STUDY OF THE WIDELY LINEAR WIENER FILTER FOR NOISE REDUCTION

Jacob Benesty[1], *Jingdong Chen*[2], *and Yiteng (Arden) Huang*[2]

[1]: INRS-EMT, University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada

[2]: WeVoice, Inc.
9 Sylvan Dr.
Bridgewater, NJ 08807, USA

ABSTRACT

This paper develops a new widely linear noise-reduction Wiener filter based on the variance and pseudo-variance of the short-time Fourier transform coefficients of speech signals. We show that this new noise-reduction filter has many interesting properties, including but not limited to: 1) it causes less speech distortion as compared to the classical noise-reduction Wiener filter; 2) its minimum mean-squared error (MSE) is smaller than that of the classical Wiener filter; 3) it can increase the subband signal-to-noise ratio (SNR), while the classical Wiener filter has no effect on the subband SNR for any given signal frame and subband.

Index Terms— Noise reduction, Wiener filter, widely linear Wiener filter, circularity, noncircularity.

1. INTRODUCTION

Noise reduction is often formulated as a linear filtering problem in the frequency domain. When we work in the frequency domain, we generally deal with complex random variables even though the original time-domain signals are real in the context of speech applications. The main concern, then, is how to design the optimal noise-reduction filters that can fully exploit the different statistics of the complex components obtained via the short-time Fourier transform (STFT). Theoretically, all the different orders of statistics should be considered during the design of the optimal noise-reduction filter. In practice, however, higher-order (higher than 2) statistics are in general difficult to estimate, and as a result, most of today's noise-reduction algorithms only consider the second-order statistics. For a zero-mean complex random variable, there are two basic types of second-order statistics depending on whether the random variable is circular or noncircular.

A complex random variable A is said to be second-order circular if $E\left(A^2\right) = 0$, where $E(\cdot)$ denotes mathematical expectation, $E(AA^*) = E\left(|A|^2\right) \neq 0$, and * denotes complex conjugation. This indicates that the second-order behavior of a circular complex random variable (CCRV) is well described by its variance. Note that the Fourier components of stationary signals are CCRVs [1]. Another powerful aspect of the second-order CCRV is that the classical linear mean-squared estimation technique for real random variables can easily be applied to CCRVs. As a matter of fact, many of the existing frequency-domain noise-reduction filters are derived based on the classical linear mean-squared estimation theory and use only the variance information while assuming that $E\left(A^2\right) = 0$. However, the STFT coefficients of a nonstationary signal like speech are not circular variables (see the example in Fig. 2, Section 4). Many natural questions then arise: is the noncircularity useful for noise reduction? If so, how do we use the noncircularity? How much it can improve noise-reduction performance? This paper attempts to answer these questions. We will show and study how to fully exploit the second-order statistics of a noncircular complex random variable (see [2], [3] for a complete description of the second-order behavior of a complex noncircular random variable) for noise reduction. We will investigate the use of the so-called widely linear (WL) mean-squared estimation theory [4]–[7] to formulate noise-reduction algorithms in the frequency domain and explain the properties of the WL noise-reduction filters.

2. PROBLEM FORMULATION

The noise reduction problem considered in this paper is one of recovering the nonstationary desired signal (clean speech) $x(k)$, k being the discrete-time index, of zero mean from the noisy observation (microphone signal) [8], [9]

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

where $v(k)$ is the unwanted additive noise, which is assumed to be a zero-mean random process (white or colored, stationary or not) and uncorrelated with $x(k)$. In the STFT domain, (1) can be rewritten as

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

where $Y(n,m)$, $X(n,m)$, and $V(n,m)$ are respectively the STFTs of $y(k)$, $x(k)$, and $v(k)$, at time-frame n and frequency-bin (or subband) m (with $m = 0, 1, \ldots, M-1$).

Using the fact that $x(k)$ and $v(k)$ are assumed to be uncorrelated, we can write the variance of the noisy spectral coefficients as

$$\phi_y(n,m) = \phi_x(n,m) + \phi_v(n,m), \tag{3}$$

where $\phi_a(n,m) \triangleq E\left[|A(n,m)|^2\right]$ is the variance of $A(n,m)$, $A(n,m)$ is the STFT coefficients of the signal $a(k)$ at time-frame n and frequency-bin m, and $a \in \{x, v, y\}$.

