Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 52 (2010) 427–439
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

On widely linear Wiener and tradeoff filters for noise reduction

Jacob Benesty [a], Jingdong Chen [b,*], Yiteng (Arden) Huang [b]

[a] INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, Quebec, Canada H5A 1K6
[b] WeVoice, Inc., 9 Sylvan Dr., Bridgewater, NJ 08807, USA

Received 13 August 2009; received in revised form 25 November 2009; accepted 3 February 2010

Abstract

Noise reduction is often formulated as a linear filtering problem in the frequency domain. With this formulation, the core issue of noise reduction becomes how to design an optimal frequency-domain filter that can significantly suppress noise without introducing perceptually noticeable speech distortion. While higher-order information can be used, most existing approaches use only second-order statistics to design the noise-reduction filter because they are relatively easier to estimate and are more reliable. When we transform nonstationary speech signals into the frequency domain and work with the short-time discrete Fourier transform coefficients, there are two types of second-order statistics, i.e., the variance and the so-called pseudo-variance due to the noncircularity of the signal. So far, only the variance information has been exploited in designing different noise-reduction filters while the pseudo-variance has been neglected. In this paper, we attempt to shed some light on how to use noncircularity in the context of noise reduction. We will discuss the design of optimal and suboptimal noise reduction filters using both the variance and pseudo-variance and answer the basic question whether noncircularity can be used to improve the noise-reduction performance.

Keywords: Noise reduction; Wiener filter; Widely linear Wiener filter; Circularity; Noncircularity

1. Introduction

Noise reduction, which aims at estimating the desired clean speech signal from noisy observations, is a very important problem and has attracted a significant amount of research and engineering attention over the past few decades (Benesty et al., 2005, 2009; Loizou, 2007; Vary and Martin, 2006; Huang et al., 2006). Typically, the noise-reduction process is formulated as a filtering problem where the clean speech estimate is obtained by passing the noisy speech through a noise-reduction filter. With such a formulation, the core issue of noise reduction becomes how to design an optimal filter that can fully exploit the speech and noise statistics to achieve maximum noise suppression without introducing perceptually noticeable speech distortion. While good filters can be designed in the time domain, most widely used approaches so far work in the frequency domain. The reason for working in the frequency domain are multiple, including (but not limited to): (1) the filtering process can be implemented very efficiently thanks to the fast Fourier transform; (2) the filters at different frequencies (or frequency bands) can be designed and handled independently of each other, which offers tremendous flexibility in dealing with colored noise; and (3) most of our knowledge and understanding of speech production and perception is related to frequencies, so in the frequency domain, our knowledge can be easily used to help optimize noise-reduction performance.

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

* Corresponding author. Address: 9 Iroquois Trail, Branchburg, NJ 08876, USA. Tel.: +1 908 722 5945.
E-mail addresses: benesty@emt.inrs.ca (J. Benesty), jingdongchen@ieee.org (J. Chen), ardenhuang@gmail.com (Y. (Arden) Huang).

0167-6393/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2010.02.003

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 consider only 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 circular if its probability density function (PDF) is the same as the PDF of Ae^{jr} (Amblard et al., 1996a, Amblard et al., 1996b), where j and r are the imaginary unit ($j^2 = -1$) and any real number, respectively. This is equivalent to saying that the PDF of a circular complex random variable (CCRV) is a function of the product AA^* only (Amblard et al., 1996a), where $*$ denotes complex conjugation. An important consequence of this is that the only nonnull moments and cumulants of a CCRV are the moments and cumulants constructed with the same power in A and A^* (Amblard et al., 1996a). Now let us confine our discussion and study to the second-order issues. With the general definition of circularity, we can readily define the second-order circularity: a zero-mean complex random variable A is said to be second-order circular if its pseudo-variance is equal to zero, i.e., $E(A^2) = 0$, where $E(\cdot)$ denotes mathematical expectation and $E(AA^*) = E(|A|^2) \neq 0$. This indicates that the second-order behavior of a CCRV is well described by its variance. Note that the Fourier components of stationary signals are CCRVs (Picinbono et al., 1994). Another powerful aspect of the second-order CCRV is that the classical linear estimation theory for real random variables can easily be applied to CCRVs. As a matter of fact, most of the existing frequency-domain noise-reduction filters are derived based on the classical mean-squared estimation approach and use only the variance information while assuming that $E(A^2) = 0$.

However, the STFT coefficients of a nonstationary signal like speech are not circular variables. To illustrate this, we take a speech signal that is recorded from a female speaker with an 8-kHz sampling rate and a 16-bit quantization and partition it into overlapping frames. The overlapping factor is 75% and the frame length is 8 ms. Each frame is then transformed into the frequency domain using a 64-point FFT. For each frequency band (except the 1st and 33rd bands where the coefficients are real), we treat the coefficients as a complex random variable (for ease of exposition, let us use A to denote this random variable) and estimate its variance and pseudo-variance. Because speech is nonstationary, we cannot simply replace the mathematical expectation with a sample average. Instead, we use the recursive estimator given in Eq. (88) of (Chen et al., 2006) to estimate both the variance and pseudo-variance (more discussion on how to estimate the variance and pseudo-variance parameters will be given in Section 7).

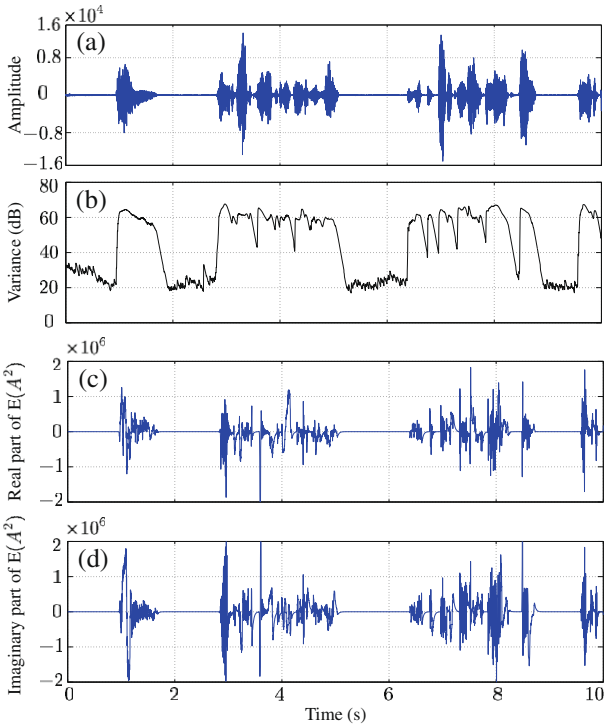

Fig. 1. Illustration of the noncircularity of the STFT coefficients of a speech signal: (a) a speech signal; (b) the $E(|A|^2)$ estimate; (c) the real part of the $E(A^2)$ estimate; and (d) the imaginary part of the $E(A^2)$ estimate.

Fig. 1 plots the estimation results for the 2nd frequency band. It is clearly seen that the pseudo-variance $E(A^2)$ of the STFT coefficients of the speech signal are not zero, so STFT coefficients of speech signals are noncircular random variables. 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 study and show how to fully exploit the second-order statistics of a noncircular complex random variable (see Neeser and Massey, 1993; Schreier and Scharf, 2003 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 (Picinbono et al., 1995; Eriksson et al., 2009; Mandic and Goh, 2009; Ollila, 2008) to formulate noise-reduction algorithms in the frequency domain and explain the benefits that can be achieved with this new formulation.

The rest of this paper is organized as follows. In Section 2, we formulate the single-channel noise reduction problem in the STFT domain and give some useful definitions and explanations that will be of great help for the rest of the paper. Section 3 explains the different performance measures for noise reduction with WL estimation. In Section 4, we write the WL mean-squared error (MSE), which is a simple and powerful tool for deriving the different optimal WL filters. In Section 5, we derive the WL Wiener filter and explains its differences from the classical Wiener filter. Section 6 deals with the WL and classical tradeoff filters.

Section 7 presents some experiments confirming the theoretical derivations. Finally, we give our conclusions in Section 8.

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) (Benesty et al., 2009; Chen et al., 2006)

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

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 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

$$\phi_a(n,m) \triangleq E\left[|A(n,m)|^2\right] \tag{4}$$

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 (Benesty et al., 2005, 2009; Loizou, 2007; Vary and Martin, 2006; Huang et al., 2006). 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 (Picinbono et al., 1995)

$$\begin{aligned} Z(n,m) &= H(n,m)Y(n,m) + H'(n,m)Y^*(n,m) \\ &= \mathbf{h}^H(n,m)\mathbf{y}(n,m), \end{aligned} \tag{5}$$

