{T}_{60}$ and MTF.

### 3.3.2. Quasi-parametric complementary Wiener filter

Dereverberation using Eq. (3) has performance limitations when there are longer values for $T_{60}$ [15]. As $T_{60}$ approaches huge values, such as $\infty$, dereverberation spectral gain approaches 1: $\lim_{T_{60}\to\infty} G(k, m) = 1$. This means that
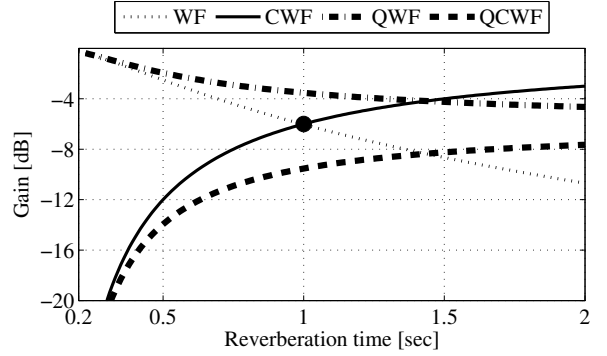


**Fig. 2**. Theoretical performance of CWF and QCWF.

$G(k, m)$ from Eq. (3) does not suppress LR when $T_{60}$ is very long. Quasi- and quasi-parametric Wiener filters (QWF and QPWF) were proposed in [24, 25] to achieve flexible noise reduction. If we introduce the concept of a QWF into the CWF dereverberation method, spectral gain can be reformulated as:

$$\mathcal{F}(k, m) = \min\left\{1, \frac{R_X^{(S)}(k, m)}{R_X^{(L)}(k, m) + R_X^{(S)}(k, m)}\right\}. \quad (11)$$

For huge $T_{60}$ values, this quasi-complementary Wiener filter (QCWF) satisfies $\lim_{T_{60}\to\infty} \mathcal{F}(k, m) = 0.5$. Fig. 2 shows the theoretical dereverberation performance of a CWF and a QCWF. The theoretical CWF appears as Eq.(11) in [15]. The theoretical QCWF is derived in the same manner as the CWF, using the statistical RIR model and Eq. (11):

$$1/\{e^{2\Delta N_E/T_{60}} + 1\}. \quad (12)$$

In Fig. 2, a crosspoint is found for the WF and CWF curves, and this point is indicated by a black circle. When $T_{60}$ is over 1 second, CWF performance exceeds WF performance, which means that the dereverberation performance of the CWF deteriorates with longer reverberation times. On the other hand, the performance of QWF always exceeds that of the QCWF. When $T_{60}$ is 1 second, reverberation is reduced by 6 dB. When $T_{60}$ is 2 seconds, reverberation is reduced by only 3 dB. Therefore, for a $T_{60}$ of under 1 second, the proposed QWF-based method works properly, however for values over 2 seconds, it can be assumed that the proposed method will not perform well. By introducing an additional parameter to control the 'strength' of the QCWF, a quasi-parametric CWF (QPCWF) can be obtained as follows:

$$\mathcal{G}(k, m) = \min\left\{1, \frac{R_X^{(S)}(k, m)}{R_X^{(L)}(k, m) + w(T_{60})R_X^{(S)}(k, m)}\right\}, \quad (13)$$

where $w(T_{60})$ is a weighting function, for example $w(T_{60}) = T_{60}$. Intuitively, for shorter $T_{60}$ values, the 'strength' term

$w(T_{60})R_X^{(S)}(k,m)$ should be small to prevent excessive LR suppression. For longer $T_{60}$ values, the weighting should be large in order to increase dereverberation performance.

