SINGLE-CHANNEL NOISE REDUCTION IN THE STFT DOMAIN FROM THE FULLBAND OUTPUT SNR PERSPECTIVE

Yingke Zhao[1]*, Jacob Benesty*[2]*, and Jingdong Chen*[1]

[1]CIAIC and School of Marine Science and Technology
Northwestern Polytechnical University
127 Youyi West Rd. Xi'an, Shaanxi 710072, China

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

ABSTRACT

This paper develops a single-channel noise reduction algorithm in the short-time Fourier transform (STFT) domain, which attempts to optimize the fullband output signal-to-noise ratio (SNR). We show that the conventional Wiener filter, the maximum SNR filter, and the ideal binary mask based method are particular cases of the developed algorithm. Simulations are presented to illustrate the properties of this algorithm.

***Index Terms*—**Noise reduction, speech enhancement, STFT domain, optimal gains, fullband output SNR, ideal binary mask.

1. INTRODUCTION

Although significant efforts have been devoted to it over the last four decades [1]–[13], noise reduction remains an open research problem. In this paper, we investigate this problem in the short-time Fourier transform (STFT) domain and develop an algorithm that can optimize the fullband output signal-to-noise ratio (SNR). We show that this algorithm is a generalized form of the widely used approaches in the STFT domain such as the Wiener and maximum SNR filters.

2. SIGNAL MODEL AND PROBLEM FORMULATION

The noise reduction or speech enhancement problem considered in this study is one of recovering the desired signal (or clean speech) $x(t)$, t being the time index, of zero mean from the noisy observation (microphone signal) [4],[5]:

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

where $v(t)$ is the unwanted additive noise, which is assumed to be a zero-mean random process white or colored but uncorrelated with $x(t)$. All signals are considered to be real and broadband.

Using the STFT, (1) can be rewritten in the time-frequency domain as [6]

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

where the zero-mean complex random variables $Y(k,n)$, $X(k,n)$, and $V(k,n)$ are the STFTs of $y(t)$, $x(t)$, and $v(t)$, respectively, at the frequency bin $k \in \{0,1,\ldots,K-1\}$ and the time frame n. Since $x(t)$ and $v(t)$ are uncorrelated by assumption, the variance of $Y(k,n)$ is

$$\begin{aligned}\phi_Y(k,n) &= E\left[|Y(k,n)|^2\right] \\ &= \phi_X(k,n) + \phi_V(k,n),\end{aligned} \tag{3}$$

where $E[\cdot]$ denotes mathematical expectation, and $\phi_X(k,n) = E\left[|X(k,n)|^2\right]$ and $\phi_V(k,n) = E\left[|V(k,n)|^2\right]$ are the variances of $X(k,n)$ and $V(k,n)$, respectively.

This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under grant No. 61425005.

Then, the objective of single-channel noise reduction in the STFT domain is to estimate the desired signal, $X(k,n)$, from the observation signal, $Y(k,n)$, in the best possible way.

Before leaving this section, let us define the subband and fullband input SNRs. The subband input SNR is defined as

$$\text{iSNR}(k,n) = \frac{\phi_X(k,n)}{\phi_V(k,n)} \tag{4}$$

and the fullband input SNR is

$$\text{iSNR}(n) = \frac{\sum_{k=0}^{K-1}\phi_X(k,n)}{\sum_{k=0}^{K-1}\phi_V(k,n)}. \tag{5}$$

It can be checked that

$$\min_k \text{iSNR}(k,n) \leq \text{iSNR}(n) \leq \max_k \text{iSNR}(k,n) \tag{6}$$

In other words, the fullband input SNR can never exceed the maximum of the subband input SNRs and can never be smaller than the minimum of the subband input SNRs.