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, superscript H denotes transpose conjugate, $*$ denotes complex conjugation as already defined in Section 1, and

$$\mathbf{h}(n,m) \triangleq \begin{bmatrix} H^*(n,m) \\ H'^*(n,m) \end{bmatrix},$$

$$\mathbf{y}(n,m) \triangleq \begin{bmatrix} Y(n,m) \\ Y^*(n,m) \end{bmatrix}.$$

If $H'(n,m) = 0$ for any n and m, (5) degenerates to the classical linear estimation theory (Benesty et al., 2009). This, however, will not happen in general for noncircular complex random variables. With the signal model given in (2), we can rewrite (5) as

$$Z(n,m) = X_{\mathrm{f}}(n,m) + V_{\mathrm{rn}}(n,m), \tag{6}$$

where

$$X_{\mathrm{f}}(n,m) \triangleq \mathbf{h}^H(n,m)\mathbf{x}(n,m),$$

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

are, respectively, the filtered version of the desired signal and its complex conjugate, and the residual noise. Vectors $\mathbf{x}(n,m)$ and $\mathbf{v}(n,m)$ are defined in a similar way to $\mathbf{y}(n,m)$. Since $X(n,m)$ and $V(n,m)$ are uncorrelated by assumption, so are $X_{\mathrm{f}}(n,m)$ and $V_{\mathrm{rn}}(n,m)$. From (6), we deduce the variance of the spectral coefficients of the signal $z(k)$ at time-frame n and frequency-bin m:

$$\phi_z(n,m) = \phi_{x_{\mathrm{f}}}(n,m) + \phi_{v_{\mathrm{rn}}}(n,m), \tag{7}$$

where

$$\phi_{x_{\mathrm{f}}}(n,m) \triangleq E\left[|X_{\mathrm{f}}(n,m)|^2\right] = \mathbf{h}^H(n,m)\mathbf{\Phi}_{\mathbf{x}}(n,m)\mathbf{h}(n,m), \tag{8}$$

$$\phi_{v_{\mathrm{rn}}}(n,m) \triangleq E\left[|V_{\mathrm{rn}}(n,m)|^2\right] = \mathbf{h}^H(n,m)\mathbf{\Phi}_{\mathbf{v}}(n,m)\mathbf{h}(n,m), \tag{9}$$

and

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{a}}(n,m) &\triangleq E\left[\mathbf{a}(n,m)\mathbf{a}^H(n,m)\right] \\ &= \phi_a(n,m)\begin{bmatrix} 1 & \gamma_a(n,m) \\ \gamma_a^*(n,m) & 1 \end{bmatrix} = \phi_a(n,m)\mathbf{\Gamma}_{\mathbf{a}}(n,m) \end{aligned} \tag{10}$$

is the covariance matrix of $\mathbf{a}(n,m) = [A(n,m) \quad A^*(n,m)]^T$ with

$$\gamma_a(n,m) \triangleq \frac{E\left[A^2(n,m)\right]}{E\left[|A(n,m)|^2\right]} \tag{11}$$

being the (second-order) circularity quotient (Ollila, 2008) and $\mathbf{\Gamma}_{\mathbf{a}}(n,m)$ being the circularity matrix. It can easily be shown that (Ollila, 2008)

$$0 \leqslant |\gamma_a(n,m)| \leqslant 1. \tag{12}$$

The circularity coefficient $|\gamma_a(n,m)|$ conveys information about the degree of circularity of the signal $A(n,m)$. In particular, if $A(n,m)$ is a (second-order) CCRV then $\gamma_a(n,m) = 0$ and $\mathbf{\Gamma}_{\mathbf{a}}(n,m) = \mathbf{I}$, where

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = [\mathbf{i}_1 \quad \mathbf{i}_2] \tag{13}$$

is the 2×2 identity matrix.

The signal $X_{\mathrm{f}}(n,m)$ consists of components from both the desired signal $X(n,m)$ and its conjugate. But not all these components are what we want. It is, therefore, necessary and important to distinguish between the filtered desired signal and the residual interference that both may exist in $X_{\mathrm{f}}(n,m)$ at the same time. Specifically, $H(n,m)X(n,m)$ is

part of the overall filtered desired signal, but $H'(n,m)X^*(n,m)$ is not. If $\gamma_x(n,m)=0$ for any n and m, $X(n,m)$ and $X^*(n,m)$ are uncorrelated and the overall filtered desired signal is indeed $H(n,m)X(n,m)$. But for $\gamma_x(n,m)\neq 0$, $X^*(n,m)$ is correlated with $X(n,m)$ and contains both the desired signal and an interference component. Following the idea proposed in (Chevalier et al., 2009), we can decompose $X^*(n,m)$ into two orthogonal components:

$$X^*(n,m)=\gamma_x^*(n,m)X(n,m)+X'(n,m), \tag{14}$$

where

$$X'(n,m)=X^*(n,m)-\gamma_x^*(n,m)X(n,m), \tag{15}$$

$$E\left[X(n,m)X'^*(n,m)\right]=0, \tag{16}$$

and

$$E\left[\left|X'(n,m)\right|^2\right]=\phi_x(n,m)\left[1-\left|\gamma_x(n,m)\right|^2\right]. \tag{17}$$

We can then rewrite (6) as

$$Z(n,m)=X_{\mathrm{fd}}(n,m)+X'_{\mathrm{ri}}(n,m)+V_{\mathrm{rn}}(n,m), \tag{18}$$

where

$$\begin{aligned}X_{\mathrm{fd}}(n,m)&\triangleq \mathbf{h}^H(n,m)\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1X(n,m)\\&=H(n,m)X(n,m)+\gamma_x^*(n,m)H'(n,m)X(n,m),\end{aligned} \tag{19a}$$

$$X'_{\mathrm{ri}}(n,m)\triangleq \mathbf{h}^H(n,m)\mathbf{i}_2X'(n,m), \tag{19b}$$

$$V_{\mathrm{rn}}(n,m)\triangleq \mathbf{h}^H(n,m)\mathbf{v}(n,m), \tag{19c}$$

are, respectively, the overall filtered desired signal, the residual interference, and the residual additive noise. 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 define the different performance measures and design different noise-reduction filters.

The three terms of the right-hand side of (18) are mutually uncorrelated. Therefore, we have

$$\phi_z(n,m)=\phi_{x_{\mathrm{fd}}}(n,m)+\phi_{x'_{\mathrm{ri}}}(n,m)+\phi_{v_{\mathrm{rn}}}(n,m), \tag{20}$$

where

$$\begin{aligned}\phi_{x_{\mathrm{fd}}}(n,m)&\triangleq E\left[\left|X_{\mathrm{fd}}(n,m)\right|^2\right]\\&=\phi_x(n,m)\mathbf{h}^H(n,m)\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1\mathbf{i}_1^H\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{h}(n,m),\end{aligned} \tag{21}$$

$$\begin{aligned}\phi_{x'_{\mathrm{ri}}}(n,m)&\triangleq E\left[\left|X'_{\mathrm{ri}}(n,m)\right|^2\right]\\&=\phi_x(n,m)\left[1-\left|\gamma_x(n,m)\right|^2\right]\mathbf{h}^H(n,m)\mathbf{i}_2\mathbf{i}_2^H\mathbf{h}(n,m),\end{aligned} \tag{22}$$

and $\phi_{v_{\mathrm{rn}}}(n,m)$ is defined in (9).

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. Performance measures

In this section, we present some useful measures that are necessary to properly design optimal gains for noise reduction and help us better understand their behaviors. Most of these measures are directly deduced from (20).

The (fullband) input signal-to-noise ratio (SNR) is defined as the ratio of the intensity of the signal of interest (speech) over the intensity of the background noise. With the signal model given in (2), the input SNR can be written as

$$\mathrm{iSNR}\triangleq\frac{\sum_n\sum_{m=0}^{M-1}\phi_x(n,m)}{\sum_n\sum_{m=0}^{M-1}\phi_v(n,m)}. \tag{23}$$

We also define the segmental and subband input SNRs, respectively, as (Benesty et al., 2009)

$$\mathrm{iSNR}(n)\triangleq\frac{\sum_{m=0}^{M-1}\phi_x(n,m)}{\sum_{m=0}^{M-1}\phi_v(n,m)}, \tag{24}$$

$$\mathrm{iSNR}(n,m)\triangleq\frac{\phi_x(n,m)}{\phi_v(n,m)},\quad m=0,1,\ldots,M-1. \tag{25}$$

After noise reduction with the model given in (18), the fullband, segmental, and subband output SNRs[1] are

$$\mathrm{oSNR}(\mathbf{h})\triangleq\frac{\sum_n\sum_{m=0}^{M-1}\phi_{x_{\mathrm{fd}}}(n,m)}{\sum_n\sum_{m=0}^{M-1}\left[\phi_{x'_{\mathrm{ri}}}(n,m)+\phi_{v_{\mathrm{rn}}}(n,m)\right]}, \tag{26}$$

$$\mathrm{oSNR}[\mathbf{h}(n)]\triangleq\frac{\sum_{m=0}^{M-1}\phi_{x_{\mathrm{fd}}}(n,m)}{\sum_{m=0}^{M-1}\left[\phi_{x'_{\mathrm{ri}}}(n,m)+\phi_{v_{\mathrm{rn}}}(n,m)\right]}, \tag{27}$$

$$\begin{aligned}\mathrm{oSNR}[\mathbf{h}(n,m)]&\triangleq\frac{\phi_{x_{\mathrm{fd}}}(n,m)}{\phi_{x'_{\mathrm{ri}}}(n,m)+\phi_{v_{\mathrm{rn}}}(n,m)},\\&m=0,1,\ldots,M-1.\end{aligned} \tag{28}$$

In the particular case where $H'(n,m)=0$ (which is true for the classical linear estimation), we have

$$\mathrm{iSNR}(n,m)=\mathrm{oSNR}[\mathbf{h}(n,m)],\quad m=0,1,\ldots,M-1, \tag{29}$$

so, the subband SNR cannot be improved.

Another important measure in noise reduction is the noise-reduction factor (Benesty et al., 2005, 2009; Chen et al., 2006), which quantifies the amount of noise being attenuated by the complex filter. With the STFT formulation, the fullband noise-reduction factor is defined as

$$\xi_{\mathrm{nr}}(\mathbf{h})\triangleq\frac{\sum_n\sum_{m=0}^{M-1}\phi_v(n,m)}{\sum_n\sum_{m=0}^{M-1}\left[\phi_{x'_{\mathrm{ri}}}(n,m)+\phi_{v_{\mathrm{rn}}}(n,m)\right]}. \tag{30}$$

Similarly, we can define the segmental and subband noise-reduction factors $\xi_{\mathrm{nr}}[\mathbf{h}(n)]$ and $\xi_{\mathrm{nr}}[\mathbf{h}(n,m)]$. With optimal gains, these noise-reduction factors are expected to be greater than 1.