If $Y(n,m)$ is real, the estimation of $X(n,m)$ can be achieved using the classical techniques, which has already been covered in the rich literature, e.g., [8]–[10]. Here we consider the case where $Y(n,m)$ is complex. In this situation, an estimate of $X(n,m)$ can be obtained using the widely linear (WL) estimation technique as [4]

$$\begin{aligned} Z(n,m) &= H(n,m)Y(n,m) + H'(n,m)Y^*(n,m) \\ &= \mathbf{h}^H(n,m)\mathbf{y}(n,m), \end{aligned} \tag{4}$$

where $Z(n,m)$ is the STFT of the signal $z(k)$ [which is an estimate of $x(k)$], $H(n,m)$ and $H'(n,m)$ are two complex gains, $\mathbf{h}(n,m) \triangleq \left[H^*(n,m) \;\; H'^*(n,m)\right]^T$, $\mathbf{y}(n,m) \triangleq \left[Y(n,m) \;\; Y^*(n,m)\right]^T$, and superscripts ^{H} and ^{T} denote, respectively, transpose conjugate and transpose. If $H'(n,m) = 0$ for any n and m, (4) degenerates to the classical linear estimation theory [9]. This, however, will not happen in general for noncircular complex random variables.

Unlike the classical noise-reduction filters where the signal estimate consists of only the filtered desired signal and residual noise, the signal estimate $Z(n,m)$ in (4) consists of an additional term called interference. To see how the interference occurs, let us introduce three concepts: the circularity quotient [7], the circularity matrix, and the covariance matrix, which are defined, respectively, as

$$\gamma_a(n,m) \triangleq \frac{E\left[A^2(n,m)\right]}{E\left[|A(n,m)|^2\right]}, \tag{5}$$

$$\boldsymbol{\Gamma}_\mathbf{a}(n,m) \triangleq \begin{bmatrix} 1 & \gamma_a(n,m) \\ \gamma_a^*(n,m) & 1 \end{bmatrix}, \tag{6}$$

$$\boldsymbol{\Phi}_\mathbf{a}(n,m) \triangleq E\left[\mathbf{a}(n,m)\mathbf{a}^H(n,m)\right] = \phi_a(n,m)\boldsymbol{\Gamma}_\mathbf{a}(n,m), \tag{7}$$

where $\mathbf{a}(n,m) \triangleq \begin{bmatrix} A(n,m) & A^*(n,m) \end{bmatrix}^T$. It is easy to check that $0 \leq |\gamma_a(n,m)| \leq 1$. Now following the idea proposed in [11], we can decompose $X^*(n,m)$ into two orthogonal components:

$$X^*(n,m) = \gamma_x^*(n,m)X(n,m) + X'(n,m), \tag{8}$$

where

$$X'(n,m) = X^*(n,m) - \gamma_x^*(n,m)X(n,m), \tag{9}$$

$$E\left[X(n,m)X'^*(n,m)\right] = 0, \tag{10}$$

and

$$E\left[\left|X'(n,m)\right|^2\right] = \phi_x(n,m)\left[1 - |\gamma_x(n,m)|^2\right]. \tag{11}$$

We can then rewrite (4) as

$$Z(n,m) = X_{\mathrm{fd}}(n,m) + X'_{\mathrm{ri}}(n,m) + V_{\mathrm{rn}}(n,m), \tag{12}$$

where

$$\begin{aligned} X_{\mathrm{fd}}(n,m) &\triangleq \mathbf{h}^H(n,m)\boldsymbol{\Gamma}_\mathbf{x}(n,m)\mathbf{i}_1X(n,m) \\ &= H(n,m)X(n,m) + \gamma_x^*(n,m)H'(n,m)X(n,m), \end{aligned} \tag{13a}$$

$$X'_{\mathrm{ri}}(n,m) \triangleq \mathbf{h}^H(n,m)\mathbf{i}_2X'(n,m), \tag{13b}$$

$$V_{\mathrm{rn}}(n,m) \triangleq \mathbf{h}^H(n,m)\mathbf{v}(n,m), \tag{13c}$$

are the overall filtered desired signal, the residual interference, and the residual additive noise respectively; $\mathbf{i}_1 \triangleq [1 \quad 0]^T$ and $\mathbf{i}_2 \triangleq [0 \quad 1]^T$. Note that the above decomposition of the signal $X^*(n,m)$ is a key part of this paper in order to be able to properly design and evaluate the optimal noise-reduction filter.