### 3.3.3. Iterative complementary Wiener filtering

For dereverberation, the LR spectrum is usually estimated using a delaying operation which delays the averaged power spectrum, as discussed in [11, 26], for example. A QPCWF can estimate the LR component without a delaying operation, however, decreasing memory consumption, which is always beneficial, especially when dereverberation is performed using digital signal processors. As a further refinement, the proposed method uses an iterative spectral modification technique known as iterative, weak, spectral subtraction (IWSS) [18–20]. In the $i$-th iterative stage, the QPCWF estimates the LR component as follows:

$$\tilde{P}_{R,X}^{(i)}(k,m) = \left(1 - \mathcal{G}^{(i)}(k,m)\right) P_X^{(i-1)}(k,m), \qquad (14)$$

where $\tilde{P}_{R,X}^{(i)}(k,m)$ represents the LR component in the $i$-th stage, the $k$-th frequency bin, and the $m$-th frame. $\mathcal{G}^{(i)}(k,m)$ is the QPCWF in $i$-th stage. $P_X^{(i)}(k,m)$, which is the enhanced power spectrum in the $i$-th stage, is represented as follows:

$$P_X^{(i)}(k,m) = \qquad (15)$$
$$\max\left\{ P_X^{(i-1)}(k,m) - \beta\tilde{P}_{R,X}^{(i)}(k,m), \eta R_X^{(L,i)}(k,m) \right\},$$

where $R_X^{(L,i)}(k,m)$ represents the EMA of the power spectra at the $i$-th stage, which is calculated in the same manner of Eq. (4). Dereverberation gain is calculated as: $\mathcal{A}^{(i)}(k,m) = P_X^{(i)}(k,m)/P_X(k,m)$ in the $i$-th stage. Finally, all gains are multiplied by each other: $\mathcal{A}(k,m) = \prod_i \mathcal{A}^{(i)}(k,m)$.

## 4. DEREVERBERATION EXPERIMENTS

In this section, the proposed method (IQPCWF) is compared to the following conventional dereverberation methods: Spectral Subtraction (SS), proposed by Lebart [11], and Optimal Modified Minimum Mean-Square Error Log-Spectral Amplitude (OM-LSA), proposed by Habets [26], as well as to our previously proposed method (CWF) [15]. These methods are evaluated with our proposed noise reduction method incorporated, except for OM-LSA, which includes its own noise reduction technique. In the case of OM-LSA, minimum statistics are used for stationary noise power estimation.

### 4.1. Simulation conditions

The REVERB challenge dataset consists of simulated data (SimData) [27] and real recordings (RealData) [28]. The sampling frequency is 16 kHz. SimData includes three types

**Table 1**. Dereverberation parameters

| CWF [15] | |
|---|---|
| $\alpha_{(\cdot)}$ | calculated by estimated $T_{60}$ |
| IQPCWF (proposed) | |
| $\alpha_{(\cdot)}$ | calculated by estimated $T_{60}$ |
| $\beta$ | 1.0 (subtraction) |
| $\eta$ | 0.64 (flooring) |
| num. of iterations | 5 |
| OM-LSA [26] | |
| $q$ | 0.2 (speech absence) |
| $\eta_z^{\mathrm{d}}$ | 0.95 (smoothing) |
| $\eta_z^{\mathrm{a}}$ | 0.8 ($0 \le \eta_z^{\mathrm{a}} \le \eta_z^{\mathrm{d}}$) |
| $T_l$ | 50 msec |
| SS [11] | |
| $\beta$ | 0.9 (smoothing) |
| $T$ | 50 msec |
| $\lambda$ | 0.1 (flooring) |

(Parameters taken directly from each study.)

of rooms, and the $T_{60}$ of the three rooms are 0.25, 0.5 and 0.7 seconds. RealData includes one type of room, and $T_{60}$ is 0.7 seconds. Two microphone positions, near and far, are included in both SimData and RealData. All of the dereverberation methods employ STFT for time-frequency analysis. The STFT parameters are the same for all the methods; the window size is 1024, the FFT size is 2048, and the shift size is 256. For noise reduction, the number of the sub-blocks is 9, which means that block length is about 3 seconds. $T_{60}$ is estimated every 30 frames, which corresponds to 480 msec., and envelopes are kept for a maximum of 2 seconds. The parameters of each method's dereverberation algorithms are determined based on each method's reference literature, as shown in Table 1). For CWF and IQPCWF, the EMA parameters are calculated using the proposed method as described in Section 3.3.1.