The complex gains add distortion to the desired signal. In order to evaluate this distortion, we define the subband speech-distortion index (Benesty et al., 2009) as

[1] In these definitions, the interference is considered as part of the noise at the output of the filter. The same applies to the noise-reduction factors.

$$\upsilon_{\mathrm{sd}}[\mathbf{h}(n,m)] \triangleq \frac{E\left[|X_{\mathrm{fd}}(n,m)-X(n,m)|^2\right]}{\phi_x(n,m)}, \quad m=0,1,\ldots,M-1. \tag{31}$$

This index is lower bounded by 0 and is expected to be smaller than 1 for optimal filters. The higher the value of $\upsilon_{\mathrm{sd}}[\mathbf{h}(n,m)]$, the more the speech is distorted at time-frame n and frequency-bin m. In a similar way, we can derive the segmental speech-distortion index $\upsilon_{\mathrm{sd}}[\mathbf{h}(n)]$ and fullband speech-distortion index $\upsilon_{\mathrm{sd}}(\mathbf{h})$ (Benesty et al., 2009).

Another interesting measure, which is somewhat similar to the noise-reduction factor, is the speech-reduction factor (Benesty et al., 2009). It measures the amount of the desired signal reduced by the complex gains. The subband speech-reduction factor is

$$\xi_{\mathrm{sr}}[\mathbf{h}(n,m)] \triangleq \frac{\phi_x(n,m)}{\phi_{x_{\mathrm{fd}}}(n,m)}, \quad m=0,1,\ldots,M-1. \tag{32}$$

Similarly, one can define the segmental and fullband speech-reduction factors $\xi_{\mathrm{sr}}[\mathbf{h}(n)]$ and $\xi_{\mathrm{sr}}(\mathbf{h})$. The larger the value of the speech-reduction factor, the more the desired signal is reduced (or distorted). This factor should be lower bounded by 1.

It is easy to verify that

$$\frac{\mathrm{oSNR}[\mathbf{h}(n,m)]}{\mathrm{iSNR}(n,m)} = \frac{\xi_{\mathrm{nr}}[\mathbf{h}(n,m)]}{\xi_{\mathrm{sr}}[\mathbf{h}(n,m)]}, \tag{33}$$

$$\frac{\mathrm{oSNR}[\mathbf{h}(n)]}{\mathrm{iSNR}(n)} = \frac{\xi_{\mathrm{nr}}[\mathbf{h}(n)]}{\xi_{\mathrm{sr}}[\mathbf{h}(n)]}, \tag{34}$$

$$\frac{\mathrm{oSNR}(\mathbf{h})}{\mathrm{iSNR}} = \frac{\xi_{\mathrm{nr}}(\mathbf{h})}{\xi_{\mathrm{sr}}(\mathbf{h})}. \tag{35}$$

Hence, a subband, segmental, or fullband increase in the SNR, i.e., $\mathrm{oSNR}[\mathbf{h}(n,m)] > \mathrm{iSNR}(n,m)$, $\mathrm{oSNR}[\mathbf{h}(n)] > \mathrm{iSNR}(n)$, or $\mathrm{oSNR}(\mathbf{h}) > \mathrm{iSNR}$, can be obtained only when the subband, segmental, or fullband noise-reduction factor is larger than the corresponding speech-reduction factor, i.e., $\xi_{\mathrm{nr}}[\mathbf{h}(n,m)] > \xi_{\mathrm{sr}}[\mathbf{h}(n,m)]$, $\xi_{\mathrm{nr}}[\mathbf{h}(n)] > \xi_{\mathrm{sr}}[\mathbf{h}(n)]$, or $\xi_{\mathrm{nr}}(\mathbf{h}) > \xi_{\mathrm{sr}}(\mathbf{h})$.

4. Widely linear mean-squared error

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

$$\begin{aligned}\mathscr{E}(n,m) &\triangleq Z(n,m)-X(n,m)\\ &= \mathbf{h}^H(n,m)\mathbf{y}(n,m)-X(n,m),\end{aligned} \tag{36}$$

which can be written as the sum of two error signals:

$$\mathscr{E}(n,m) = \mathscr{E}_{\mathrm{d}}(n,m)+\mathscr{E}_{\mathrm{r}}(n,m), \tag{37}$$

where

$$\begin{aligned}\mathscr{E}_{\mathrm{d}}(n,m) &\triangleq X_{\mathrm{fd}}(n,m)-X(n,m)\\ &= \left[\mathbf{h}^H(n,m)\boldsymbol{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1-1\right]X(n,m)\end{aligned} \tag{38}$$

is the signal distortion due to the complex filter and

$$\begin{aligned}\mathscr{E}_{\mathrm{r}}(n,m) &\triangleq X'_{\mathrm{ri}}(n,m)+V_{\mathrm{rn}}(n,m)\\ &= \mathbf{h}^H(n,m)\mathbf{i}_2X'(n,m)+\mathbf{h}^H(n,m)\mathbf{v}(n,m)\end{aligned} \tag{39}$$

represents the interference and noise residuals.

The subband mean-squared error (MSE) is then

$$J[\mathbf{h}(n,m)] \triangleq E\left[|\mathscr{E}(n,m)|^2\right] = J_{\mathrm{d}}[\mathbf{h}(n,m)]+J_{\mathrm{r}}[\mathbf{h}(n,m)], \tag{40}$$

where

$$\begin{aligned}J_{\mathrm{d}}[\mathbf{h}(n,m)] &\triangleq E\left[|\mathscr{E}_{\mathrm{d}}(n,m)|^2\right]\\ &= E\left[|X_{\mathrm{fd}}(n,m)-X(n,m)|^2\right]\\ &= \phi_x(n,m)\left|\mathbf{h}^H(n,m)\boldsymbol{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1-1\right|^2\end{aligned} \tag{41}$$

and

$$\begin{aligned}J_{\mathrm{r}}[\mathbf{h}(n,m)] &\triangleq E\left[|\mathscr{E}_{\mathrm{r}}(n,m)|^2\right]\\ &= E\left[|X'_{\mathrm{ri}}(n,m)|^2\right]+E\left[|V_{\mathrm{rn}}(n,m)|^2\right]\\ &= \phi_{x'_{\mathrm{ri}}}(n,m)+\phi_{v_{\mathrm{rn}}}(n,m).\end{aligned} \tag{42}$$

Let us take the particular filter $\mathbf{h}(n,m)=\mathbf{i}_1,\ \forall n,m$. In this case, the subband MSE is

$$J(\mathbf{i}_1) = \phi_v(n,m), \tag{43}$$

so there will be neither noise reduction nor speech distortion. We now define the subband normalized MSE (NMSE) as

$$\begin{aligned}\widetilde{J}[\mathbf{h}(n,m)] &\triangleq \frac{J[\mathbf{h}(n,m)]}{J(\mathbf{i}_1)} = \mathrm{iSNR}(n,m)\cdot\upsilon_{\mathrm{sd}}[\mathbf{h}(n,m)]\\ &\quad+\frac{1}{\xi_{\mathrm{nr}}[\mathbf{h}(n,m)]},\end{aligned} \tag{44}$$

where

$$\upsilon_{\mathrm{sd}}[\mathbf{h}(n,m)] \triangleq \frac{J_{\mathrm{d}}[\mathbf{h}(n,m)]}{\phi_x(n,m)}, \tag{45}$$

$$\xi_{\mathrm{nr}}[\mathbf{h}(n,m)] \triangleq \frac{\phi_v(n,m)}{J_{\mathrm{r}}[\mathbf{h}(n,m)]}. \tag{46}$$

This shows clearly how the two WL subband MSEs [Eqs. (41) and (42)] are related to some of the performance measures. Similar relations also hold for the fullband and segmental measures.

5. Widely linear Wiener filter

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

$$\begin{aligned}\mathbf{h}_{\mathrm{WLW}}(n,m) &= \boldsymbol{\Phi}_{\mathbf{y}}^{-1}(n,m)\boldsymbol{\Phi}_{\mathbf{x}}(n,m)\mathbf{i}_1 \\ &= \frac{\phi_x(n,m)}{\phi_y(n,m)}\cdot\boldsymbol{\Gamma}_{\mathbf{y}}^{-1}(n,m)\boldsymbol{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1 \\ &= \left[\mathbf{I}-\frac{\phi_v(n,m)}{\phi_y(n,m)}\cdot\boldsymbol{\Gamma}_{\mathbf{y}}^{-1}(n,m)\boldsymbol{\Gamma}_{\mathbf{v}}(n,m)\right]\mathbf{i}_1. \end{aligned} \tag{47}$$

It follows immediately that

$$H_{\mathrm{WLW}}(n,m) = \frac{1-\gamma_x(n,m)\gamma_y^*(n,m)}{1-\left|\gamma_y(n,m)\right|^2}\cdot\frac{\phi_x(n,m)}{\phi_y(n,m)}, \tag{48a}$$

$$H'_{\mathrm{WLW}}(n,m) = \frac{\gamma_x(n,m)-\gamma_y(n,m)}{1-\left|\gamma_y(n,m)\right|^2}\cdot\frac{\phi_x(n,m)}{\phi_y(n,m)}. \tag{48b}$$

Using (4) and (11), one can deduce the following relation:

$$\begin{aligned}\gamma_y(n,m)\phi_y(n,m) &= \gamma_x(n,m)\phi_x(n,m) \\ &\quad+\gamma_v(n,m)\phi_v(n,m).\end{aligned} \tag{49}$$

By using (49), the WL Wiener complex gains in (48), can be rearranged as

$$H_{\mathrm{WLW}}(n,m) = 1-\frac{1-\gamma_v(n,m)\gamma_y^*(n,m)}{1-\left|\gamma_y(n,m)\right|^2}\cdot\frac{\phi_v(n,m)}{\phi_y(n,m)}, \tag{50a}$$

$$H'_{\mathrm{WLW}}(n,m) = \frac{\gamma_y(n,m)-\gamma_v(n,m)}{1-\left|\gamma_y(n,m)\right|^2}\cdot\frac{\phi_v(n,m)}{\phi_y(n,m)}. \tag{50b}$$

We recall that the classical Wiener gain (Benesty et al., 2009) is

$$H_{\mathrm{W}}(n,m) = \frac{\phi_x(n,m)}{\phi_y(n,m)} = 1-\frac{\phi_v(n,m)}{\phi_y(n,m)}. \tag{51}$$

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.

