INVESTIGATION OF A PARAMETRIC GAIN APPROACH TO SINGLE-CHANNEL SPEECH ENHANCEMENT

Gongping Huang[1], *Jingdong Chen*[1], *and Jacob Benesty*[2]

[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 investigates a parametric gain approach to single-channel noise reduction in the frequency domain. In comparison with the traditional parametric Wiener gain, the major novelty of this presented approach is that the parametric gain is formulated to estimate the noise by using the mean-squared error (MSE) between the noise and the noise estimate. The enhanced signal is then obtained by subtracting the noise estimate from the noisy observation signal. We show that this new method is more practical to implement and can produce better noise reduction performance as compared to the traditional parametric Wiener filtering techniques if the order of the parametric gain is not equal to 1. If the order is 1, the parametric gain is similar to the traditional Wiener gain. Simulation results are presented to illustrate the properties of this new approach.

Index Terms— Noise reduction, speech enhancement, single-channel, frequency domain, Wiener gain, parametric gain.

1. INTRODUCTION

Noise is ubiquitous and can cause significant degradation in speech quality and intelligibly in speech communication systems. To reduce the impact of the noise, noise reduction (or speech enhancement) is needed to "clean" the noisy signal before it is stored, compressed, transmitted or played back [1–4]. There are many different ways to achieve noise reduction, including filtering techniques, spectral restoration, model based methods [5–16], etc. However, one of the most widely used methods so far is the gain approach in the frequency domain or more precisely in the short-time Fourier transform (STFT) domain. In such a method, the noisy signal that is to be enhanced is partitioned into small frames and transformed into the frequency domain using the STFT. Then a gain is estimated and applied to the noisy speech spectrum in each subband to achieve noise reduction. Many different algorithms were developed over the last three decades to estimate the noise reduction gain [10–16]. Those methods differ from each other in the form of the gain as a function of the signal statistics. But they all share a common basis, i.e., the gain is formulated to directly estimate the clean speech using statistics of the observed noisy signal and estimated noise statistics.

In this paper, we investigate a parametric gain approach to noise reduction. Unlike the traditional techniques that estimate the desired, clean speech directly, we formulate the problem as to estimate the noise first, and the speech estimate is then obtained by subtracting the noise estimate from the noisy signal. We show that this new method is more practical to implement and can produce better noise reduction performance as compared to the traditional parametric Wiener filtering techniques if the order of the parametric gain is not equal to 1. If the order is 1, the parametric gain is similar to the traditional Wiener gain.

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

2. SIGNAL MODEL AND PROBLEM FORMULATION

The noise reduction (or speech enhancement) problem considered in this paper is one of recovering the desired signal $x(t)$, t being the time index, of zero mean from the noisy observation (microphone signal) [1, 3]:

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

where the zero-mean random process $v(t)$ is the unwanted additive noise, which is assumed to be independent of $x(t)$. All signals are considered to be real, stationary, and broadband.

In the frequency domain, at frequency index f, (1) can be expressed as

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

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

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

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

The objective of single-channel noise reduction in the frequency domain is then to find from the observation a "good" estimate of $X(f)$ in the sense that the additive noise is significantly reduced while the desired signal is not substantially distorted.

Traditionally, an estimate of the desired signal, $X(f)$, is obtained by applying a gain, $H(f)$, to the observation, $Y(f)$. The minimization of the corresponding mean-squared error (MSE) of this well-known concept leads to the classical Wiener gain [3]:

$$H_{\mathrm{W}}(f) = \frac{\phi_X(f)}{\phi_Y(f)} = \frac{\mathrm{iSNR}(f)}{1 + \mathrm{iSNR}(f)}, \tag{4}$$

where

$$\mathrm{iSNR}(f) = \frac{\phi_X(f)}{\phi_V(f)} \tag{5}$$

is the narrowband input signal-to-noise ratio (SNR). It is clear that this gain is always real and $0 \leq H_{\mathrm{W}}(f) \leq 1$. Therefore, the optimal estimate of $X(f)$ in the minimum MSE sense is

$$\widehat{X}_{\mathrm{W}}(f) = H_{\mathrm{W}}(f)Y(f) = e^{\jmath\theta_Y(f)} H_{\mathrm{W}}(f) \left|Y(f)\right|, \tag{6}$$

where $\jmath$ is the imaginary unit with $\jmath^2 = -1$ and $\theta_Y(f)$ is the phase of $Y(f)$.