3. NOISE REDUCTION WITH GAINS

In the most widely used approaches to noise reduction in the STFT domain, a complex gain, $H(k,n)$, is applied to the observation, $Y(k,n)$, i.e.,

$$\begin{aligned}Z(k,n) &= H(k,n)Y(k,n) \\ &= X_{\text{fd}}(k,n) + V_{\text{rn}}(k,n),\end{aligned} \tag{7}$$

where $Z(k,n)$ is the estimate of $X(k,n)$, $X_{\text{fd}}(k,n) = H(k,n)X(k,n)$ is the filtered desired signal, and $V_{\text{rn}}(k,n) = H(k,n)V(k,n)$ is the residual noise. The variance of $Z(k,n)$ is then

$$\begin{aligned}\phi_Z(k,n) &= |H(k,n)|^2\,\phi_Y(k,n) \\ &= \phi_{X_{\text{fd}}}(k,n) + \phi_{V_{\text{rn}}}(k,n),\end{aligned} \tag{8}$$

where $\phi_{X_{\text{fd}}}(k,n) = |H(k,n)|^2\,\phi_X(k,n)$ and $\phi_{V_{\text{rn}}}(k,n) = |H(k,n)|^2\,\phi_V(k,n)$ are the variances of $X_{\text{fd}}(k,n)$ and $V_{\text{rn}}(k,n)$, respectively.

It is clear that the subband input and output SNRs are equal. However, the fullband output SNR, which is given by

$$\text{oSNR}\left[H(:,n)\right] = \frac{\sum_{k=0}^{K-1}\phi_{X_{\text{fd}}}(k,n)}{\sum_{k=0}^{K-1}\phi_{V_{\text{rn}}}(k,n)}, \tag{9}$$

is generally different from the fullband input SNR.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

The objective of this work is to find the K subband gains, $H(k,n)$, $k=0,1,\ldots,K-1$, in such a way that the fullband output SNR is as large as possible or, at least, greater than the fullband input SNR, i.e., $\text{oSNR}\left[H(:,n)\right] > \text{iSNR}(n)$.

To simplify the notation, we denote $\text{iSNR}\left(k,n\right)$ by $\lambda\left(k,n\right)$ from now on. We also re-order the subband input SNRs in the following way:

$$\lambda\left(k_0,n\right) \geq \lambda\left(k_1,n\right) \geq \cdots \geq \lambda\left(k_{K-1},n\right), \tag{10}$$

where k_i, $i=0,1,\ldots,K-1$ and $k_i \in \{0,1,\ldots,K-1\}$.

Given (7) and (10), one can write the fullband output SNR as

$$\begin{aligned}\text{oSNR}\left[\mathbf{h}(n)\right] &= \frac{\mathbf{h}^H(n)\mathbf{D}_X(n)\mathbf{h}(n)}{\mathbf{h}^H(n)\mathbf{D}_V(n)\mathbf{h}(n)} \\ &= \frac{\sum_{i=0}^{K-1}\left|H\left(k_i,n\right)\right|^2\phi_X\left(k_i,n\right)}{\sum_{i=0}^{K-1}\left|H(k_i,n)\right|^2\phi_V\left(k_i,n\right)},\end{aligned} \tag{11}$$

where

$$\mathbf{h}(n) = \left[H\left(k_0,n\right) \;\; H\left(k_1,n\right) \;\; \ldots \;\; H\left(k_{K-1},n\right)\right]^T \tag{12}$$

is a filter of length K containing all the subband gains, the superscripts ^{T} and ^{H} are the transpose and conjugate-transpose operators, respectively, and

$$\mathbf{D}_X(n) = \text{diag}\left[\phi_X\left(k_0,n\right),\phi_X\left(k_1,n\right),\ldots,\phi_X\left(k_{K-1},n\right)\right], \tag{13}$$

$$\mathbf{D}_V(n) = \text{diag}\left[\phi_V\left(k_0,n\right),\phi_V(k_1,n),\ldots,\phi_V\left(k_{K-1},n\right)\right], \tag{14}$$

are two diagonal matrices. It can be checked that

$$\mathbf{D}_V^{-1}(n)\mathbf{D}_X(n) = \text{diag}\left[\lambda\left(k_0,n\right),\lambda\left(k_1,n\right),\ldots,\lambda\left(k_{K-1},n\right)\right]$$

is also a diagonal matrix containing all the K subband input SNRs ordered from the largest to the smallest.

Now, we give two interesting properties.