In practical situations, $\gamma_x(n,m)$ is in general not zero because speech is nonstationary. But noise is relatively stationary as compared to the speech signal and $\gamma_v(n,m)$ may be close to 0. Now let us assume that $\gamma_v(n,m)=0$. With this assumption and using the subband SNR definition given in (25), we can write the WL Wiener complex gains in (48) [or (50)] as

$$H_{\mathrm{WLW}}(n,m) = \eta_{\mathrm{WLW}}(n,m)H_{\mathrm{W}}(n,m), \tag{52a}$$

$$H'_{\mathrm{WLW}}(n,m) = \eta'_{\mathrm{WLW}}(n,m)H_{\mathrm{W}}(n,m), \tag{52b}$$

where

$$\eta_{\mathrm{WLW}}(n,m) = \frac{1-\frac{\mathrm{iSNR}(n,m)}{1+\mathrm{iSNR}(n,m)}\cdot\left|\gamma_x(n,m)\right|^2}{1-\left[\frac{\mathrm{iSNR}(n,m)}{1+\mathrm{iSNR}(n,m)}\right]^2\cdot\left|\gamma_x(n,m)\right|^2}, \tag{53a}$$

$$\eta'_{\mathrm{WLW}}(n,m) = \frac{\frac{1}{1+\mathrm{iSNR}(n,m)}\cdot\gamma_x(n,m)}{1-\left[\frac{\mathrm{iSNR}(n,m)}{1+\mathrm{iSNR}(n,m)}\right]^2\cdot\left|\gamma_x(n,m)\right|^2}. \tag{53b}$$

It is easy to check that $\eta_{\mathrm{WLW}}(n,m)$ is always real and it satisfies $0\leqslant\eta_{\mathrm{WLW}}(n,m)\leqslant 1$. But $\eta'_{\mathrm{WLW}}(n,m)$ is in general complex, with its magnitude being in the range between 0 and 1, i.e., $0\leqslant|\eta_{\mathrm{WLW}}(n,m)|\leqslant 1$. The difference between the WL Wiener and classical Wiener filters depends on the two weighting functions $\eta_{\mathrm{WLW}}(n,m)$ and $\eta'_{\mathrm{WLW}}(n,m)$. Fig. 2 plots both $\eta_{\mathrm{WLW}}(n,m)$ and $|\eta'_{\mathrm{WLW}}(n,m)|$ as a function of the subband input SNR, $\mathrm{iSNR}(n,m)$, and the speech noncircularity $\gamma_x(n,m)$. There are three circumstances:

1. Large values of $\mathrm{iSNR}(n,m)$ (e.g., $\geqslant 15$ dB). In this case, $\eta_{\mathrm{WLW}}(n,m)$ approaches 1 while $\eta'_{\mathrm{WLW}}(n,m)$ is close to 0. This indicates that the WL Wiener filter converges to the classical Wiener filter in high SNR conditions regardless of the degree of speech noncircularity.
2. Small values of $\mathrm{iSNR}(n,m)$ (e.g., $\leqslant -15$ dB). In this situation, $\eta_{\mathrm{WLW}}(n,m)$ also approaches 1 regardless of the value of $\gamma_x(n,m)$. This shows that $H_{\mathrm{WLW}}(n,m)$ converges to the classical Wiener gain. But the value of $\gamma_x(n,m)$ plays an important role in $\eta_{\mathrm{WLW}}(n,m)$. So, in this circumstance, the WL Wiener is different from the classical Wiener filter, but the difference mainly comes from filtering the signal conjugate.
3. Moderate values of $\mathrm{iSNR}(n,m)$ (e.g., in the range between -15 dB and 15 dB). Both the subband input SNR and noncircularity play an important role in the WL Wiener filter.

From a practical viewpoint, the third case is the most interesting one since if the subband input SNR is very large, the WL Wiener filter will be similar to the classical

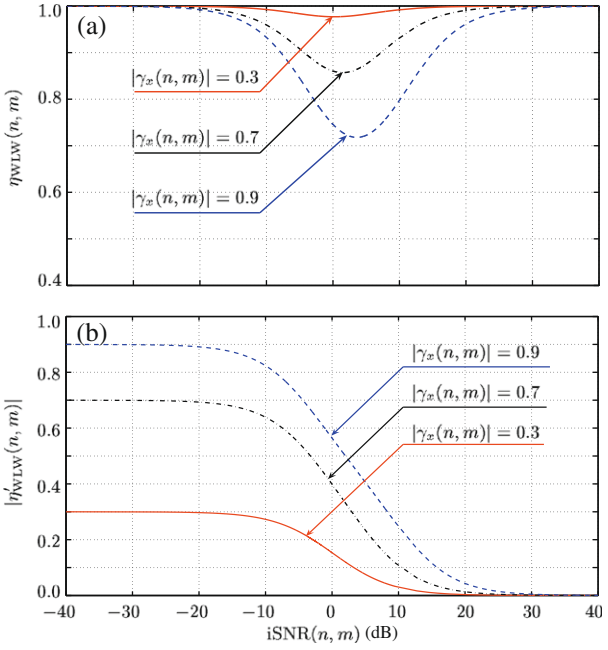

Fig. 2. (a) Relationship between $\eta_{\mathrm{WLW}}(n,m)$, $\mathrm{iSNR}(n,m)$, and $\gamma_x(n,m)$. (b) Relationship between $|\eta'_{\mathrm{WLW}}(n,m)|$, $\mathrm{iSNR}(n,m)$, and $\gamma_x(n,m)$.

Wiener filter; but if the input SNR is very small, then it would be very difficult to estimate the speech noncircularity even though the WL Wiener filter is superior to the classical Wiener filter.

Now let us take a slightly different angle by looking at the filtered desired signal and residual noise and interference. Substituting (52) into (19a), we can deduce the filtered desired signal due to the WL Wiener filter [assuming that $\gamma_v(n,m)=0$] as

$$X_{\text{fd, WLW}}(n,m)=G_{\text{fd, WLW}}(n,m)X(n,m), \tag{54}$$

where

$$\begin{aligned}G_{\text{fd, WLW}}(n,m)&=H_{\text{WLW}}(n,m)+\gamma_x^*(n,m)H'_{\text{WLW}}(n,m)\\&=\frac{1+\frac{1-\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)}\cdot|\gamma_x(n,m)|^2}{1-\left[\frac{\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)}\right]^2\cdot|\gamma_x(n,m)|^2}\\&\quad\cdot\frac{\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)}.\end{aligned} \tag{55}$$

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

$$G_{\text{fd, W}}(n,m)=H_{\text{W}}(n,m)=\frac{\text{iSNR}(n,m)}{1+\text{iSNR}(n,m)}. \tag{56}$$

It can be checked that $G_{\text{fd, WLW}}(n,m)\geqslant G_{\text{fd, W}}(n,m)$, meaning that the WL Wiener filter will introduce less speech distortion. Fig. 3 plots both $G_{\text{fd, WLW}}(n,m)$ and $G_{\text{fd, W}}(n,m)$ as a function of $\text{iSNR}(n,m)$. It is seen that both $G_{\text{fd, WLW}}(n,m)$ and $G_{\text{fd, W}}(n,m)$ increase with the subband input SNR. So, less speech distortion is added to the enhanced signal by 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 (55). 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. 3 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 $\text{iSNR}(n,m)=5$ dB, if we want the WL Wiener filter to have 5% less speech distortion than the classical Wiener filter, we would need $|\gamma_x(n,m)|\geqslant 0.76$.

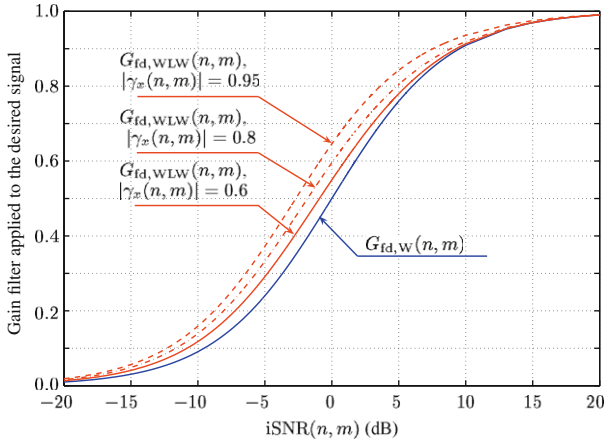

Fig. 3. Comparison between the WL and classical Wiener filters for their gain applied to filter the desired speech signal.