The three terms on the right-hand side of (12) are mutually uncorrelated. Therefore, we have

$$\phi_z(n,m) = \phi_{x_{\mathrm{fd}}}(n,m) + \phi_{x'_{\mathrm{ri}}}(n,m) + \phi_{v_{\mathrm{rn}}}(n,m), \tag{14}$$

where

$$\begin{aligned} \phi_{x_{\mathrm{fd}}}(n,m) &\triangleq E\left[|X_{\mathrm{fd}}(n,m)|^2\right] \\ &= \phi_x(n,m)\mathbf{h}^H(n,m)\boldsymbol{\Gamma}_\mathbf{x}(n,m)\mathbf{i}_1\mathbf{i}_1^H\boldsymbol{\Gamma}_\mathbf{x}(n,m)\mathbf{h}(n,m), \end{aligned} \tag{15a}$$

$$\begin{aligned} \phi_{x'_{\mathrm{ri}}}(n,m) &\triangleq E\left[\left|X'_{\mathrm{ri}}(n,m)\right|^2\right] \\ &= \phi_x(n,m)\left[1 - |\gamma_x(n,m)|^2\right]\mathbf{h}^H(n,m)\mathbf{i}_2\mathbf{i}_2^H\mathbf{h}(n,m), \end{aligned} \tag{15b}$$

$$\begin{aligned} \phi_{v_{\mathrm{rn}}}(n,m) &\triangleq E\left[|V_{\mathrm{rn}}(n,m)|^2\right] \\ &= \mathbf{h}^H(n,m)\boldsymbol{\Phi}_\mathbf{v}(n,m)\mathbf{h}(n,m). \end{aligned} \tag{15c}$$

The objective of noise reduction in the frequency domain is then to find optimal gains $H(n,m)$ and $H'(n,m)$ at each time-frame n and frequency-bin m that would attenuate the noise as much as possible with as little distortion as possible to the desired signal (speech).

3. WIDELY LINEAR WIENER FILTER

We define the subband error signal between the estimated and desired signals as

$$\begin{aligned} \mathcal{E}(n,m) &\triangleq Z(n,m) - X(n,m) \\ &= \mathbf{h}^H(n,m)\mathbf{y}(n,m) - X(n,m). \end{aligned} \tag{16}$$

The subband mean-squared error (MSE) is then

$$\begin{aligned} J\left[\mathbf{h}(n,m)\right] &\triangleq E\left[|\mathcal{E}(n,m)|^2\right] \\ &= \phi_x(n,m)\left|\mathbf{h}^H(n,m)\boldsymbol{\Gamma}_\mathbf{x}(n,m)\mathbf{i}_1 - 1\right|^2 \\ &+ \phi_{x'_{\mathrm{ri}}}(n,m) + \phi_{v_{\mathrm{rn}}}(n,m). \end{aligned} \tag{17}$$

Taking the gradient of the subband MSE, $J\left[\mathbf{h}(n,m)\right]$, with respect to $\mathbf{h}^H(n,m)$ and equating the result to zero give us the WL Wiener filter:

$$\begin{aligned} \mathbf{h}_{\mathrm{WLW}}(n,m) &= \boldsymbol{\Phi}_\mathbf{y}^{-1}(n,m)\boldsymbol{\Phi}_\mathbf{x}(n,m)\mathbf{i}_1 \\ &= \frac{\phi_x(n,m)}{\phi_y(n,m)}\cdot\boldsymbol{\Gamma}_\mathbf{y}^{-1}(n,m)\boldsymbol{\Gamma}_\mathbf{x}(n,m)\mathbf{i}_1 \\ &= \left[\mathbf{I} - \frac{\phi_v(n,m)}{\phi_y(n,m)}\cdot\boldsymbol{\Gamma}_\mathbf{y}^{-1}(n,m)\boldsymbol{\Gamma}_\mathbf{v}(n,m)\right]\mathbf{i}_1. \end{aligned} \tag{18}$$