### 4.2. Processing delay

The processing delay can be separated into two delays: the signal delay for input/output, and the $T_{60}$ estimation delay. The proposed method is a real-time operation, since the signal is processed frame-by-frame. Processing delays for the proposed method are as follows:

- signal delay: 64 msec (equal to the signal delay of spectral subtraction),

- $T_{60}$ estimation delay: 480 msec (30 frames)

The proposed method uses STFT, and the signal delay equals the window size. This delay is the same as that of conventional noise reduction methods such as spectral subtraction, for example. The proposed method outputs $\hat{T}_{60}$ every 30

frames. The $T_{60}$ estimation can store 2 seconds of envelope data, therefore a stable $\hat{T}_{60}$ value can be expected 2 seconds from the beginning of the observed signal.

## 4.3. Discussion of experimental results and computational costs

Four mandatory objective measures are required for the REVERB Challenge: Cepstral distance (CD) [29], Log-likelihood ratio (LLR) [29], Frequency-weighted segmental SNR (FWSegSNR) [29] and Speech-to-reverberation modulation energy ratio (SRMR) [30]. Fig. 3 and Fig. 4 show our experimental results for the far and near positions, respectively. Results are obtained from real-time operation, and are averaged for all utterances. Bars in the figures represent experimental blind dereverberation results. Circles in the figures show the results under the 'oracle $T_{60}$' condition, which is a fixed $T_{60}$ condition designated in the experimental instructions of the REVERB challenge.

The proposed IQPCWF method achieved higher SRMRs than the CWF and SS methods, meaning that the proposed method can more effectively suppress LR than the other two methods. With respect to the distortion measures CD, LLR and FWSegSNR, the proposed method achieved similar performance to CWF. When comparing SS, CWF and the proposed method, levels of distortion vary according to the acoustic conditions, making it difficult to determine which method is more effective overall. Under all conditions, OM-LSA achieved high SRMR values, however levels of distortion were significantly worse when using OM-LSA. These findings suggest that the proposed method achieves a significantly better balance of dereverberation performance and distortion reduction than conventional methods.

The proposed method's $T_{60}$ estimation is computationally efficient, but inaccurate. For OM-LSA and SS, there are significant differences between real-time $T_{60}$ estimation and estimation under the 'oracle' condition, as shown in Fig. 3 and Fig. 4. This is important, because OM-LSA and SS are sensitive to the accuracy of $T_{60}$ estimation. In spite of its 'rough' $T_{60}$ estimation, the proposed method proved to be rather robust, because only negligible differences are observed between real-time $T_{60}$ estimation results and $T_{60}$ estimation results under the 'oracle' condition.

The computational cost of each of these dereverberation methods was evaluated using the real-time factor (RTF), and the results appear in Table 2. About three times the RTF is required by the proposed method, in comparison to minimum variance distortionless response (MVDR), which is shown as a reference. CWF's cost is less three times as large as the cost of MVDR, and is thus the most computationally efficient of the examined methods. In comparison, the RTF of the proposed method is 1.3 times that of CWF. Although the proposed method uses iterative spectral processing, it requires only 1.1 times the RTF of the SS method, which involves only

**Table 2**. Computational cost

| Method | Real Time Factor | |
|---|---|---|
| | RealData | SimData |
| MVDR (reference) | 0.017 | 0.016 |
| CWF | 0.038 | 0.037 |
| IQPCWF (proposed) | 0.050 | 0.048 |
| OM-LSA | 0.070 | 0.067 |
| SS | 0.045 | 0.045 |

one-shot spectral subtraction. The OM-LSA estimator uses an incomplete gamma function which includes numerical integration, which is why the computational cost of OM-LSA is so high.