Now let us examine the MSE due to the WL Wiener filter. The subband minimum NMSE is found by replacing the WL Wiener filter in (44):

$$\begin{aligned}\widetilde{J}[\mathbf{h}_{\text{WLW}}(n,m)]&=\text{iSNR}(n,m)\\&\quad\times\left[1-\frac{\phi_x(n,m)}{\phi_y(n,m)}\cdot\mathbf{i}_1^H\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{\Gamma}_{\mathbf{y}}^{-1}(n,m)\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1\right].\end{aligned} \tag{57}$$

The subband NMSE for the classical Wiener filter is

$$\widetilde{J}[\mathbf{h}_{\text{W}}(n,m)]=\text{iSNR}(n,m)\left[1-\frac{\phi_x(n,m)}{\phi_y(n,m)}\right]. \tag{58}$$

Obviously,

$$\widetilde{J}[\mathbf{h}_{\text{WLW}}(n,m)]\leqslant\widetilde{J}[\mathbf{h}_{\text{W}}(n,m)]\leqslant 1,\ \forall n,m. \tag{59}$$

We have shown previously that the WL Wiener filter adds less distortion to the desired speech signal. The fact that the subband NMSE of the WL Wiener filter is smaller than that of the Wiener filter shows the advantage of the WL Wiener filter from another viewpoint. From (57) and (58), we also deduce that

$$1\leqslant\mathbf{i}_1^H\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{\Gamma}_{\mathbf{y}}^{-1}(n,m)\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1\leqslant 1+\frac{1}{\text{iSNR}(n,m)},\ \forall n,m. \tag{60}$$

The quantity $\mathbf{i}_1^H\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{\Gamma}_{\mathbf{y}}^{-1}(n,m)\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1$ determines the amount of noncircularity in the signals. In particular, for noncircular signals and when subband input SNRs are low, this quantity can be large and, as a result, the subband NMSE of the WL Wiener filter can be much smaller than the subband NMSE of the classical Wiener filter. For high subband input SNRs, we should not expect, obviously, much difference between the two filters.

Having shown that the WL Wiener filter introduces less speech distortion and has smaller NMSE 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* (47), *the subband output SNR is always greater than or equal to the subband input SNR, i.e.*, $\text{oSNR}[\mathbf{h}_{\text{WLW}}(n,m)]\geqslant\text{iSNR}(n,m)$, $\forall n,m$.

Proof. From the definition of the subband output SNR given in (28) and according to (9), (21), (22), we can readily obtain:

$$\mathrm{oSNR}[\mathbf{h}_{\mathrm{WLW}}(n,m)]=\frac{\phi_x\left(|H_{\mathrm{WLW}}|^2+\gamma_x H_{\mathrm{WLW}}H'^{*}_{\mathrm{WLW}}+\gamma_x^{*}H^{*}_{\mathrm{WLW}}H'_{\mathrm{WLW}}+|\gamma_x|^2|H'_{\mathrm{WLW}}|^2\right)}{\phi_v\left(|H_{\mathrm{WLW}}|^2+\gamma_v H_{\mathrm{WLW}}H'^{*}_{\mathrm{WLW}}+\gamma_v^{*}H^{*}_{\mathrm{WLW}}H'_{\mathrm{WLW}}+|H'_{\mathrm{WLW}}|^2\right)+\phi_x|H'_{\mathrm{WLW}}|^2\left(1-|\gamma_x|^2\right)}. \tag{61}$$

Note that in (61) we have dropped the (n,m) from $\phi_x(n,m)$, $\phi_v(n,m)$, $\gamma_x(n,m)$, $\gamma_v(n,m)$, $H_{\mathrm{WLW}}(n,m)$, and $H'_{\mathrm{WLW}}(n,m)$ to make the equation more compact. Substituting (48) into (61) and with some mathematical manipulation (see Appendix), we can show that

$$\mathrm{oSNR}[\mathbf{h}_{\mathrm{WLW}}(n,m)]\geqslant\frac{\phi_x(n,m)}{\phi_v(n,m)}=\mathrm{iSNR}(n,m), \tag{62}$$

where the equality holds when $\gamma_v(n,m)=\gamma_x(n,m)=0$. This completes the proof. □

Recall that for the classical Wiener filter, the input and output subband SNRs are equal, i.e.,

$$\mathrm{oSNR}[\mathbf{h}_{\mathrm{W}}(n,m)]=\mathrm{oSNR}[\mathbf{h}_{\mathrm{W}}(n,m)]=\frac{\phi_x(n,m)}{\phi_v(n,m)}. \tag{63}$$

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

6. Widely linear tradeoff filter

As can be seen from (47), not much flexibility is associated with the WL Wiener filter in the sense that we do not know in advance neither by how much the output SNR will be improved nor by how much the desired signal will be affected. This optimal filter tries to find, on its own, a compromise between noise reduction and speech distortion. However, in many practical situations, we wish to have some flexibility to control the compromise between noise reduction and speech distortion and the best way to do this is via the so-called tradeoff filter.

The subband NMSE as shown is (44) is the sum of two terms. One depends on the speech distortion while the other depends on the noise reduction. Instead of minimizing the subband NMSE as we already did in finding the WL Wiener filter, we can minimize the speech-distortion index with the constraint that the noise-reduction factor is equal to a positive value that is greater than one. Mathematically, this is equivalent to

$$\min_{\mathbf{h}(n,m)} J_{\mathrm{d}}[\mathbf{h}(n,m)] \quad \text{subject to} \quad J_{\mathrm{r}}[\mathbf{h}(n,m)]=\beta\phi_v(n,m), \tag{64}$$

where $0<\beta<1$ in order to have some noise reduction. If we use a Lagrange multiplier, $\mu\geqslant 0$, to adjoin the constraint to the cost function, we easily derive the WL tradeoff filter:

$$\mathbf{h}_{\mathrm{WLT}}(n,m)=\left[\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1\mathbf{i}_1^H\mathbf{\Gamma}_{\mathbf{x}}(n,m)+\mu\frac{\mathbf{\Phi}_{\mathrm{in}}(n,m)}{\phi_x(n,m)}\right]^{-1}\times\mathbf{\Gamma}_{\mathbf{x}}(n,m)\mathbf{i}_1, \tag{65}$$

where

$$\mathbf{\Phi}_{\mathrm{in}}(n,m)=\phi_x(n,m)\left[1-|\gamma_x(n,m)|^2\right]\mathbf{i}_2\mathbf{i}_2^H+\mathbf{\Phi}_{\mathbf{v}}(n,m) \tag{66}$$

is the covariance matrix of the interference plus noise, and the Lagrange multiplier, μ, satisfies $J_{\mathrm{r}}[\mathbf{h}_{\mathrm{WLT}}(n,m)]=\beta\phi_v(n,m)$. However, in practice it is not easy to determine the optimal μ. Therefore, when this parameter is chosen in an ad-hoc way, we can see that for

- $\mu=1$, $\mathbf{h}_{\mathrm{WLT}}(n,m)=\mathbf{h}_{\mathrm{WLW}}(n,m)$, which is the WL Wiener filter;
- $\mu=0$, $\mathbf{h}_{\mathrm{WLT}}(n,m)=\mathbf{i}_1$. This filter does not add any distortion to the desired speech signal, but does not reduce any noise either;
- $\mu>1$, results in an aggressive filter (compared to the WL Wiener case), which leads to a low level of residual noise, but at the expense of a high level of speech distortion;
- $\mu<1$, results in a conservative filter (compared to the WL Wiener case) with low speech distortion but high residual noise.

We recall that the classical tradeoff filter (Benesty et al., 2009) is

$$\mathbf{h}_{\mathrm{T}}(n,m)=\frac{\phi_x(n,m)}{\phi_x(n,m)+\mu\phi_v(n,m)}\cdot\mathbf{i}_1=\left[\frac{\phi_y(n,m)-\phi_v(n,m)}{\phi_y(n,m)-(\mu-1)\phi_v(n,m)}\right]\mathbf{i}_1. \tag{67}$$

If $\gamma_x(n,m)=\gamma_v(n,m)=0$, the WL tradeoff filter will degenerate to the classical tradeoff filter. While the tradeoff filter is always real, the WL tradeoff filter is, in general, complex.

Finding the two filters $\mathbf{h}_{\mathrm{WLT}}(n,m)$ and $\mathbf{h}_{\mathrm{T}}(n,m)$ in such a way that

$$J_{\mathrm{r}}[\mathbf{h}_{\mathrm{WLT}}(n,m)]=J_{\mathrm{r}}[\mathbf{h}_{\mathrm{T}}(n,m)]=\beta\phi_v(n,m), \tag{68}$$

implies that

$$J_{\mathrm{d}}[\mathbf{h}_{\mathrm{WLT}}(n,m)]\leqslant J_{\mathrm{d}}[\mathbf{h}_{\mathrm{T}}(n,m)]. \tag{69}$$

Therefore, with the same level of noise reduction, the WL tradeoff filter will cause less distortion to the speech signal than the classical tradeoff filter. Moreover, following the proof given in the Appendix, we can prove that with the WL tradeoff filter given in (65), the subband output SNR is always greater than or equal to the subband input SNR, i.e.,

$$\mathrm{oSNR}[\mathbf{h}_{\mathrm{WLT}}(n,m)]\geqslant\mathrm{iSNR}(n,m)=\mathrm{oSNR}[\mathbf{h}_{\mathrm{T}}(n,m)],\ \forall n,m. \tag{70}$$

This shows that the WL tradeoff filter may improve the subband SNR, while the classical tradeoff filter has no effect on the subband SNR for any given frame n and subband m.

