Constrained Wiener Gains and Filters for Single-Channel and Multichannel Noise Reduction

Tao Long* Jacob Benesty† and Jingdong Chen*
* Northwestern Polytechnical University, Xi'an, China
E-mail: jingdongchen@ieee.org Tel/Fax: +86-029-88491079
† University of Quebec, Montreal, Canada
E-mail:benesty@emt.inrs.ca Tel/Fax: +01-514-2287019

***Abstract*—Noise reduction has long been an active research topic in signal processing and many algorithms have been developed over the last four decades. These algorithms were proved to be successful in some degree to improve the signal-to-noise ratio (SNR) and speech quality. However, there is one problem common to all these algorithms: the volume of the enhanced signal after noise reduction is often perceived lower than that of the original signal. This phenomenon is particularly serious when SNR is low. In this paper, we develop two constrained Wiener gains and filters for noise reduction in the short-time Fourier transform (STFT) domain. These Wiener gains and filters are deduced by minimizing the mean-squared error (MSE) between the clean speech and the speech estimate with the constraint that the sum of the variances of the filtered speech and residual noise is equal to the variance of the noisy observation.**

I. Constrained Wiener Gains For Single- Channel Noise Reduction

A. Signal Model

Let us first consider the single-channel noise reduction problem, which is to recover the zero-mean desired signal $x(t)$ [1], [2], [3] from the following noisy signal:

$$y(t) = x(t) + v(t), \tag{1}$$

where t is the time index and $v(t)$ is the zero-mean unwanted additive noise. The noise signal $v(t)$ is assumed to be uncorrelated with $x(t)$.

In the frequency domain and at the frequency index f, the signal model in (1) is expressed as

$$Y(f) = X(f) + V(f), \tag{2}$$

where $Y(f)$, $X(f)$, and $V(f)$ are the frequency-domain representations of $y(t)$, $x(t)$, and $v(t)$, respectively. Since $x(t)$ and $v(t)$ are uncorrelated and zero mean by assumption, the variance of $Y(f)$ is

$$\phi_Y(f) = E\left[|Y(f)|^2\right] = \phi_X(f) + \phi_V(f), \tag{3}$$

where $E[\cdot]$ denotes mathematical expectation, and $\phi_X(f)$ and $\phi_V(f)$ are the variances of $X(f)$ and $V(f)$, respectively.

B. Conventional Wiener Gains

Traditionally, an estimate of the desired signal, $X(f)$, is obtained by applying a gain, $H(f)$, to the observation, $Y(f)$, i.e.

$$\widehat{X}(f) = Y(f)H(f). \tag{4}$$

The most widely used way to find an optimal gain so that $\widehat{X}(f)$ is a good estimate of $X(f)$ is via the mean-squared error (MSE) criterion given by

$$J_X\left[H(f)\right] = E\left[\left|\widehat{X}(f) - X(f)\right|^2\right]. \tag{5}$$

Minimizing $J_X\left[H(f)\right]$ gives the conventional Wiener gain [1]:

$$H_{\mathrm{W}}(f) = \frac{\phi_X(f)}{\phi_Y(f)} = \frac{\mathrm{iSNR}(f)}{1+\mathrm{iSNR}(f)}, \tag{6}$$

where $\mathrm{iSNR}(f) = \phi_X(f)/\phi_V(f)$ is the narrowband input SNR. It is clear that this gain is always real and $0 \leq H_{\mathrm{W}}(f) \leq 1$. Therefore, the optimal estimate (in the minimum MSE sense) of $X(f)$ and the minimum MSE (MMSE) are, respectively,

$$\widehat{X}_{\mathrm{W}}(f) = H_{\mathrm{W}}(f)Y(f) \tag{7}$$

and

$$J_X\left[H_{\mathrm{W}}(f)\right] = \phi_X(f) - \phi_{\widehat{X}_{\mathrm{W}}}(f), \tag{8}$$

where $\phi_{\widehat{X}_{\mathrm{W}}}(f) = \phi_X^2(f)/\phi_Y(f)$ is the variance of $\widehat{X}_{\mathrm{W}}(f)$.