## 5. CONCLUSION

In this study, a single-channel, computationally restrained, blind dereverberation method was proposed. The proposed method is a real-time operation, consisting of iterative, weak, sub-block-wise spectral subtraction for noise reduction, $T_{60}$ estimation, parameter optimization using MTF, and employs an iterative, quasi-parametric, complementary Wiener filter for dereverberation. Experimental evaluation showed that the proposed method achieves a better balance of dereverberation and distortion performance than conventional methods. Additionally, in spite of its 'rough' estimation of $T_{60}$, the proposed method is significantly robust under various acoustic conditions. Even though the proposed method involves iterative processing, computational cost is sufficiently constrained. Signal delay occurs as the result of STFT processing, but this delay is equivalent to the delay caused by conventional noise reduction methods such as spectral subtraction. Future work includes parameter optimization for the iterative method, as well as evaluation of the resulting 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*, 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*, Mar. 2010, pp. 4710–4713.
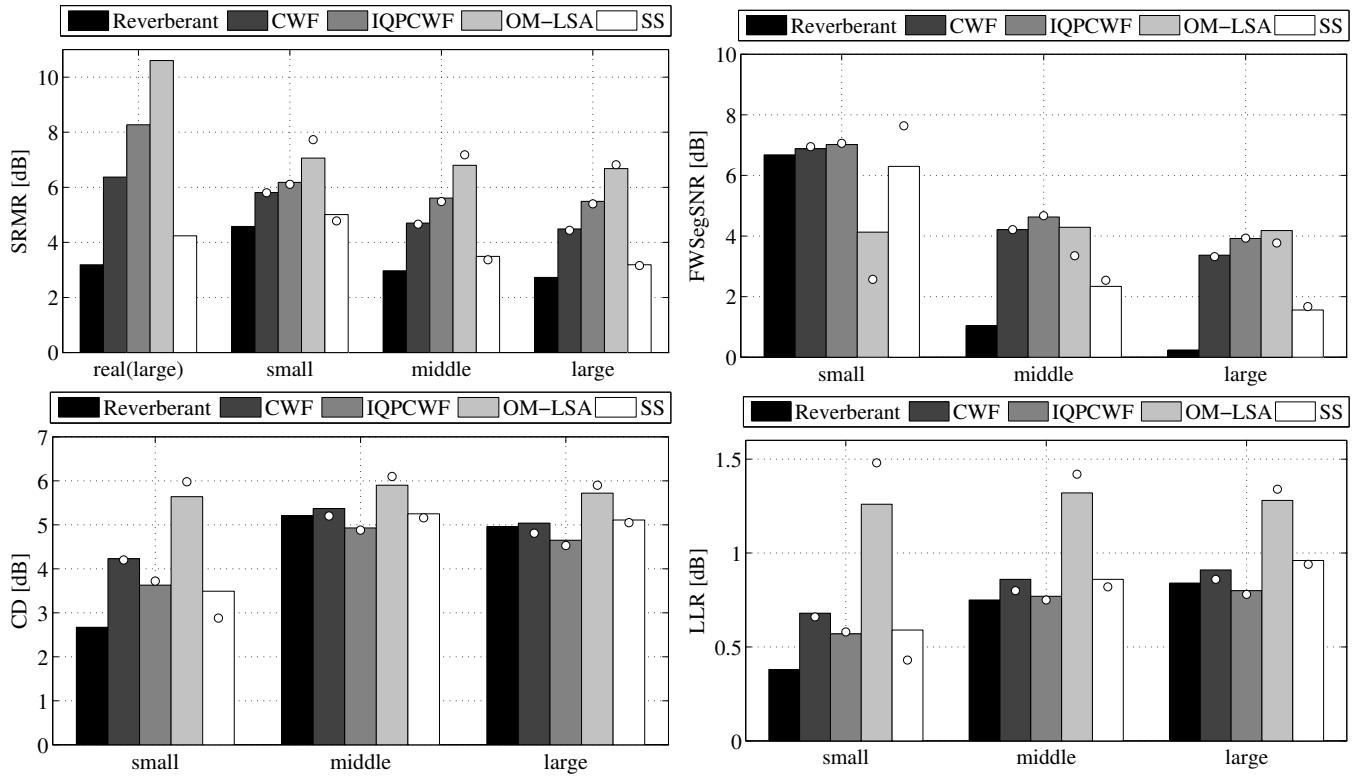
**Fig. 3**. Dereverberation performance for far position. Circles represent 'oracle' condition. Higher SRMR and FWSegSNR values indicate better performance; lower CD and LLR values indicate better performance.