7. Experimental results

We have developed a WL noise-reduction Wiener filter in Section 5 and a WL tradeoff filter in Section 6. Through the theoretical analysis, we have shown that the WL Wiener filter introduces less speech distortion and has a smaller minimum MSE than the classical Wiener filter. Furthermore, The WL Wiener filter can improve the subband SNR, while the classical Wiener filter has no effect on the subband SNR for any given frame and subband. Similarly, the WL tradeoff filter has many advantages over the classical tradeoff filter. We have carried out a number of experiments to study the performance of the developed WL noise-reduction filters. In this section, we present some results, which justify what we learned through the theoretical analysis in the previous sections and highlight the merits and limitations of the WL Wiener filter.

The critical issue in implementing the WL Wiener filter given in (47) or (48) lies in the estimation of the variance parameters (or power spectra) $\phi_y(n,m), \phi_x(n,m)$, and $\phi_v(n,m)$ and the noncircularity parameters $\gamma_y(n,m)$, $\gamma_x(n,m)$, and $\gamma_v(n,m)$. The estimation of the variance parameters has been well addressed in the literature (Benesty et al., 2009, Chen et al., 2006, Martin, 2001; Hirsch and Ehrlicher, 1995; Stahl et al., 2000; Diethorn, 2004). In this section, we will put our emphasis on the estimation of the noncircularity parameters. We consider two approaches: short-time average and recursive estimation. The former basically approximates the mathematical expectation in (11) with a short-time sample average, while the latter uses a recursive method.

7.1. Estimation of the noncircularity quotient using a short-time average method

The clean speech used in this experiment is from the TIMIT database (DARPA TIMIT, 1990; Lee and Hon, 1989), which was designed to provide speech data for acoustic-phonetic studies and for the development and evaluation of automatic speech recognition (ASR) systems. This database consists of a total of 6300 sentences spoken by 630 speakers (10 sentences by each). Each speech signal in the database is recorded using a 16-kHz sampling rate (with a 16-bit quantization) and is accompanied by manually segmented phonetic (based on 61 phonemes) transcripts. Fig. 4 (the upper trace) plots one speech signal from the speaker FAKS0 and both the phonetic transcription and phoneme boundaries are also visualized.

In this experiment, we take the ten sentences from the speaker FAKS0 and use them as the clean speech signals, and the corresponding noisy signals are generated by adding a white Gaussian noise into the clean speech with different

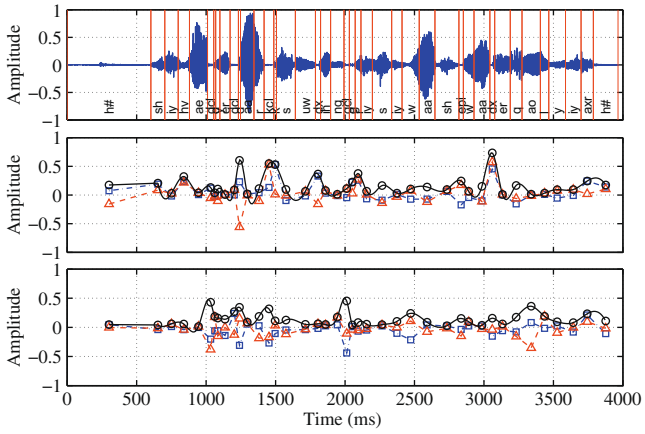

Fig. 4. A speech signal selected from the TIMIT database and the corresponding $\gamma_x(n,m)$ estimated with a short-time sample average. The upper trace: waveform with phoneme labeling and phoneme boundaries. The middle trace: the real (□), imaginary (△), and magnitude (○) parts of $\gamma_x(n,3)$ 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.

SNRs. To perform noise reduction in the frequency domain, each speech signal is partitioned into overlapping frames with a frame width of 4 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 average is used to replace the mathematical expectation in (4) and (11) to compute the variance parameters and circularity quotients. Note that the parameters $\phi_x(n,m)$ and $\gamma_x(n,m)$, and $\phi_y(n,m)$ and $\gamma_y(n,m)$ are directly computed from the clean and noisy signals respectively. Fig. 4 (the lower two traces) shows the estimated γ_x at the third and sixth subbands. Similar to the case in Fig. 1, it is clearly seen that $\gamma_x(n,m)$ is not equal to zero, which again shows that the complex STFT coefficients of speech are not circular variables.

With the computed variance and noncircularity parameters, we construct a WL Wiener filter for each phoneme at each subband. For the purpose of comparison, a classical Wiener gain is also constructed according to (51). After passing the noisy speech spectrum through the constructed Wiener filters, the inverse STFT (with overlap add) is used to obtain the time-domain speech estimate. Fig. 5 shows one clean speech signal, its noise corrupted counterpart (with iSNR = 10 dB), and the clean speech estimates using both the WL and classical Wiener filters. The spectrograms of the four signals in Fig. 5 are visualized in Fig. 6. It is clearly seen that a significant amount of noise reduction has been achieved with the use of the Wiener filters. Comparing Fig. 6c with Fig. 6d, one may notice that the WL Wiener filter achieves more noise reduction in some frequency bands.

To objectively assess the noise-reduction performance of the WL and Wiener filters, we evaluate three measures: the speech-distortion index, the noise-reduction factor, and the output SNR. These three performance measures can be

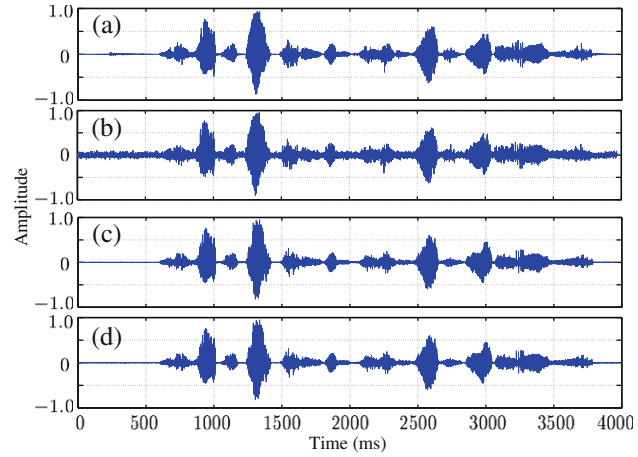

Fig. 5. Noise reduction using the WL and classical Wiener filters: (a) the clean speech, (b) the noisy speech with iSNR = 10 dB, (c) the output of the WL Wiener filter, and (d) the output of the classical Wiener filter.

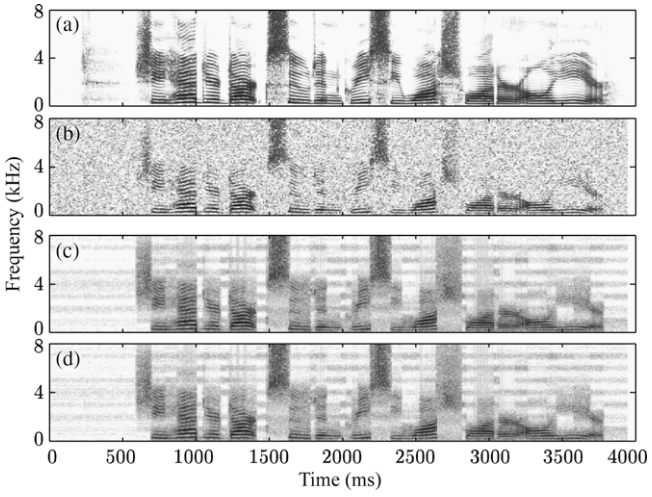

Fig. 6. Spectrograms of (a) the clean speech signal, (b) the noisy speech with iSNR = 10 dB, (c) the clean speech estimate obtained using the WL Wiener filter, and (d) the clean speech estimate obtained using the classical Wiener filter.

evaluated either in the fullband, or in the subband or segmental levels. But here we only compute the fullband measures.

Table 1 presents the average results computed over the ten speech signals. Note that the performance measures were computed for each individual signal first and the results were averaged out then. The results for the speech-distortion index and the noise-reduction factor were averaged in the linear scale while the output SNRs were averaged in the dB scale.

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. At the same time, the WL Wiener filter has achieved more noise reduction. However, as the input SNR increases, the WL Wiener filter tends to have a similar performance to the classical Wiener filter. This coincides with the theoretical analysis that the WL Wiener filter converges to the classical Wiener filter in high SNR conditions. We also see that, in general, the WL Wiener filter yields a higher output SNR as compared to the classical Wiener filter. However, the difference is somehow marginal. The underlying reasons are multiple. First, 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. Second, as we have shown previously that a significant amount of noncircularity (with $|\gamma_x(n,m)| \geqslant 0.7$) is needed in order for the WL Wiener filter to achieve noticeable better performance than the classical Wiener filter. Through experiments, we find that while the noncircularity for some subbands are strong, there are also many

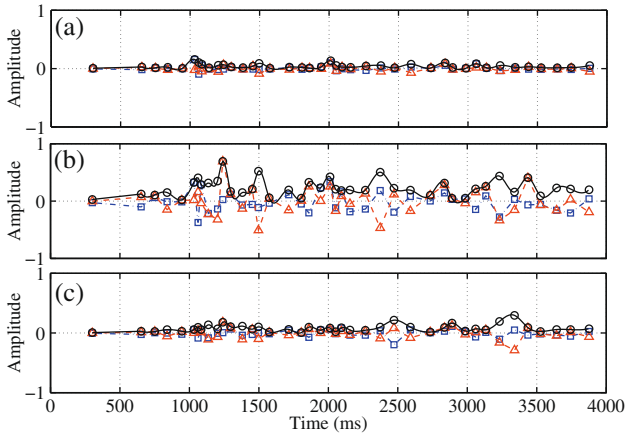

Fig. 7. The noncircularity quotient $\gamma_y(n,m)$ (□: real part, △: imaginary part, ○: magnitude) estimated with a short-time average method in white Gaussian noise with iSNR = 10 dB: (a) $m = 4$, (b) $m = 5$, and (c) $m = 6$.