It follows immediately that

$$H_{\mathrm{WLW}}(n,m) = \frac{1 - \gamma_x(n,m)\gamma_y^*(n,m)}{1 - |\gamma_y(n,m)|^2}\cdot\frac{\phi_x(n,m)}{\phi_y(n,m)}, \tag{19a}$$

$$H'_{\mathrm{WLW}}(n,m) = \frac{\gamma_x(n,m) - \gamma_y(n,m)}{1 - |\gamma_y(n,m)|^2}\cdot\frac{\phi_x(n,m)}{\phi_y(n,m)}. \tag{19b}$$

From the definitions of variance and circularity quotient, one can easily verify the following relation:

$$\begin{aligned} \gamma_y(n,m)\phi_y(n,m) &= \gamma_x(n,m)\phi_x(n,m) \\ &+ \gamma_v(n,m)\phi_v(n,m). \end{aligned} \tag{20}$$

By using (20), the WL Wiener complex gains in (19) can be rearranged as

$$H_{\mathrm{WLW}}(n,m) = 1 - \frac{1 - \gamma_v(n,m)\gamma_y^*(n,m)}{1 - |\gamma_y(n,m)|^2}\cdot\frac{\phi_v(n,m)}{\phi_y(n,m)}, \tag{21a}$$

$$H'_{\mathrm{WLW}}(n,m) = \frac{\gamma_y(n,m) - \gamma_v(n,m)}{1 - |\gamma_y(n,m)|^2}\cdot\frac{\phi_v(n,m)}{\phi_y(n,m)}. \tag{21b}$$

We recall that the classical Wiener filter [9] is

$$H_{\mathrm{W}}(n,m) = \frac{\phi_x(n,m)}{\phi_y(n,m)} = 1 - \frac{\phi_v(n,m)}{\phi_y(n,m)}. \tag{22}$$

Of course, taking $\gamma_x(n,m) = \gamma_v(n,m) = 0$ in the WL Wiener filter, we obtain the classical Wiener filter. While the Wiener filter is always real, the WL Wiener filter is, in general, complex.

Now let us examine the minimum MSE for the WL Wiener filter. The subband minimum MSE is found by substituting the WL Wiener filter given in (18) into (17):

$$\begin{aligned} J\left[\mathbf{h}_{\mathrm{WLW}}(n,m)\right] &= \phi_x(n,m)\cdot \\ &\left[1 - \frac{\phi_x(n,m)}{\phi_y(n,m)}\cdot\mathbf{i}_1^H\boldsymbol{\Gamma}_\mathbf{x}(n,m)\boldsymbol{\Gamma}_\mathbf{y}^{-1}(n,m)\boldsymbol{\Gamma}_\mathbf{x}(n,m)\mathbf{i}_1\right]. \end{aligned} \tag{23}$$

The subband MSE for the classical Wiener filter is

$$J\left[\mathbf{h}_{\mathrm{W}}(n,m)\right] = \phi_x(n,m)\left[1 - \frac{\phi_x(n,m)}{\phi_y(n,m)}\right]. \tag{24}$$

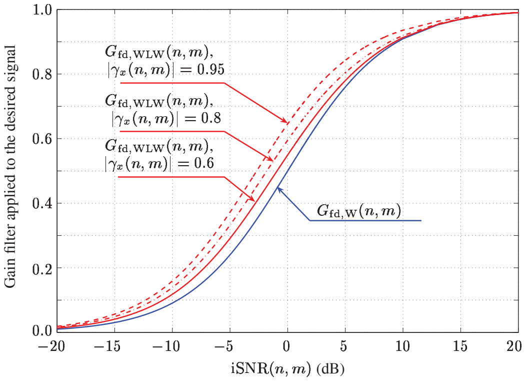

Fig. 1. Comparison between the WL and classical Wiener filters for their gain applied to filter the desired speech signal.