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

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

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

[7] M. Miyoshi, M. Delcroix, and K. Kinoshita, "Calculating inverse filters for speech dereverberation," *IEICE Transactions on 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 Transactions on Acoustics, Speech and Signal Processing*, 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 Transactions on Acoustics, Speech and Signal Processing*, 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 Transactions on Audio, Speech and Language Processing*, 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] 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.

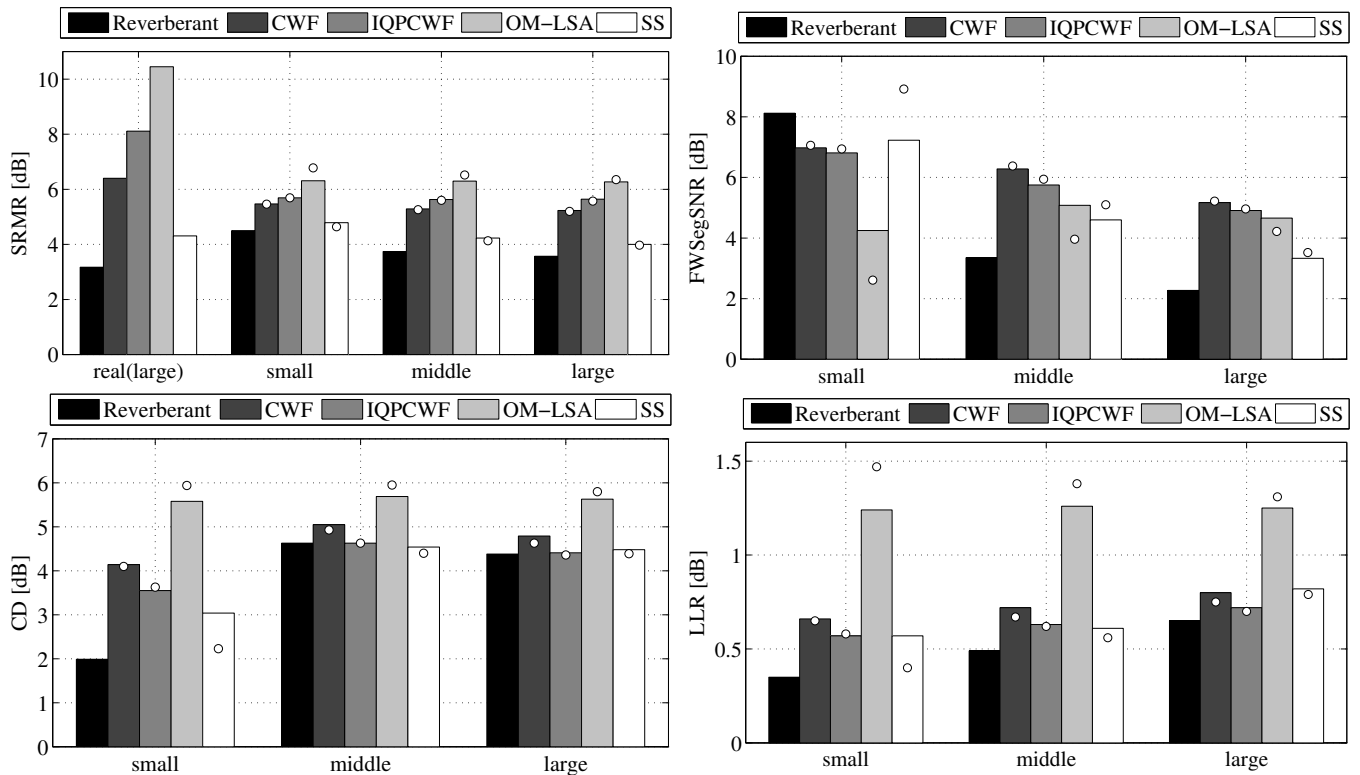[15] K. Kondo, Y. Takahashi, T. Komatsu, T. Nishino, and K. Takeda, "Computationally efficient single channel