Alternatively, we can also estimate the noise signal, $V(f)$, by applying a gain, $H'(f)$, to the observation, $Y(f)$, i.e., $\widehat{V}(f) = Y(f)H'(f)$. Using the MSE criterion:

$$J_V\left[H'(f)\right] = E\left[\left|\widehat{V}(f) - V(f)\right|^2\right], \tag{9}$$

we find the optimal gain and estimator:

$$H'_{\mathrm{W}}(f) = \frac{\phi_V(f)}{\phi_Y(f)} = \frac{1}{1+\mathrm{iSNR}(f)} \tag{10}$$

and

$$\widehat{V}_{\mathrm{W}}(f) = H'_{\mathrm{W}}(f)Y(f). \tag{11}$$

The corresponding MMSE is

$$J_V\left[H'_{\mathrm{W}}(f)\right] = \phi_V(f) - \phi_{\widehat{V}_{\mathrm{W}}}(f), \tag{12}$$

where $\phi_{\widehat{V}_{\mathrm{W}}}(f) = \phi_V^2(f)/\phi_Y(f)$ is the variance of $\widehat{V}_{\mathrm{W}}(f)$. Giving the estimate of $V(f)$, we can estimate $X(f)$ as follows:

$$\widehat{X}'_{\mathrm{W}}(f) = Y(f) - \widehat{V}_{\mathrm{W}}(f) = \widehat{X}_{\mathrm{W}}(f). \tag{13}$$

So, the two estimators in (7) and (13) are strictly equivalent here. It can be shown that

$$J_X\left[H_{\mathrm{W}}(f)\right] = J_V\left[H'_{\mathrm{W}}(f)\right] = E\left[\widehat{X}_{\mathrm{W}}(f)\widehat{V}^*_{\mathrm{W}}(f)\right], \tag{14}$$

where the superscript * denotes the complex conjugation. Also, it is interesting to observe that the sum of the estimated speech and noise signals is equal to the observation, i.e.,

$$\widehat{X}_{\mathrm{W}}(f) + \widehat{V}_{\mathrm{W}}(f) = Y(f), \tag{15}$$

which implies that $H_{\rm W}(f) + H'_{\rm W}(f) = 1$. However, the sum of the variances of the estimated speech and noise signals is not equal to the variance of the observation, i.e.,

$$\phi_{\widehat{X}_{\rm W}}(f) + \phi_{\widehat{V}_{\rm W}}(f) = \frac{\phi_X^2(f) + \phi_V^2(f)}{\phi_Y(f)} \neq \phi_Y(f). \quad (16)$$

This is due to the fact that $\widehat{X}_{\rm W}(f)$ and $\widehat{V}_{\rm W}(f)$ are correlated as shown in (14).

C. Constrained Wiener Gains

Let us define the MSE criterion:

$$J\left[H(f), H'(f)\right] = J_X\left[H(f)\right] + J_V\left[H'(f)\right]. \quad (17)$$

The minimization of $J\left[H(f), H'(f)\right]$ without any constraint or with the constraint that $\widehat{X}(f)+\widehat{V}(f) = Y(f)$ [i.e., $H(f)+H'(f) = 1$] leads to $H_{\rm W}(f)$ and $H'_{\rm W}(f)$.

Another interesting possibility is to minimize $J\left[H(f), H'(f)\right]$ with the constraint that the sum of the variances of the estimated speech and noise signals is equal to the variance of the observation, i.e., $\phi_{\widehat{X}}(f)+\phi_{\widehat{V}}(f) = \phi_Y(f)$, or, equivalently, $|H(f)|^2 + |H'(f)|^2 = 1$. By using the Lagrange multiplier technique, we find that the constrained Wiener gains for the estimation of the speech and noise signals are, respectively,

$$H_{\rm cW}(f) = \frac{\phi_X(f)}{\sqrt{\phi_X^2(f) + \phi_V^2(f)}} = \sqrt{\frac{\mathrm{iSNR}^2(f)}{1 + \mathrm{iSNR}^2(f)}}, \quad (18)$$

$$H'_{\rm cW}(f) = \frac{\phi_V(f)}{\sqrt{\phi_X^2(f) + \phi_V^2(f)}} = \sqrt{\frac{1}{1 + \mathrm{iSNR}^2(f)}}. \quad (19)$$