It is easy to check that

$$J\left[\mathbf{h}_{\text{WLW}}(n,m)\right] \le J\left[\mathbf{h}_{\text{W}}(n,m)\right], \quad \forall n,m. \tag{25}$$

So, the subband minimum MSE of the WL Wiener filter is smaller than that of the classical Wiener filter, which shows one advantage of the WL Wiener filter.

Now let us take a slightly different angle to study the WL Wiener filter by examining the filtered desired signal and residual noise and interference. Since in most situations noise is relatively more stationary than speech, it is reasonable to assume that $\gamma_v(n,m) = 0$. In this case, if substituting (19) into (13a) and using some simple mathematical manipulation, we can deduce the filtered desired signal due to the WL Wiener filter as

$$X_{\text{fd,WLW}}(n,m) = G_{\text{fd,WLW}} X(n,m), \tag{26}$$

where

$$G_{\text{fd,WLW}}(n,m) = H_{\text{WLW}}(n,m) + \gamma_x^*(n,m) H'_{\text{WLW}}(n,m) \tag{27}$$
$$= \frac{1 + \dfrac{1-\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)} \cdot |\gamma_x(n,m)|^2}{1 - \left[\dfrac{\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)}\right]^2 \cdot |\gamma_x(n,m)|^2} \cdot \frac{\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)},$$

and

$$\text{iSNR}(n,m) \triangleq \frac{\phi_x(n,m)}{\phi_v(n,m)} \tag{28}$$

is the subband input SNR at frame n and frequency-bin m.

Recall that for the classical Wiener filter, the gain filter applied to the desired speech is

$$G_{\text{fd,W}}(n,m) = H_{\text{W}}(n,m) = \frac{\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)}. \tag{29}$$

It can be checked that $1 \ge G_{\text{fd,WLW}}(n,m) \ge G_{\text{fd,W}}(n,m) \ge 0$, meaning that the WL Wiener filter introduces less speech distortion. Figure 1 plots both the $G_{\text{fd,WLW}}(n,m)$ and $G_{\text{fd,W}}(n,m)$ as a function of iSNR(n,m). It is seen that both the $G_{\text{fd,WLW}}(n,m)$ and $G_{\text{fd,W}}(n,m)$ increase with the subband input SNR. So, less speech distortion is added to the enhanced signal by both the WL and classical Wiener filters as the subband input SNR increases. It is also seen that $G_{\text{fd,WLW}}(n,m)$ increases as $|\gamma_x(n,m)|$ increases, which can be easily checked from (27). Therefore, the more the desired signal is noncircular, the less is the signal distortion caused by the WL Wiener filter. However, when SNR is either very large (e.g., > 15 dB) or very small (e.g., < -15 dB), $G_{\text{fd,WLW}}(n,m)$ converges to $G_{\text{fd,W}}(n,m)$ regardless of the degree of speech noncircularity. In this case, both the WL Wiener and classical Wiener have a similar amount of speech distortion. We also notice from Fig. 1 that a significant degree of noncircularity is needed in order for the WL Wiener filter to have noticeable less speech distortion than the classical Wiener filter. For instance, when iSNR$(n,m) = 5$ dB, if we want the WL Wiener filter to have 5% less speech distortion than the classical Wiener filter, we would need $|\gamma_x(n,m)| \ge 0.76$.

Following the same line of analysis, we can compare the WL and classical Wiener filters for their noise reduction performance, which will not be presented here due to the space limit.

Having shown that the WL Wiener filter introduces less speech distortion and has a smaller minimum MSE than the classical Wiener filter, we now analyze the subband output SNR of the WL Wiener filter. We have the following theorem.

Theorem: With the WL Wiener filter given in (18), the subband output SNR is always greater than or at least equal to the subband input SNR, i.e., oSNR $\left[\mathbf{h}_{\text{WLW}}(n,m)\right] \ge$ iSNR(n,m), $\forall n,m$.