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

Performance measure	Noise reduction filter	iSNR (Input SNR) (dB)				
		−10	−5	0	5	10
Speech distortion υ_{sd}	WF	0.213	0.113	0.051	0.020	0.007
	WL	0.211	0.111	0.050	0.020	0.007
Noise reduction ξ_{nr}	WF	89.65	39.29	19.05	9.94	6.25
	WL	90.38	39.71	19.11	10.01	6.29
Output SNR	WF	6.33 dB	9.18 dB	11.85 dB	14.50 dB	17.74 dB
	WL	6.43 dB	9.23 dB	11.88 dB	14.53 dB	17.77 dB

subbands where the noncircularity quotients can be small. Fig. 7 plots the $\gamma_y(n,m)$ at the 4th, 5th, and 6th subbands for the signal shown in Fig. 5. It is seen that the noncircularity quotient in the 5th subband is large, but the quotients for the 4th and 6th subbands are small. So, for the 4th and 6th subbands, the WL Wiener filter would have a similar performance to the classical Wiener filter. In practice, we should check the $\gamma_y(n,m)$ or $\gamma_x(n,m)$ estimates. If their magnitude is small, we can simply replace the WL Wiener filter with the classical Wiener filter at that subband without dramatically affecting the noise-reduction performance.

We also studied the WL and classical Wiener filters for their performance in different noise conditions. Table 2 shows the results in a car-noise condition where the noise was recorded in a car running at 50 miles/h on a high way. Similar to the previous experiment, we see that the WL Wiener filter consistently outperforms the classical Wiener filter.

7.2. Estimation of the noncircularity quotient using a recursive method

In the previous experiment, the variance and noncircularity parameters $\phi_x(n,m)$ and $\gamma_x(n,m)$ are directly computed from the clean speech signal. In practice, however, the clean speech signal is not accessible. As a result, $\phi_x(n,m)$ and $\gamma_x(n,m)$ have to be estimated from the noisy observation. A normal practice is to estimate the variance and noncircularity parameters of the noisy and noise signals first, and the parameters of the clean speech can then be derived using the relations given in (3) and (49). In this section, we consider a recursive algorithm (which was originally developed to estimate the spectra of the noisy and noise signals (Diethorn, 2004; Chen et al., 2003) to estimate both the variance and pseudo-variance parameters. Specifically, the variance and noncircularity quotient of the noise signal at the mth subband is estimated as

$$\hat{\phi}_v(n,m)=\begin{cases}\alpha_{v,\mathrm{a}}\hat{\phi}_v(n-1,m)+(1-\alpha_{v,\mathrm{a}})|Y(n,m)|^2,\\ \quad \text{if } |Y(n,m)|^2\geqslant\hat{\phi}_v(n-1,m)\\ \alpha_{v,\mathrm{d}}\hat{\phi}_v(n-1,m)+(1-\alpha_{v,\mathrm{d}})|Y(n,m)|^2,\\ \quad \text{if } |Y(n,m)|^2<\hat{\phi}_v(n-1,m)\end{cases}, \tag{71}$$

where $\alpha_{v,\mathrm{a}}$ and $\alpha_{v,\mathrm{d}}$ are, respectively, the "attack" and the "decay" coefficients, and

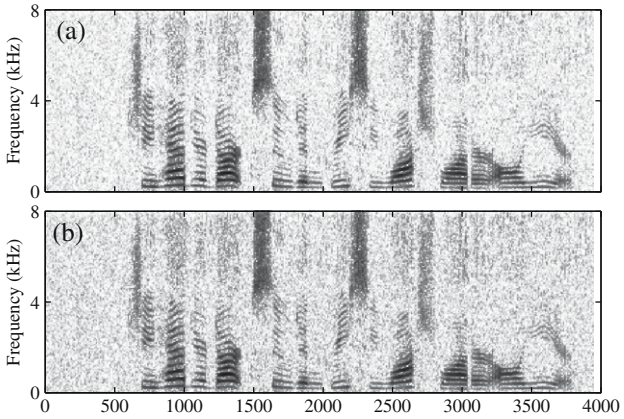

Fig. 8. Spectrograms of (a) the enhanced signal by the WL Wiener filter and (b) the enhanced signal by the classical Wiener filter. The clean and noisy speech signals are the same as shown in Fig. 6 with iSNR = 10 dB.

$$\hat{\gamma}_v(n,m)=\frac{\overline{V^2}(n,m)}{\hat{\phi}_v(n,m)}, \tag{72}$$

with

$$\overline{V^2}(n,m)=\begin{cases}\beta_{v,\mathrm{a}}\overline{V^2}(n-1,m)+(1-\beta_{v,\mathrm{a}})Y^2(n,m),\\ \quad \text{if } |Y(n,m)|^2\geqslant\hat{\phi}_v(n-1,m)\\ \beta_{v,\mathrm{d}}\overline{V^2}(n-1,m)+(1-\beta_{v,\mathrm{d}})Y^2(n,m),\\ \quad \text{if } |Y(n,m)|^2<\hat{\phi}_v(n-1,m)\end{cases}, \tag{73}$$

where again $\beta_{v,\mathrm{a}}$ is an "attack" coefficient and $\beta_{v,\mathrm{d}}$ a "decay" coefficient. Similarly, the noisy variance $\phi_y(n,m)$ is computed from the noisy spectrum $Y(n,m)$ using the following recursion:

$$\hat{\phi}_y(n,m)=\begin{cases}\alpha_{y,\mathrm{a}}\hat{\phi}_y(n-1,m)+(1-\alpha_{y,\mathrm{a}})|Y(n,m)|^2,\\ \quad \text{if } |Y(n,m)|^2\geqslant\hat{\phi}_y(n-1,m)\\ \alpha_{y,\mathrm{d}}\hat{\phi}_y(n-1,m)+(1-\alpha_{y,\mathrm{d}})|Y(n,m)|^2,\\ \quad \text{if } |Y(n,m)|^2<\hat{\phi}_y(n-1,m)\end{cases}. \tag{74}$$

Note that typically the values of $\alpha_{y,\mathrm{a}}$ and $\alpha_{y,\mathrm{d}}$ should be smaller than those of $\alpha_{v,\mathrm{a}}$ and $\alpha_{v,\mathrm{d}}$ because the noise is assumed to be more stationary than the desired speech signal. With the $\hat{\phi}_y(n,m)$ estimated using (74), we can then compute the noncircularity quotient $\hat{\gamma}_y(n,m)$ using a recursion

Table 2
Performance of the classical and WL Wiener filters in car noise.

Performance measure	Noise reduction filter	iSNR (Input SNR) (dB)				
		−10	−5	0	5	10
Speech distortion υ_{sd}	WF	0.089	0.052	0.029	0.014	0.006
	WL	0.086	0.049	0.027	0.013	0.005
Noise reduction ξ_{nr}	WF	214.66	111.77	54.41	23.94	10.60
	WL	225.80	118.63	57.66	24.83	10.80
Output SNR oSNR	WF	12.07 dB	14.73 dB	16.88 dB	18.49 dB	20.08 dB
	WL	12.37 dB	15.03 dB	17.17 dB	18.68 dB	20.18 dB

Table 3
Performance of the classical and WL Wiener filters in white Gaussian noise.

Performance measure	Noise reduction filter	iSNR (Input SNR) (dB)				
		−10	−5	0	5	10
Speech distortion υ_{sd}	WF	0.502	0.352	0.198	0.122	0.066
	WL	0.491	0.340	0.194	0.121	0.066
Output SNR oSNR	WF	−8.81 dB	−0.313 dB	8.06 dB	13.80 dB	18.41 dB
	WL	−8.63 dB	0.326 dB	8.74 dB	14.42 dB	19.00 dB

similar to (72), but with a different set of attack and decay coefficients, i.e., $\beta_{y,a}$, and $\beta_{y,d}$.

The proper values of all the attack and decay coefficients will no doubt play a critical role in affecting the noise-reduction performance. Unfortunately, it is very difficult to determine the optimal values of these coefficients using analytical methods. So, we carried out a series of noise-reduction experiments based on a 30-second long clean speech signal and several different types of noise. By varying each attack or decay coefficient from 0 and 1, good performance has been achieved with $\alpha_{v,a} = \beta_{v,a} = 0.999$, $\alpha_{v,d} = \beta_{v,d} = 0.98$, $\alpha_{y,a} = \beta_{y,a} = 0.67$, $\alpha_{y,d} = \beta_{y,d} = 0.92$ for both the WL and classical Wiener filters. These values are then used in this experiment for performance evaluation.

Fig. 8 presents an example of the outputs of the WL and classical Wiener filters where the original clean and noisy speech signals are the same as in Fig. 5. Comparing Fig. 8 with the noisy speech spectrogram shown in Fig. 6, one can see that a significant amount of noise has been reduced with the estimated classical and WL Wiener filters.