Then, we deduce two different estimators for $X(f)$:

$$\widehat{X}_{\rm cW}(f) = H_{\rm cW}(f)Y(f) \quad (20)$$

and

$$\widehat{X}'_{\rm cW}(f) = Y(f) - \widehat{V}_{\rm cW}(f) = \overline{H}_{\rm cW}(f)Y(f), \quad (21)$$

where

$$\widehat{V}_{\rm cW}(f) = H'_{\rm cW}(f)Y(f) \quad (22)$$

and

$$\overline{H}_{\rm cW}(f) = 1 - H'_{\rm cW}(f). \quad (23)$$

Contrary to the conventional Wiener approach, $\widehat{X}_{\rm cW}(f) \neq \widehat{X}'_{\rm cW}(f)$. It can be verified that

$$E\left[\widehat{X}_{\rm cW}(f)\widehat{V}^*_{\rm cW}(f)\right] \geq E\left[\widehat{X}_{\rm W}(f)\widehat{V}^*_{\rm W}(f)\right] \quad (24)$$

$$\overline{H}_{\rm cW}(f) \leq H_{\rm W}(f) \leq H_{\rm cW}(f). \quad (25)$$

As a consequence, we can state that $\widehat{X}_{\rm cW}(f)$ [resp. $\widehat{X}'_{\rm cW}(f)$] is more (resp. less) noisy but less (resp. more) distorted than $\widehat{X}_{\rm W}(f) = \widehat{X}'_{\rm W}(f)$.

II. Constrained Wiener Filters For Multichannel Noise Reduction

A. Signal Model

Now, we consider the multichannel signal model in which an M-element microphone array captures a convolved source signal in some noise field. The received signals, at the time index t, are expressed as [4], [5]

$$\begin{aligned} y_m(t) &= g_m(t) * s(t) + v_m(t) \\ &= x_m(t) + v_m(t),\ m = 1, 2, \ldots, M, \end{aligned} \quad (26)$$

where $g_m(t)$ is the room impulse response from the unknown speech source, $s(t)$, to the mth microphone, $*$ stands for linear convolution, and $v_m(t)$ is the additive noise at microphone m. We assume that the signals $x_m(t) = g_m(t) * s(t)$ and $v_m(t)$ are uncorrelated and zero mean.

In this section, our desired signal is designated by the clean (but convolved) speech signal received at microphone 1, namely $x_1(t)$ [4], i.e., we attempt to recover $x_1(t)$ given $y_m(t)$, $m = 1, 2, \ldots, M$.

Expression (26) can be written in the frequency domain, at the frequency index f, as

$$Y_m(f) = X_m(f) + V_m(f),\ m = 1, 2, \ldots, M, \quad (27)$$

where $Y_m(f)$, $X_m(f)$and $V_m(f)$ are the frequency-domain representations of $y_m(t)$, $x_m(t)$, and $v_m(t)$, respectively. It is more convenient to write the M frequency-domain microphone signals in a vector notation as

$$\begin{aligned} \mathbf{y}(f) &= \begin{bmatrix} Y_1(f) & Y_2(f) & \cdots & Y_M(f) \end{bmatrix}^T \\ &= \mathbf{x}(f) + \mathbf{v}(f), \end{aligned} \quad (28)$$

where the superscript ^{T} denotes the transpose of a vector or a matrix, and $\mathbf{x}(f)$ and $\mathbf{v}(f)$ are defined in a similar way to $\mathbf{y}(f)$. From (28), we easily obtain the correlation matrix of $\mathbf{y}(f)$:

$$\mathbf{\Phi}_{\mathbf{y}}(f) = \mathbf{\Phi}_{\mathbf{x}}(f) + \mathbf{\Phi}_{\mathbf{v}}(f) \quad (29)$$

where $\mathbf{\Phi}_{\mathbf{x}}(f)$ and $\mathbf{\Phi}_{\mathbf{v}}(f)$ are the correlation matrices of $\mathbf{x}(f)$ and $\mathbf{v}(f)$, respectively.