Property 3.1: With $\lambda\left(k_0,n\right) \geq \lambda\left(k_1,n\right) \geq \cdots \geq \lambda\left(k_{K-1},n\right) \geq 0$, we have

$$\begin{aligned}\frac{\sum_{i=0}^{K-1}\left|\alpha_i(n)\right|^2\lambda\left(k_i,n\right)}{\sum_{i=0}^{K-1}\left|\alpha_i(n)\right|^2} &\leq \frac{\sum_{i=0}^{K-2}\left|\alpha_i(n)\right|^2\lambda\left(k_i,n\right)}{\sum_{i=0}^{K-2}\left|\alpha_i(n)\right|^2} \leq \cdots \\ \cdots &\leq \frac{\sum_{i=0}^{1}\left|\alpha_i(n)\right|^2\lambda\left(k_i,n\right)}{\sum_{i=0}^{1}\left|\alpha_i(n)\right|^2} \leq \lambda\left(k_0,n\right)\end{aligned} \tag{15}$$

or, equivalently,

$$\begin{aligned}\frac{\sum_{i=0}^{K-1}\left|\alpha_i(n)\right|^2\phi_X\left(k_i,n\right)}{\sum_{i=0}^{K-1}\left|\alpha_i(n)\right|^2\phi_V\left(k_i,n\right)} &\leq \frac{\sum_{i=0}^{K-2}\left|\alpha_i(n)\right|^2\phi_X\left(k_i,n\right)}{\sum_{i=0}^{K-2}\left|\alpha_i(n)\right|^2\phi_V\left(k_i,n\right)} \leq \cdots \\ \cdots &\leq \frac{\sum_{i=0}^{1}\left|\alpha_i(n)\right|^2\phi_X\left(k_i,n\right)}{\sum_{i=0}^{1}\left|\alpha_i(n)\right|^2\phi_V\left(k_i,n\right)} \leq \frac{\phi_X\left(k_0,n\right)}{\phi_V\left(k_0,n\right)},\end{aligned} \tag{16}$$

where $\alpha_i(n)$, $i=0,1,\ldots,K-1$ are arbitrary complex numbers with at least one of them different from 0.

Property 3.2: With $\lambda(k_0,n) \geq \lambda(k_1,n) \geq \cdots \geq \lambda(k_{K-1},n) \geq 0$, we have

$$\begin{aligned}\lambda\left(k_{K-1},n\right) &\leq \frac{\sum_{i=0}^{1}\left|\beta_{K-1-i}(n)\right|^2\lambda\left(k_{K-1-i},n\right)}{\sum_{i=0}^{K-1}\left|\beta_{K-1-i}(n)\right|^2} \leq \cdots \\ \cdots &\leq \frac{\sum_{i=0}^{K-2}\left|\beta_{K-1-i}(n)\right|^2\lambda\left(k_{K-1-i},n\right)}{\sum_{i=0}^{K-2}\left|\beta_{K-1-i}(n)\right|^2} \leq \\ &\frac{\sum_{i=0}^{K-1}\left|\beta_{K-1-i}(n)\right|^2\lambda\left(k_{K-1-i},n\right)}{\sum_{i=0}^{K-2}\left|\beta_{K-1-i}(n)\right|^2}\end{aligned} \tag{17}$$

or, equivalently,

$$\begin{aligned}\frac{\phi_X\left(k_{K-1},n\right)}{\phi_V\left(k_{K-1},n\right)} &\leq \frac{\sum_{i=0}^{1}\left|\beta_{K-1-i}(n)\right|^2\phi_X\left(k_{K-1-i},n\right)}{\sum_{i=0}^{1}\left|\beta_{K-1-i}(n)\right|^2\phi_V\left(k_{K-1-i},n\right)} \leq \\ \cdots &\leq \frac{\sum_{i=0}^{K-2}\left|\beta_{K-1-i}(n)\right|^2\phi_X\left(k_{K-1-i},n\right)}{\sum_{i=0}^{K-2}\left|\beta_{K-1-i}(n)\right|^2\phi_V\left(k_{K-1-i},n\right)} \leq \\ &\frac{\sum_{i=0}^{K-1}\left|\beta_{K-1-i}(n)\right|^2\phi_X\left(k_{K-1-i},n\right)}{\sum_{i=0}^{K-1}\left|\beta_{K-1-i}(n)\right|^2\phi_V\left(k_{K-1-i},n\right)},\end{aligned} \tag{18}$$

where $\beta_{K-1-i}(n)$, $i=0,1,\ldots,K-1$ are arbitrary complex numbers with at least one of them different from 0.

The previous inequalities can be easily shown by induction. It follows then that[1]

$$\lambda\left(k_{K-1},n\right) \leq \text{oSNR}\left[\mathbf{h}(n)\right] \leq \lambda\left(k_0,n\right), \forall \mathbf{h}(n), \tag{19}$$

as well as the inequalities in (6). Clearly, both the fullband input and output SNRs can never exceed the maximum subband input SNR.