Alternatively, we can also estimate the noise signal, $V(f)$, by applying a gain, $H'(f)$, to the observation, $Y(f)$. By using the MSE criterion, we easily find that the optimal gain

$$H'_{\mathrm{W}}(f) = \frac{\phi_V(f)}{\phi_Y(f)} = \frac{1}{1 + \mathrm{iSNR}(f)}, \tag{7}$$

and the corresponding estimator is

$$\widehat{V}_{\mathrm{W}}(f) = H'_{\mathrm{W}}(f)Y(f). \tag{8}$$

As a consequence, the optimal estimate of $X(f)$ is

$$\widehat{X}'_{\mathrm{W}}(f) = Y(f) - \widehat{V}_{\mathrm{W}}(f) = \widehat{X}_{\mathrm{W}}(f). \tag{9}$$

Obviously, the two methods are strictly equivalent. However, when a parametric gain is used, the two methods may generate different performance as will be discussed in the next section. Generally, the second technique is preferred in practice because $H'_{\mathrm{W}}(f)$ [which depends explicitly on the statistics of $V(f)$ and $Y(f)$], is easier to estimate than $H_{\mathrm{W}}(f)$ [which depends explicitly on the statistics of $X(f)$ and $Y(f)$]

It can be shown [3] that the broadband output SNR with the Wiener gain, which is defined as

$$\mathrm{oSNR}\left(H_{\mathrm{W}}\right) = \frac{\int_f H^2_{\mathrm{W}}(f)\phi_X(f)df}{\int_f H^2_{\mathrm{W}}(f)\phi_V(f)df} \tag{10}$$

is always greater than or equal to the broadband input SNR, which is given by

$$\mathrm{iSNR} = \frac{\int_f \phi_X(f)df}{\int_f \phi_V(f)df}. \tag{11}$$

However, the narrowband output SNR with the Wiener gain is always equal to the narrowband input SNR.

3. PARAMETRIC APPROACH

In this section, we propose to estimate $|V(f)|^\alpha$, with $\alpha > 0$ (the parametric order), by applying a gain, $H'_\alpha(f)$, to the modified observation, $|Y(f)|^\alpha$. Since $H'_\alpha(f)\left|Y(f)\right|^\alpha$ is an estimate of $|V(f)|^\alpha$, then a natural estimate of $V(f)$ is

$$\widehat{V}_\alpha(f) = e^{\jmath\theta_Y(f)} \left[H'_\alpha(f)\left|Y(f)\right|^\alpha\right]^{1/\alpha} = \sqrt[\alpha]{H'_\alpha(f)}Y(f). \tag{12}$$

We deduce that an estimate of $X(f)$ is

$$\begin{aligned}\widehat{X}_\alpha(f) &= Y(f) - \widehat{V}_\alpha(f) \\ &= \left[1 - \sqrt[\alpha]{H'_\alpha(f)}\right] Y(f) \\ &= H_\alpha(f)Y(f),\end{aligned} \tag{13}$$

where

$$H_\alpha(f) = 1 - \sqrt[\alpha]{H'_\alpha(f)} \tag{14}$$

is the equivalent gain for the estimation of $X(f)$ from $Y(f)$.