B. Conventional Wiener Filters

In the multichannel scenario, the desired signal, $X_1(f)$, is usually obtained by applying a complex-valued linear filter, $\mathbf{h}(f)$, of length M, to the observation signal vector, $\mathbf{y}(f)$, i.e., $\widehat{X}_1(f) = \mathbf{h}^H(f)\mathbf{y}(f)$. The minimization of the MSE criterion:

$$J_{X_1}\left[\mathbf{h}(f)\right] = E\left[\left|\widehat{X}_1(f) - X_1(f)\right|^2\right], \quad (30)$$

gives the multichannel Wiener filter [4]:

$$\mathbf{h}_{\rm W}(f) = \mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{x}}(f)\mathbf{i}, \quad (31)$$

where $\mathbf{i}$ is the first column of the $M \times M$ identity matrix, $\mathbf{I}_M$. We deduce that the optimal estimate (in the minimum MSE sense) of $X_1(f)$ and the MMSE are, respectively,

$$\widehat{X}_{1,\rm W}(f) = \mathbf{h}_{\rm W}^H(f)\mathbf{y}(f), \quad (32)$$

$$J_{X_1}\left[\mathbf{h}_{\rm W}(f)\right] = \phi_{X_1}(f) - \phi_{\widehat{X}_{1,\rm W}}(f), \quad (33)$$

where $\phi_{\widehat{X}_{1,\rm W}}(f) = \mathbf{i}^T\mathbf{\Phi}_{\mathbf{x}}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{x}}(f)\mathbf{i}$ is the variance of $\widehat{X}_{1,\rm W}(f)$.

Alternatively, we can also estimate the noise signal at the first microphone, i.e., $V_1(f)$, by applying a complex-valued linear filter, $\mathbf{h}'(f)$, of length M, to the observation signal vector, $\mathbf{y}(f)$, i.e.,

$$\widehat{V}_1(f) = \mathbf{h}'^H(f)\mathbf{y}(f). \tag{34}$$

From the MSE criterion:

$$J_{V_1}[\mathbf{h}'(f)] = E\left[\left|\widehat{V}_1(f) - V_1(f)\right|^2\right], \tag{35}$$

we find that the optimal filter and estimator are, respectively,

$$\mathbf{h}'_{\mathrm{W}}(f) = \mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{i} \tag{36}$$

and

$$\widehat{V}_{1,\mathrm{W}}(f) = \mathbf{h}'^H_{\mathrm{W}}(f)\mathbf{y}(f). \tag{37}$$

The corresponding MMSE is

$$J_{V_1}[\mathbf{h}'_{\mathrm{W}}(f)] = \phi_{V_1}(f) - \phi_{\widehat{V}_{1,\mathrm{W}}}(f), \tag{38}$$

where $\phi_{\widehat{V}_{1,\mathrm{W}}}(f) = \mathbf{i}^T\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{i}$ is the variance of $\widehat{V}_{1,\mathrm{W}}(f)$. Now that we have an optimal estimate of $V_1(f)$, the estimate $X_1(f)$ is obtained as follows:

$$\widehat{X}'_{1,\mathrm{W}}(f) = Y_1(f) - \widehat{V}_{1,\mathrm{W}}(f) = \widehat{X}_{1,\mathrm{W}}(f). \tag{39}$$

So, the two estimators in (32) and (39) are strictly equivalent. It is easy to show that

$$\begin{aligned} J_{X_1}[\mathbf{h}_{\mathrm{W}}(f)] = J_{V_1}[\mathbf{h}'_{\mathrm{W}}(f)] &= E\left[\widehat{X}_{1,\mathrm{W}}(f)\widehat{V}^*_{1,\mathrm{W}}(f)\right] \\ &= \mathbf{i}^T\mathbf{\Phi}_{\mathbf{x}}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{i}. \end{aligned} \tag{40}$$

Also, it is important to see that the sum of the estimated speech and noise signals is equal to the observation at the reference microphone, $\widehat{X}_{1,\mathrm{W}}(f) + \widehat{V}_{1,\mathrm{W}}(f) = Y_1(f)$. As a result, $\mathbf{h}_{\mathrm{W}}(f) + \mathbf{h}'_{\mathrm{W}}(f) = \mathbf{i}$.