4. DETERMINATION OF THE GAINS FROM THE FULLBAND OUTPUT SNR

The filter, $\mathbf{h}(n)$, that maximizes the fullband output SNR given in (11) is simply the eigenvector corresponding to the maximum eigenvalue of the matrix $\mathbf{D}_V^{-1}(n)\mathbf{D}_X(n)$. Since this matrix is diagonal, its maximum eigenvalue is its largest diagonal element, i.e., $\lambda\left(k_0,n\right)$. As a consequence, the maximum SNR filter is

$$\mathbf{h}_{\alpha_0}(n) = \alpha_0(n)\mathbf{i}_0, \tag{20}$$

where $\alpha_0(n) \neq 0$ is an arbitrary complex number and $\mathbf{i}_0$ is the first column of the $K \times K$ identity matrix, $\mathbf{I}_K$. Equivalently, we can write (20) as

$$\begin{cases} H_{\alpha_0}\left(k_0,n\right) = \alpha_0(n) \\ H\left(k_i,n\right) = 0, \quad i=1,2,\ldots,K-1 \end{cases}. \tag{21}$$

With (20), we get the maximum possible fullband output SNR, which is

$$\begin{aligned}\text{oSNR}\left[\mathbf{h}_{\alpha_0}(n)\right] &= \lambda\left(k_0,n\right) \\ &= \max_k \text{iSNR}(k,n) \geq \text{iSNR}(n).\end{aligned} \tag{22}$$

As a result,

$$\text{oSNR}\left[\mathbf{h}_{\alpha_0}(n)\right] \geq \text{oSNR}\left[\mathbf{h}(n)\right], \forall \mathbf{h}(n). \tag{23}$$

The estimate of the desired signal with the filter given (20) is

$$\begin{cases} \widehat{X}_{\alpha_0}\left(k_0,n\right) = H_{\alpha_0}\left(k_0,n\right)Y(k_0,n) \\ \widehat{X}\left(k_i,n\right) = 0, i=1,2,\ldots,K-1 \end{cases}. \tag{24}$$

Now, we need to determine $\alpha_0(n)$. There are at least two ways to find this parameter. The first one is by minimizing the mean-squared error (MSE) between $X(k_0,n)$ and $\widehat{X}_{\alpha_0}(k_0,n)$, i.e.,

$$J\left[H_{\alpha_0}\left(k_0,n\right)\right] = E\left[\left|X\left(k_0,n\right) - H_{\alpha_0}\left(k_0,n\right)Y\left(k_0,n\right)\right|^2\right]. \tag{25}$$

[1] This is also a consequence of the fullband output SNR in (11), whose form is the generalized Rayleigh quotient.

The second possibility is to minimize the distortion-based MSE, i.e.,

$$J_{\rm d}\left[H_{\alpha_0}\left(k_0,n\right)\right]=E\left[\left|X\left(k_0,n\right)-H_{\alpha_0}\left(k_0,n\right)X\left(k_0,n\right)\right|^2\right]. \tag{26}$$

The minimization of $J[H_{\alpha_0}(k_0,n)]$ leads to

$$H_{\alpha_0,{\rm W}}(n)=\frac{{\rm iSNR}\left(k_0,n\right)}{1+{\rm iSNR}\left(k_0,n\right)}, \tag{27}$$

which is the Wiener gain at the frequency bin k_0. Similarly, the minimization of $J_{\rm d}\left[H_{\alpha_0}\left(k_0,n\right)\right]$ gives

$$H_{\alpha_0,{\rm U}}(n)=1, \tag{28}$$

which is the unitary gain at the frequency bin k_0.