**Fig. 4**. Dereverberation performance for near position. Circles represent 'oracle' condition. Higher SRMR and FWSegSNR values indicate better performance; lower CD and LLR values indicate better performance.

dereverberation based on complementary wiener filter," in *Proc. of ICASSP*, May 2013, pp. 7452–7456.

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

[17] R. Martin, "An efficient algorithm to estimate the instantaneous snr of speech signals.," in *Proc. Eurospeech*, Sep. 1993, pp. 1093–1096.

[18] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, and K. Kondo, "Theoretical analysis of iterative weak spectral subtraction via higher-order statistics," in *Proc. of MLSP*, Sep. 2010, pp. 220–225.

[19] R. Miyazaki, H. Saruwatari, T. Inoue, K. Shikano, and K. Kondo, "Musical-noise-free speech enhancement: Theory and evaluation," in *Proc. of ICASSP*, Mar. 2012, pp. 4565–4568.

[20] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, T. Inoue, K. Shikano, and K. Kondo, "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 7, pp. 2080–2094, Sep. 2012.

[21] M. Rettinger, *Acoustic Design and Noise Control*, Chemical Publishing, New York, N.Y., 1977.

[22] T. Houtgast and H. J. M. Steeneken, "A review of the mtf concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.*, vol. 77, no. 3, pp. 1069–1077, Mar. 1985.

[23] M. Unoki, M. Furukawa, K. Sakata, and M. Akagi, "An improved method based on the mtf concept for restoring the power envelope from a reverberant signal," *Acoustical Science and Technology*, vol. 25, no. 4, pp. 232–242, July 2004.

[24] J. Even, H. Saruwatari, K. Shikano, and T. Takatani, "Speech enhancement in presence of diffuse background noise: Why using blind signal extraction?," in *Proc. ICASSP*, Mar. 2010, pp. 4770–4773.

[25] T. Inoue, H. Saruwatari, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise in wiener filtering family via higher-order statistics," in *Proc. of ICASSP*, May 2011, pp. 5076–5079.

[26] E. A. P. Habets, *Single- and multi-microphone speech dereverberation using spectral enhancement*, Ph.D. thesis, Technische Universiteit Eindhoven, 2007.

[27] T. Robinson, J. Fransen, D. Pye, J. Foote, and S. Renals, "Wsjcam0: A british english speech corpus for large vocabulary continuous speech recognition," in *Proc. of ICASSP*, May 1995, pp. 81–84.

[28] M. Lincoln, I. McCowan, J. Vepa, and H.K. Maganti, "The multi-channel wall street journal audio visual corpus (mc-wsj-av): Specification and initial experiments," in *Proc. of IEEE Workshop on Automatic Speech Recognition and Understanding*, Nov. 2005, pp. 357–362.

[29] Y. Hu and P.C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[30] T.H. Falk, C. Zheng, and W.Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 7, pp. 1766–1774, Sep. 2010.