Taking into account the interference component, we can write the output subband SNR of a WL filter as

$$\text{oSNR}\left[\mathbf{h}(n,m)\right] \triangleq \frac{\phi_{x_{\text{fd}}}(n,m)}{\phi_{x'_{\text{ri}}}(n,m) + \phi_{v_{\text{rn}}}(n,m)}. \tag{30}$$

Applying (15) and (19) to (30) and using some mathematical manipulation, we can show that oSNR $\left[\mathbf{h}_{\text{WLW}}(n,m)\right] \ge \phi_x(n,m)/\phi_v(n,m) =$ iSNR(n,m). The detailed proof is not presented here due to the space limit. Recall that for the classical Wiener filter, the input and output subband SNRs are equal, i.e.,

$$\text{oSNR}\left[\mathbf{h}_{\text{W}}(n,m)\right] = \frac{\phi_x(n,m)}{\phi_v(n,m)} = \text{iSNR}(n,m). \tag{31}$$

So, the classical Wiener filter cannot improve the subband SNR. But the WL Wiener filter can improve the subband SNR, which, again, shows the advantage of the WL Wiener filter over the classical Wiener filter.

4. EXPERIMENTS

The clean speech used in this experiment is from the TIMIT database [12], which was designed to provide speech data for acoustic-phonetic studies and for the development and evaluation of automatic speech recognition (ASR) systems. Each speech signal in this database is recorded using a 16-kHz sampling rate and is accompanied by manually segmented phonetic (based on 61 phonemes) transcripts. In this experiment, we took one signal from the speaker FAKS0 and downsampled it into 8 kHz. This signal is then used as the clean speech. Figure 2 (the upper trace) plots this signal and also visualizes both the phonetic transcription and phoneme boundaries. The corresponding noisy signal is generated by adding white Gaussian noise into the clean speech with different SNRs.

To perform noise reduction in the frequency domain, the input speech signal is partitioned into overlapping frames with a frame width of 8 ms 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 frequency domain using a 64-point FFT. At each subband and for each phoneme, a short-time sample average is used to replace the mathematical expectation to compute the variance parameters and circularity quotients. Note that the parameters $\phi_x(n,m)$ and $\gamma_x(n,m)$, $\phi_y(n,m)$ and $\gamma_y(n,m)$, and $\phi_v(n,m)$ and $\gamma_v(n,m)$ are directly computed from the clean speech, the noisy, and the noise signals respectively. Figure 2 (the

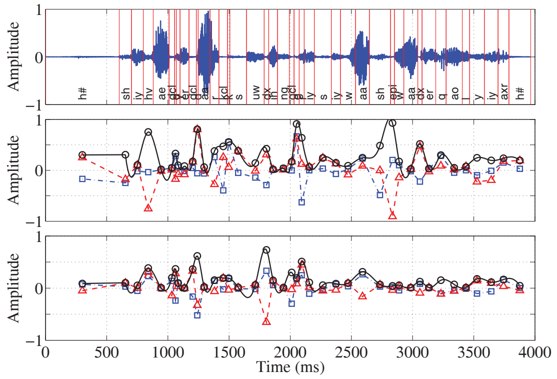

Fig. 2. A speech signal selected from the TIMIT database and the corresponding $\gamma_x(n,m)$ estimated with a short-time sample average. The upper trace: waveform with phoneme labeling and phoneme boundaries. The middle trace: the real(□), imaginary (△), and magnitude (○) parts of $\gamma_x(n,3)$ estimated with a short-time sample average. The lower trace: the real(□), imaginary (△), and magnitude (○) of $\gamma_x(n,6)$ estimated with a short-time sample average.

lower two traces) shows the estimated γ_x at the third and sixth subbands. It is clearly seen that $\gamma_x(n,m)$ is not equal to zero, which illustrates that the complex STFT coefficients of speech are not circular variables.

With the computed variance and noncircularity parameters, we constructed a WL Wiener filter for each phoneme at each subband according to (21). After passing the noisy speech spectrum through the constructed WL Wiener filter, the inverse STFT (with overlap add) is used to obtain the time-domain speech estimate. For the purpose of comparison, we also constructed the classical Wiener filter [eq. (22)] using the estimated variance parameters. To assess the noise-reduction performance of the WL and classical Wiener filters, we evaluate two objective measures: the speech-distortion index and the output SNR. The speech-distortion index is defined as [9]

$$\upsilon_{\mathrm{sd}}(\mathbf{h}) \triangleq \frac{\sum_n \sum_{m=0}^{M-1} E\left[|X_{\mathrm{fd}}(n,m) - X(n,m)|^2\right]}{\sum_n \sum_{m=0}^{M-1} \phi_x(n,m)}. \tag{32}$$