However, the sum of the variances of the estimated speech and noise signals is not equal to the variance of the observation at the reference microphone, i.e.,

$$\phi_{\widehat{X}_{1,\mathrm{W}}}(f) + \phi_{\widehat{V}_{1,\mathrm{W}}}(f) \neq \phi_{Y_1}(f). \tag{41}$$

C. Constrained Wiener Filters

Similar to the single-channel case, let us define the following MSE criterion:

$$J[\mathbf{h}(f), \mathbf{h}'(f)] = J_{X_1}[\mathbf{h}(f)] + J_{V_1}[\mathbf{h}'(f)]. \tag{42}$$

It is clear that the minimization of the previous criterion without any constraint leads to $\mathbf{h}_{\mathrm{W}}(f)$ and $\mathbf{h}'_{\mathrm{W}}(f)$. Now, we wish to minimize $J[\mathbf{h}(f), \mathbf{h}'(f)]$ subject to

$$\begin{aligned} \phi_{Y_1}(f) &= \phi_{\widehat{X}_1}(f) + \phi_{\widehat{V}_1}(f) \\ &= \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{h}(f) + \mathbf{h}'^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{h}'(f), \end{aligned} \tag{43}$$

which means that the variance of the observation is equal to the sum of the variances of the estimated speech and noise signal at the reference microphone. Combining this constraint to the MSE criterion with a Lagrange multiplier, we find that the constrained Wiener filters for the estimation of the speech and noise signals are, respectively,

$$\mathbf{h}_{\mathrm{cW}}(f) = \mathcal{S}(f)\mathbf{h}_{\mathrm{W}}(f), \tag{44}$$

$$\mathbf{h}'_{\mathrm{cW}}(f) = \mathcal{S}(f)\mathbf{h}'_{\mathrm{W}}(f), \tag{45}$$

where $\mathcal{S}(f)$ is

$$\sqrt{\frac{\phi_{Y_1}(f)}{\mathbf{i}^T\mathbf{\Phi}_{\mathbf{x}}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{x}}(f)\mathbf{i} + \mathbf{i}^T\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{i}}}.$$

It can be shown that $\mathcal{S}(f) \geq 1$. From these results, we can deduce two different estimators for $X_1(f)$:

$$\widehat{X}_{1,\mathrm{cW}}(f) = \mathbf{h}^H_{\mathrm{cW}}(f)\mathbf{y}(f), \tag{46}$$

$$\widehat{X}'_{1,\mathrm{cW}}(f) = Y_1(f) - \widehat{V}_{1,\mathrm{cW}}(f) = \overline{\mathbf{h}}^H_{\mathrm{cW}}(f)\mathbf{y}(f), \tag{47}$$

where $\widehat{V}_{1,\mathrm{cW}}(f) = \mathbf{h}'^H_{\mathrm{cW}}(f)\mathbf{y}(f)$ and $\overline{\mathbf{h}}_{\mathrm{cW}}(f) = \mathbf{i} - \mathbf{h}'_{\mathrm{cW}}(f)$.

Contrary to the Wiener case, $\widehat{X}_{1,\mathrm{cW}}(f) \neq \widehat{X}'_{1,\mathrm{cW}}(f)$. We can state that $\widehat{X}_{1,\mathrm{cW}}(f)$ [resp. $\widehat{X}'_{1,\mathrm{cW}}(f)$] is more (resp. less) noisy but less (resp. more) distorted than $\widehat{X}_{1,\mathrm{W}}(f) = \widehat{X}'_{1,\mathrm{W}}(f)$.

III. Simulations

In our simulations, the clean signal is partitioned into overlapping frames with a frame size of $K = 256$ and an overlapping factor of 75%. A Kaiser window is then applied to each frame and the windowed frame signal is subsequently transformed into the STFT domain using a 256-point FFT. The clean speech used is recorded in a quiet office room. It is sampled at 8 kHz. The overall length of the signal is approximately 30-s long. The noisy speech is obtained by adding white noise to the clean speech (the noise signal is properly scaled to control the input SNR level).

To compute the traditional and constrained Wiener gains/filters, we need to know $\phi_Y(f)$, $\phi_X(f)$, and $\phi_V(f)$. In this work, an estimate of $\phi_Y(f)$ at the nth frame, which is denoted as $\phi_Y(f,n)$, is computed using a short-time average with samples from the most recent 200-ms signal.

The estimate of the noise variance at the nth frame, i.e., $\phi_V(f,n)$, is obtained using the minima controlled recursive averaging (MCRA) method [6], [7]. Then, $\phi_X(f,n)$ is computed as

$$\phi_X(f,n) = \phi_Y(f,n) - \phi_V(f,n), \tag{48}$$

where $\phi_X(f,n)$ is forced to be zero when negative.