Expressions (27) and (28) give two different maximum fullband SNR filters. While they both maximize the fullband output SNR, these two filters may introduce a significant amount of distortion to the desired signal, since all its frequency bins are forced to be 0 except at k_0. A much better approach when we deal with broadband signals such as speech is to form the filter from a linear combination of the eigenvectors corresponding to the $P(\leq K)$ largest eigenvalues of $\mathbf{D}_V^{-1}(n)\mathbf{D}_X(n)$, i.e.,

$$\begin{aligned}\mathbf{h}_{\alpha_{0:P-1}}(n)&=\sum_{p=0}^{P-1}\alpha_p(n)\mathbf{i}_p\\&=\mathbf{h}_{\alpha_{0:P-2}}(n)+\alpha_{P-1}(n)\mathbf{i}_{P-1},\end{aligned} \tag{29}$$

where $\alpha_p(n)$, $p=0,1,\ldots,P-1$ are arbitrary complex numbers with at least one of them different from 0 and $\mathbf{i}_p$ is the $(p+1)$th column of $\mathbf{I}_K$. We can also write (29) as

$$\begin{cases}H_{\alpha_p}\left(k_p,n\right)=\alpha_p(n), & p=0,1,\ldots,P-1\\ H\left(k_i,n\right)=0, & i=P,P+1,\ldots,K-1\end{cases}. \tag{30}$$

Hence, an estimate of the desired signal is

$$\begin{cases}\widehat{X}_{\alpha_p}\left(k_p,n\right)=H_{\alpha_p}\left(k_p,n\right)Y\left(k_p,n\right), & p=0,1,\ldots,P-1\\ \widehat{X}\left(k_i,n\right)=0, & i=P,P+1,\ldots,K-1\end{cases}. \tag{31}$$

To find the α_p's, we can either optimize $J\left[H_{\alpha_0}\left(k_0,n\right)\right]$ or $J_{\rm d}\left[H_{\alpha_0}\left(k_0,n\right)\right]$. The first one leads to the Wiener gains at the frequency bins k_p, $p=0,1,\ldots,P-1$, i.e.,

$$H_{\alpha_p,{\rm W}}(k_p,n)=\frac{{\rm iSNR}(k_p,n)}{1+{\rm iSNR}(k_p,n)},\ p=0,1,\ldots,P-1, \tag{32}$$

while the second one gives the unitary gains at the frequency bins k_p, $p=0,1,\ldots,P-1$, i.e.,

$$H_{\alpha_p,{\rm U}}\left(k_p,n\right)=1,\ p=0,1,\ldots,P-1. \tag{33}$$

The filters (of length K) corresponding to (32) and (33) are, respectively,

$$\mathbf{h}_{{\rm W},P}(n)=\left[H_{\alpha_p,{\rm W}}\left(k_0,n\right)\ \ldots\ H_{\alpha_p,{\rm W}}\left(k_{P-1},n\right)\ 0\ \ldots\ 0\right]^T \tag{34}$$

and

$$\mathbf{h}_{{\rm U},P}(n)=\left[1\ \ldots\ 1\,0\ \ldots\ 0\right]^T. \tag{35}$$

For $P=K$, $\mathbf{h}_{{\rm W},K}(n)$ corresponds to the classical Wiener approach and $\mathbf{h}_{{\rm U},K}(n)$ is the identity filter, which does not affect the observations. Clearly, $\mathbf{h}_{{\rm U},P}(n)$ corresponds to the ideal binary mask [7], since the subband observation signals with the P largest subband input SNRs are not affected while the $K-P$ others with the smallest subband input SNRs are put to 0. We should always have

$${\rm oSNR}\left[\mathbf{h}_{{\rm U},P}(n)\right]\leq{\rm oSNR}\left[\mathbf{h}_{{\rm W},P}(n)\right]. \tag{36}$$

From property 3.1, we deduce that

$$\begin{aligned}{\rm iSNR}(n)&\leq{\rm oSNR}\left[\mathbf{h}_{{\rm W},K}(n)\right]\leq{\rm oSNR}\left[\mathbf{h}_{{\rm W},K-1}(n)\right]\\&\leq\cdots\leq{\rm oSNR}\left[\mathbf{h}_{{\rm W},1}(n)\right]=\lambda\left(k_0,n\right)\end{aligned} \tag{37}$$

and

$$\begin{aligned}{\rm iSNR}(n)&={\rm oSNR}\left[\mathbf{h}_{{\rm U},K}(n)\right]\leq{\rm oSNR}\left[\mathbf{h}_{{\rm U},K-1}(n)\right]\\&\leq\cdots\leq{\rm oSNR}\left[\mathbf{h}_{{\rm U},1}(n)\right]=\lambda\left(k_0,n\right).\end{aligned} \tag{38}$$