Same as in the previous experiment, we computed the average speech distortion index and output SNR over the 10 TIMIT speech signals. The results are shown in Table 3. More than 8-dB SNR improvement has been achieved with both the classical and WL Wiener filters. Comparatively, the WL Wiener filter has a smaller speech-distortion index and higher SNR improvement, which again coincides with what was observed from the theoretical analysis.

Notice that the difference in output SNR between the WL and classical Wiener filters is less than 1 dB. This indicates that, even though the noncircularity information can be used to improve noise reduction performance in terms of speech distortion and SNR improvement, the additional performance improvement due to the use of noncircularity may not be so significant. The major underlying reason, from our observation, is that the noncircularity quotients in many subbands are not very large to furnish dramatic performance improvement.

8. Conclusions

Noise reduction is often formulated as a filtering problem in the frequency domain. 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 will introduce less distortion to the desired speech signal and has a smaller NMSE as compared to the classical Wiener filter. Most importantly, we have proved that 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. We also compared the WL and classical Wiener filters using experiments and the results corroborate the theoretical analysis. We have also deduced a WL tradeoff filter, which can be used to adjust the compromise between the amount of noise reduction and the degree of speech distortion when it is needed. This new tradeoff filter has many advantages over the classical tradeoff filters in terms of speech distortion, noise reduction, and subband SNR improvement.

Appendix.

Lemma. *With the WL Wiener filter, we have the following inequality*:

$$\frac{\phi_x\left(|H_{\mathrm{WLW}}|^2+\gamma_x H_{\mathrm{WLW}}H'^{*}_{\mathrm{WLW}}+\gamma_x^{*}H^{*}_{\mathrm{WLW}}H'_{\mathrm{WLW}}+|\gamma_x|^2|H'_{\mathrm{WLW}}|^2\right)}{\phi_v\left(|H_{\mathrm{WLW}}|^2+\gamma_v H_{\mathrm{WLW}}H'^{*}_{\mathrm{WLW}}+\gamma_v^{*}H^{*}_{\mathrm{WLW}}H'_{\mathrm{WLW}}+|H'_{\mathrm{WLW}}|^2\right)+\phi_x|H'_{\mathrm{WLW}}|^2\left(1-|\gamma_x|^2\right)}\geqslant\frac{\phi_x}{\phi_v}. \tag{75}$$

Note that, again, for the purpose of compactness, we have dropped the (n,m) from $\phi_x(n,m), \phi_v(n,m), \gamma_x(n,m), \gamma_v(n,m), H_{\text{WLW}}(n,m)$, and $H'_{\text{WLW}}(n,m)$, which should not cause any confusion.

Proof. Using the fact that $|\gamma_y|^2 \leqslant 1$, we can easily check that

$$0 \leqslant |\gamma_y - \gamma_x|^2 \leqslant 1 - \gamma_x\gamma_y^* - \gamma_x^*\gamma_y + |\gamma_x|^2. \tag{76}$$

Slightly rearranging (76) gives

$$2 - \gamma_x\gamma_y^* - \gamma_x^*\gamma_y \geqslant 1 - |\gamma_x|^2. \tag{77}$$

It follows immediately that (assuming that $\phi_v \neq 0$)

$$\begin{aligned}&\frac{\phi_y}{\phi_v}|\gamma_x - \gamma_y|^2\left(1 - \gamma_x\gamma_y^*\right) + \frac{\phi_y}{\phi_v}|\gamma_x - \gamma_y|^2\left(1 - \gamma_x^*\gamma_y\right)\\ &\quad\geqslant \frac{\phi_y}{\phi_v}|\gamma_x - \gamma_y|^2\left(1 - |\gamma_x|^2\right).\end{aligned} \tag{78}$$

From (49), we easily obtain the following relation:

$$\gamma_x - \gamma_v = \frac{\phi_y}{\phi_v}(\gamma_x - \gamma_y). \tag{79}$$

Using (79) and the fact that $\phi_y = \phi_x + \phi_v$, we can rewrite (78) as

$$\begin{aligned}&(\gamma_x - \gamma_v)\left(\gamma_x^* - \gamma_y^*\right)\left(1 - \gamma_x\gamma_y^*\right) + \left(\gamma_x^* - \gamma_v^*\right)(\gamma_x\\ &\quad - \gamma_y)\left(1 - \gamma_x^*\gamma_y\right) + |\gamma_x - \gamma_y|^2\left(|\gamma_x|^2 - 1\right)\\ &\quad\geqslant \frac{\phi_x}{\phi_v}|\gamma_x - \gamma_y|^2\left(1 - |\gamma_x|^2\right).\end{aligned} \tag{80}$$

From the above inequality and using the WL Wiener filter given in (48b), we get

$$\begin{aligned}&(\gamma_x - \gamma_v)H_{\text{WLW}}H_{\text{WLW}}'^* + \left(\gamma_x^* - \gamma_v^*\right)H_{\text{WLW}}^*H'_{\text{WLW}}\\ &\qquad + |H'_{\text{WLW}}|^2\left(|\gamma_x|^2 - 1\right)\\ &\quad\geqslant \frac{\phi_x}{\phi_v}|H'_{\text{WLW}}|^2\left(1 - |\gamma_x|^2\right).\end{aligned} \tag{81}$$

With some simple mathematical manipulation, (81) becomes

$$\begin{aligned}&|H_{\text{WLW}}|^2 + \gamma_x H_{\text{WLW}}H_{\text{WLW}}'^* + \gamma_x^*H_{\text{WLW}}^*H'_{\text{WLW}}\\ &\quad + |H'_{\text{WLW}}|^2 - |H_{\text{WLW}}|^2 + \gamma_v H_{\text{WLW}}H_{\text{WLW}}'^*\\ &\quad + \gamma_v^*H_{\text{WLW}}^*H'_{\text{WLW}} + |H'_{\text{WLW}}|^2\\ &\quad\geqslant \frac{\phi_x}{\phi_v}|H'_{\text{WLW}}|^2\left(1 - |\gamma_x|^2\right).\end{aligned} \tag{82}$$

It is then straightforward to verify the inequality in (75). Therefore, the WL Wiener filter can improve the subband SNR. In comparison, the classical Wiener filter has no effect on the subband SNR for any given frame n and subband m. □

References

Amblard, P.O., Gaeta, M., Lacoume, J.L., 1996a. Statistics for complex variables and signals – Part I: Variables. Signal Process. 53, 1–13.

Amblard, P.O., Gaeta, M., Lacoume, J.L., 1996b. Statistics for complex variables and signals – Part II: Signals. Signal Process. 53, 15–25.

in Benesty, J., Makino, S., Chen, J. (Eds.), 2005. Speech Enhancement. Springer-Verlag, Berlin, Germany.

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

Chen, J., Huang, Y., Benesty, J., 2003. Filtering techniques for noise reduction and speech enhancement. In: Benesty, J., Huang, Y. (Eds.), Adaptive Signal Processing: Applications to Real-World Problems. Springer-Verlag, Berlin, Germany, pp. 129–154.

Chen, J., Benesty, J., Huang, Y., Doclo, S., 2006. New insights into the noise reduction Wiener filter. IEEE Trans. Audio, Speech, Lang. Process. 14, 1218–1234.

Chevalier, P., Delmas, J.-P., Oukaci, A., 2009. Optimal widely linear MVDR beamforming for noncircular signals. In: Proc. IEEE ICASSP, pp. 3573–3576.

DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT), from the NIST TIMIT Speech Disc, 1990.

Diethorn, E.J., 2004. Subband noise reduction methods for speech enhancement. In: Huang, Y., Benesty, J. (Eds.), Audio Signal Processing for Next-Generation Multimedia Communication Systems. Kluwer, Boston, MA, pp. 91–115.

Eriksson, J., Ollila, E., Koivunen, V., 2009. Statistics for complex random variables revisited. In: Proc. IEEE ICASSP, pp. 3565–3568.

Hirsch, H.G., Ehrlicher, C., 1995. Noise estimation techniques for robust speech recognition. In: Proc. IEEE ICASSP, vol. 1, pp. 153–156.

Huang, Y., Benesty, J., Chen, J., 2006. Acoustic MIMO Signal Processing. Springer-Verlag, Berlin, Germany.

Lee, K.-F., Hon, H.-W., 1989. Speaker-independent phone recognition using hidden Markov models. IEEE Trans. Acoust., Speech, Signal Process. 37, 1641–1648.

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

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

Martin, R., 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. IEEE Trans. Speech Audio Process. 9, 504–512.

Neeser, F.D., Massey, J.L., 1993. Proper complex random processes with applications to information theory. IEEE Trans. Inform. Theory 39, 1293–1302.

Ollila, E., 2008. On the circularity of a complex random variable. IEEE Signal Process. Lett. 15, 841–844.

Picinbono, B., 1994. On circularity. IEEE Trans. Signal Process. 42, 3473–3482.

Picinbono, B., Chevalier, P., 1995. Widely linear estimation with complex data. IEEE Trans. Signal Process. 43, 2030–2033.

Schreier, P.J., Scharf, L.L., 2003. Second-order analysis of improper complex random vectors and processes. IEEE Trans. Signal Process. 51, 714–725.

Stahl, V., Fischer, A., Bippus, R., 2000. Quantile based noise estimation for spectral subtraction and Wiener filtering. In: Proc. IEEE ICASSP, vol. 3, pp. 1875–1878.

Vary, P., Martin, R., 2006. Digital Speech Transmission: Enhancement, Coding and Error Concealment. John Wiley & Sons, Ltd., Chichester, England.