Substituting $\phi_Y(f,n)$, $\phi_X(f,n)$, $\phi_V(f,n)$ into (6), (18), and (23), we implemented the conventional and constrained Wiener gains. The noisy speech spectra is then passed through the Wiener gains to obtain the estimate of the desired signal, i.e.,

$$\widehat{X}(f) = H(f)Y(f) = \widehat{X}_{\mathrm{fd}}(f) + \widehat{V}_{\mathrm{rn}}(f), \tag{49}$$

where $\widehat{X}_{\mathrm{fd}}(f) = H(f)X(f)$ is the filtered desired signal, $\widehat{V}_{\mathrm{rn}}(f) = H(f)V(f)$ is the residual noise. Finally, the inverse FFT (with the overlap add technique) is used to obtain the time-domain signals $\widehat{x}(t)$, $\widehat{x}_{\mathrm{fd}}(t)$, and $\widehat{v}_{\mathrm{rn}}(t)$, which are the time-domain counterparts of $\widehat{X}(f)$, $\widehat{X}_{\mathrm{fd}}(f)$, and $\widehat{V}_{\mathrm{rn}}(f)$, respectively.

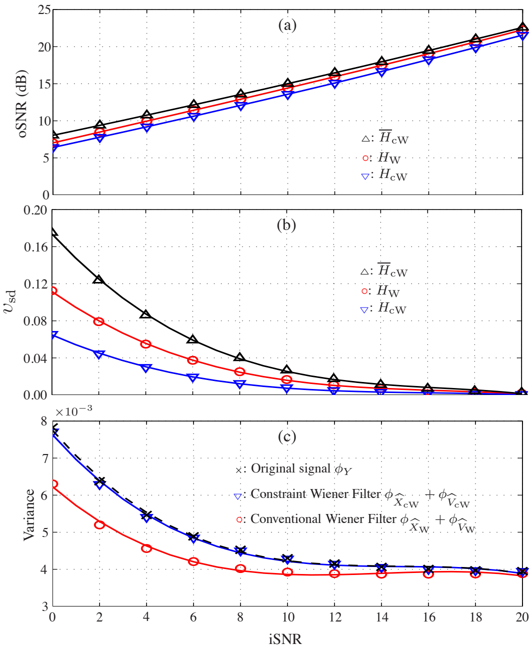

Fig. 1. Performance of the traditional ($H_{\rm W}$) and the two constrained ($\overline{H}_{\rm cW}$ and $H_{\rm cW}$) Wiener gains in the single-channel case as a function of the input SNR in white noise.

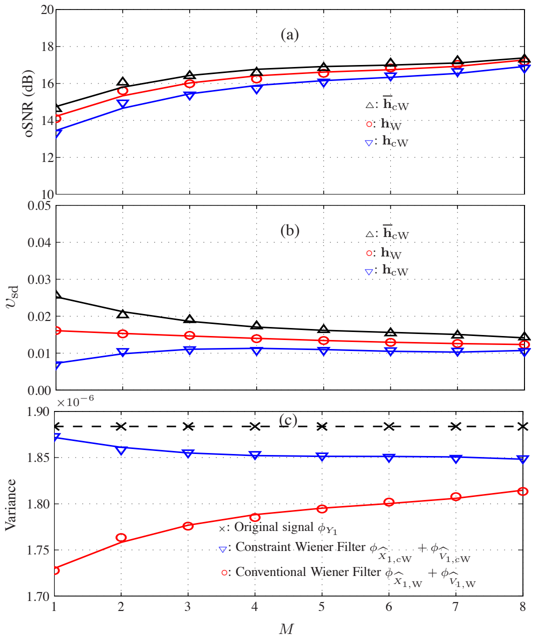

Fig. 2. Performance of the traditional Wiener ($\mathbf{h}_{\rm W}$) and the two constrained ($\overline{\mathbf{h}}_{\rm cW}$ and $\mathbf{h}_{\rm cW}$) Wiener filters as a function of the number of microphones in the multichannel case in white noise with input SNR $= 10$ dB and $T_{60} =$ 380 ms.