5. SIMULATIONS

We have presented the single-channel noise reduction problem in the STFT domain and discussed the design of the optimal gains to reduce noise and improve the fullband output SNR. In this section, we study the performance through experiments. The clean speech signal used in the experiments is taken from the TIMIT database [8]. The speech signals of twenty different speakers are used. The original sampling rate of the TIMIT database is 16 kHz; but we downsampled the signals into 8 kHz as we are interested in telecommunication applications with narrowband speech. The noisy signal is then generated by adding either a simulated white Gaussian noise or a prerecorded real-environment noise to the clean speech with the noise signal being properly scaled to control the input SNR. The signal is split into short frames with a frame length of 128 (a Kaiser window is applied) and an overlapping factor of 75%. A 128-point FFT is subsequently used to transform each frame into the STFT domain. An optimal filter is then constructed and applied to the noisy speech signal to reduce noise. Finally, the time-domain speech estimate is obtained by using the inverse STFT. In our simulations, the parameter P is determined as follows. We set a threshold, δ, and the value of P is equal to the number of subbands with subband input SNRs being larger than or equal to δ.

The estimate of the noise variance is achieved with the minima controlled recursive averaging approach (MCRA) [9]. We also need to know the variance of $Y(k,n)$, which is estimated using the following recursive method:

$$\widehat{\phi}_Y(k,n)=\alpha_y\widehat{\phi}_Y(k,n-1)+(1-\alpha_y)|Y(k,n)|^2, \tag{39}$$

where $0<\alpha_y<1$ is a forgetting factor. In our simulations, the value of α_y is set to 0.32. Once the variance of the noise and noisy signals are computed, the noise reduction filter is then constructed according to (34).

In order to evaluate the performance, we use three measures in our simulations: the fullband output SNR, the fullband speech distortion index [11], and the perceptual evaluation of speech quality (PESQ) [12]. Both the fullband output SNR and the fullband speech distortion index are computed in the time domain. We have 20 speakers and 10 sentences from each speaker. The fullband output SNR and the fullband speech distortion index are computed for each sentence with a long time average and results are then averaged over to obtain the overall performance. The PESQ result is computed according to [12], [13]. Briefly, for a given noise condition, the PESQ score is computed for every speaker. The PESQ mean opinion score