For the classical Wiener filter, its output consists of two components: the filtered desired speech and residual noise. The output SNR can be defined as

$$\mathrm{oSNR}(\mathbf{h}_{\mathrm{W}}) \triangleq \frac{\sum_n \sum_{m=0}^{M-1} \phi_{x_{\mathrm{fd}}}(n,m)}{\sum_n \sum_{m=0}^{M-1} \phi_{v_{\mathrm{rn}}}(n,m)}. \tag{33}$$

While for the WL Wiener filter, since its output also consists of an interference component, we define its output SNR as

$$\mathrm{oSNR}(\mathbf{h}_{\mathrm{WLW}}) \triangleq \frac{\sum_n \sum_{m=0}^{M-1} \phi_{x_{\mathrm{fd}}}(n,m)}{\sum_n \sum_{m=0}^{M-1} \left[\phi_{x'_{\mathrm{ri}}}(n,m) + \phi_{v_{\mathrm{rn}}}(n,m)\right]}. \tag{34}$$

Table 1 presents the results. It is seen from Table 1 that the measured speech-distortion index for the WL Wiener filter is smaller than that of the classical Wiener filter. This coincides with the theoretical analysis that the WL Wiener filter introduces less speech distortion to the desired signal. The two filters have achieved similar SNR improvement, which is somehow unexpected. The underlying reason, we suspect, could be due to the noncircularity estimation. We only compute one noncircularity quotient for each phoneme at each subband. Since speech is nonstationary and time varying, its statistics may change significantly even within one phoneme. So, the short-time average method may not necessarily be a good or reliable approach to estimating the noncircularity.

Table 1. Performance of the classical and WL Wiener filters in white Gaussian noise.

Performance Measure	NR filter	iSNR (Input SNR)				
		-10 dB	-5 dB	0 dB	5 dB	10 dB
Speech distortion υ_{sd}	WF	0.213	0.097	0.047	0.020	0.009
	WL	0.205	0.093	0.045	0.019	0.008
Output SNR oSNR (dB)	WF	5.5	8.2	11.3	14.3	17.4
	WL	5.6	8.2	11.3	14.3	17.4

5. CONCLUSIONS

When we work with the STFT coefficients in the frequency domain, we generally deal with complex random variables even though the original time-domain signals are real in the context of speech applications. A complex random variable can be either (second-order) circular or noncircular depending on whether its pseudo-variance is zero or not. Traditionally, the STFT coefficients of speech are assumed to be circular and most noise-reduction approaches design the noise-reduction filter based only on the variance of the STFT coefficients (or power spectra) of the noise and noisy signals. In this paper, we have illustrated that the STFT coefficients of speech are in general noncircular variables because speech signals are highly nonstationary. Based on the noncircularity, we have deduced a WL noise-reduction Wiener filter. We have shown through theoretical analysis that the WL Wiener filter introduces less distortion to the desired speech signal and has a smaller minimum MSE as compared to the classical Wiener filter. Most importantly, the WL Wiener filter can improve the subband SNR, which is different from the classical Wiener filter that does not change the subband SNR for any given frame and subband.

6. REFERENCES

[1] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, pp. 3473–3482, Dec. 1994.

[2] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1293–1302, July 1993.

[3] P. J. Schreier and L. L. Scharf, "Second-order anaysis of improper complex random vectors and processes," *IEEE Trans. Signal Process.*, vol. 51, pp. 714–725, Mar. 2003.

[4] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, pp. 2030–2033, Aug. 1995.

[5] J. Eriksson, E. Ollila, and V. Koivunen, "Statistics for complex random variables revisited," in *Proc. IEEE ICASSP*, 2009, pp. 3565–3568.

[6] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Wiley, 2009.

[7] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, 2008.

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

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

[11] P. Chevalier, J.-P. Delmas, and A. Oukaci, "Optimal widely linear MVDR beamforming for noncircular signals," in *Proc. IEEE ICASSP*, 2009, pp. 3573–3576.

[12] K.-F. Lee and H.-W. Hon, "Speaker-independent phone recognition using hidden Markov models," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 1641–1648, Nov. 1989.