We use the output SNR and speech distortion index as the performance criteria [8], which are defined as

$$\text{oSNR} = \frac{E\left[\widehat{x}_{\rm fd}^2(t)\right]}{E\left[\widehat{v}_{\rm rn}^2(t)\right]}, \tag{50}$$

$$\upsilon_{\rm sd} = \frac{E\left\{\left[\widehat{x}_{\rm fd}(t) - x(t)\right]^2\right\}}{E\left[x^2(t)\right]}. \tag{51}$$

We also calculate the variance of the enhanced signal in the time domain with different filters, and then compare the results with that of the noisy signal.

The first simulation assesses the single-channel noise reduction performance as a function of the input SNR, i.e., iSNR. We use the conventional Wiener gain, $H_{\rm W}$, and the two constrained Wiener gains, $H_{\rm cW}$ and $\overline{H}_{\rm cW}$. The results are plotted in Fig. 1. One can see from Fig. 1(a) that the output SNR (oSNR) of the three Wiener gains increases linearly with iSNR. Figure 1(b) shows that the speech distortion index of the Wiener gains decreases with iSNR. In comparison, $\widehat{X}_{\rm cW}(f)$ [resp. $\widehat{X}'_{\rm cW}(f)$] is more (resp. less) noisy but less (resp. more) distorted than $\widehat{X}_{\rm W}(f) = \widehat{X}'_{\rm W}(f)$. Also, one can see from Fig. 1(c) that the variance of the enhanced speech is approximately equal to the variance of the noisy signal with the constrained Wiener gains. However, the variance of the enhanced signal with the traditional Wiener method is much smaller than that of the noisy signal, particularly when the input SNR is low.

The second simulation studies the impact of the number of microphones on the noise reduction performance. The simulation is conducted with the real impulse responses measured in the Varechoic Chamber at Bell Labs [9], [10]. In this simulation, we set $T_{60} = 380$ ms and spatially white noise is added into each microphone with iSNR $= 10$ dB. We use the conventional Wiener filter, $\mathbf{h}_{\rm W}$, and the two constrained Wiener filters, $\mathbf{h}_{\rm cW}$ and $\overline{\mathbf{h}}_{\rm cW}$. The results of this simulation are plotted in Fig. 2. One can see that both the output SNR and the speech distortion index increase with the number of microphones, M. One can also see that $\widehat{X}_{1,\rm cW}(f)$ [resp. $\widehat{X}'_{1,\rm cW}(f)$] is more (resp. less) noisy but less (resp. more) distorted than $\widehat{X}_{1,\rm W}(f) = \widehat{X}'_{1,\rm W}(f)$.

IV. Conclusions

In this paper, we developed two constrained Wiener gains for single-channel noise reduction and two constrained Wiener filters for multichannel noise reduction in the frequency domain. These constrained Wiener gains and filters are deduced by minimizing the mean-squared error (MSE) between the clean speech and the speech estimate with the constraint that the variance of the enhanced signal is equal to that of the noisy signal. In comparison with the traditional Wiener gains and filters, the advantage of the deduced ones is that the volume of the enhanced signal after noise reduction is similar to that of the noisy signal.

Acknowledgment

This work is supported in part by the NSFC "Distinguished Young Scientists Fund" under Grant No. 61425005 and No. 61601370.

REFERENCES

[1] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*, Springer-Verlag, Berlin, Germany, 2009.

[2] J. Benesty and J. Chen, *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study*, Springer Briefs in Electrical and Computer Engineering, 2011.

[3] P. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, Boca Raton, FL, 2007.

[4] J. Chen, J. Benesty, and Y. Huang, *Microphone Array Signal Processing*, Springer-Verlag, Berlin, Germany, 2008.

[5] M. Brandstein and Eds. D. B. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, Germany, 2001.

[6] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *Signal Processing Letters, IEEE*, vol. 9, no. 1, pp. 12–15, Jan 2002.

[7] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 5, pp. 466–475, Sept 2003.

[8] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction wiener filter," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, pp. 1218–1234, July 2006.

[9] R. Kubli, W. Ward, G. Elko, and W. McDougald, "The new varechoic chamber at at&t bell Labs," *Proc. Wallance Clement Sabine Centennial Symposium*, 1994.

[10] A. Härmä, "Acoustic measurement data from the varechoic chamber," *Technical Memorandum, Agere Systems*, Nov 2001.