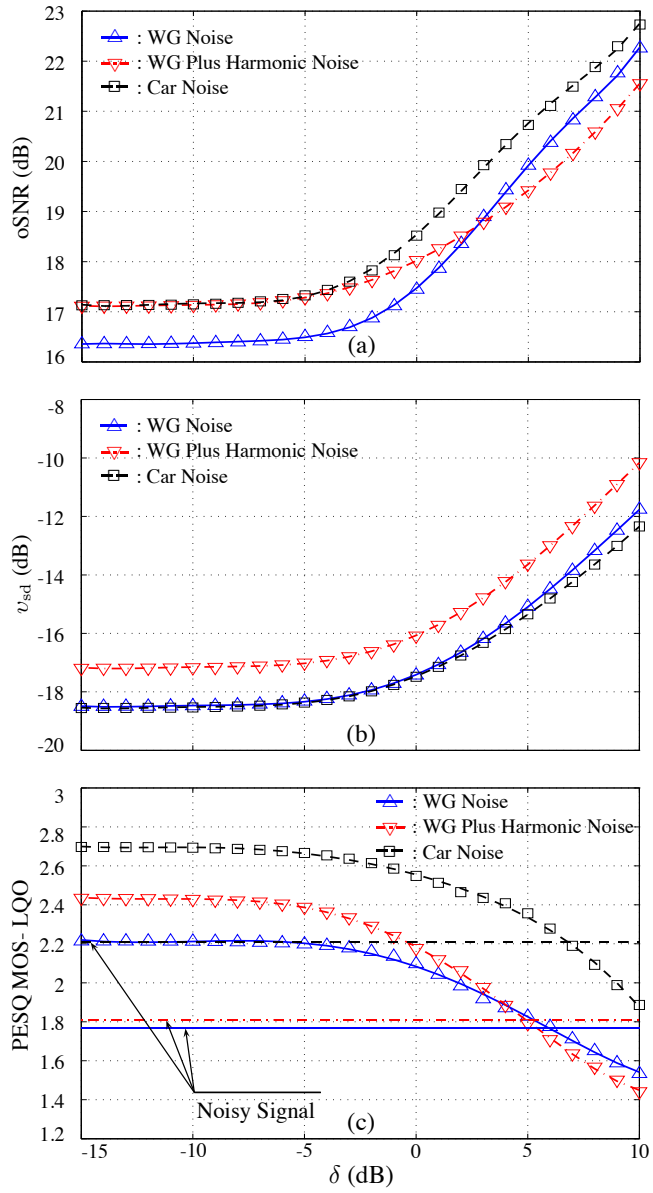

Fig. 1. Performance of the noise reduction filter, $\mathbf{h}_{\mathrm{W},P}(n)$, as a function of δ in three different types of noise (white Gaussian noise, white Gaussian plus harmonic noise, and car noise): (a) output SNR, (b) speech distortion index, and (c) PESQ MOS-LQO score. Simulation conditions: iSNR $= 10$ dB, $\alpha_y = 0.32$, and the PESQ MOS-LQO scores of the noisy signals are 1.7689, 1.8096, and 2.2097.

(MOS) is then computed by averaging all the PESQ scores. Finally, the PESQ MOS score is mapped to a PESQ MOS-LQO (listening quality objective) score with the following mapping function:

$$\mathrm{PESQ}_{\mathrm{MOS-LQO}} = 0.999 + \frac{4}{1 + e^{-1.4945 \times \mathrm{PESQ}_{\mathrm{MOS}} + 4.6607}}.$$

Various simulations were performed to evaluate the performance of the deduced algorithm and the impact of the values of different parameters on the performance. Due to space limitation, only one set of results with iSNR being 10 dB is presented, where three kinds of noise are involved, i.e., white Gaussian noise, white Gaussian plus harmonic noise (consisting of 20 sinusoid signals and their frequencies ranging from 100 Hz to 2000 Hz with an interval of 100 Hz, and the ratio between the variance of the periodic signals and that of the white noise is 6 dB), and a pre-recorded car noise.

The results of this simulation are plotted in Fig. 1. One can see that the output SNR increases dramatically with the value of δ in all the three types of noise; however, the speech distortion index also increases with δ. This is expected as when the value of δ increases, more frequency bins are forced to be zero by the noise reduction filter. Consequently, more noise is removed, but so is the speech, leading to more speech distortion. The performances of the filter in different types of noise are slightly different. In particular, one can see that the improvement of the PESQ MOS-LQO score in the harmonic noise case is higher than that in the other two types of noise, showing that the deduced algorithm is more efficient in dealing with noise that has narrowband components.

6. CONCLUSIONS

In this paper, we developed a single-channel noise reduction algorithm in the STFT domain. Unlike traditional approaches, such as the Wiener filter, that are derived by minimizing the MSE on a subband basis, the noise reduction filter in this work is derived by optimizing the fullband output SNR. The deduced filter has some resemblance to the traditional noise reduction filters in the STFT domain, but has a more general form. Indeed, the traditional Wiener, maximum SNR, and ideal binary mask filters can be viewed as particular cases.

7. REFERENCES

[1] M. R. Schroeder, "Processing of communications signals to reduce effects of noise," U.S. Patent 3 403 224, Sep., 24 1968.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, pp. 113–120, Apr. 1979.

[3] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, pp. 137–145, Apr. 1980.

[4] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise reduction in speech processing*. Berlin, Germany: Springer-Verlag, 2009.

[5] P. C. Loizou, *Speech enhancement: theory and practice*. Boca Raton, FL: CRC Press, 2013.

[6] J. Benesty, J. Chen, and E. A. P. Habets, *Speech enhancement in the STFT domain*. Springer Briefs in Electrical and Computer Engineering, 2011.

[7] D. Wang, "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech separation by humans and machines*. Pierre Divenyi, Ed., pp. 181–197, Kluwer 2005.

[8] J. W. Lyons, "DARPA TIMIT acoustic-phonetic continuous speech corpus," *Technical Report NISTIR 4930, National Institute of Standards and Technology*, 1993.

[9] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Trans. Signal Process. Lett.*, vol. 9, pp. 12–15, Jan. 2002.

[10] P. J. Wolfe and S. J. Godsill, "Simple alternatives to the ephraim and malah suppression rule for speech enhancement," in *Proc. 11th IEEE Signal Process. Workshop Statist. Signal Process.*, Aug. 2001.

[11] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, July 2006.

[12] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[13] G. Huang, J. Benesty, T. Long, and J. Chen, "A family of maximum SNR filters for noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 22, pp. 2034–2047, Dec. 2014.