We define the error signal between the signal of interest and its estimate as

$$\mathcal{E}_\alpha(f) = \left|V(f)\right|^\alpha - H'_\alpha(f)\left|Y(f)\right|^\alpha, \tag{15}$$

from which we deduce the parametric MSE criterion:

$$\begin{aligned}J\left[H'_\alpha(f)\right] &= E\left[\mathcal{E}^2_\alpha(f)\right] \\ &= \phi_{|V|,\alpha}(f) + H'^2_\alpha(f)\phi_{|Y|,\alpha}(f) - 2H'_\alpha(f)\phi_{|V||Y|,\alpha}(f),\end{aligned} \tag{16}$$

where $\phi_{|V|,\alpha}(f) = E\left[|V(f)|^{2\alpha}\right]$, $\phi_{|Y|,\alpha}(f) = E\left[|Y(f)|^{2\alpha}\right]$, and $\phi_{|V||Y|,\alpha}(f) = E\left[|V(f)|^\alpha \left|Y(f)\right|^\alpha\right]$.

The minimization of $J\left[H'_\alpha(f)\right]$ with respect to $H'_\alpha(f)$ leads to the optimal gains for the estimation of $V(f)$ and $X(f)$, respectively, i.e.,

$$H'_{\alpha,\mathrm{o}}(f) = \frac{\phi_{|V||Y|,\alpha}(f)}{\phi_{|Y|,\alpha}(f)} \tag{17}$$

and

$$H_{\alpha,\mathrm{o}}(f) = 1 - \sqrt[\alpha]{H'_{\alpha,\mathrm{o}}(f)} = 1 - \sqrt[\alpha]{\frac{\phi_{|V||Y|,\alpha}(f)}{\phi_{|Y|,\alpha}(f)}}. \tag{18}$$

The quantity $\phi_{|Y|,\alpha}(f)$ can be easily estimated from the observations while the quantity $\phi_{|V||Y|,\alpha}(f)$ can be estimated from the components of the noise during silences [even with a delay between $Y(f)$ and $V(f)$] since no phases are involved in the expression. Substituting (17) into (16), we find that the minimum MSE is

$$\begin{aligned}J\left[H'_{\alpha,\mathrm{o}}(f)\right] &= \phi_{|V|,\alpha}(f)\left[1 - \frac{\phi^2_{|V||Y|,\alpha}(f)}{\phi_{|V|,\alpha}(f)\phi_{|Y|,\alpha}(f)}\right] \\ &= \phi_{|V|,\alpha}(f)\left[1 - \gamma^2_{|V||Y|,\alpha}(f)\right],\end{aligned} \tag{19}$$

where

$$\gamma^2_{|V||Y|,\alpha}(f) = \frac{\phi^2_{|V||Y|,\alpha}(f)}{\phi_{|V|,\alpha}(f)\phi_{|Y|,\alpha}(f)}, \tag{20}$$

with $0 \leq \gamma^2_{|V||Y|,\alpha}(f) \leq 1$. Obviously, the optimal estimate of $X(f)$ in this context is

$$\widehat{X}_{\alpha,\mathrm{o}}(f) = H_{\alpha,\mathrm{o}}(f)Y(f). \tag{21}$$

We can express $H_{\alpha,\mathrm{o}}(f)$ as a function of $\gamma_{|V||Y|,\alpha}(f)$, i.e.,

$$H_{\alpha,\mathrm{o}}(f) = 1 - \sqrt[\alpha]{\gamma_{|V||Y|,\alpha}(f)\sqrt{\frac{\phi_{|V|,\alpha}(f)}{\phi_{|Y|,\alpha}(f)}}}. \tag{22}$$

We observe from the previous equation that $0 \leq H_{\alpha,\mathrm{o}}(f) \leq 1$.

4. PERFORMANCE MEASURES

In this paper, we adopt the output SNR, the noise reduction factor, the speech distortion index, and the perceptual evaluation of speech quality (PESQ) [19, 20] as the the performance measures for an objective evaluation of the parametric gain approach presented in the previous section.

The narrowband and broadband input SNRs were already defined in Section 2. Since we deal with gains, the narrowband output

SNR is equal to the narrowband input SNR. The broadband input SNR is defined as

$$\text{oSNR}\left(H_{\alpha,\text{o}}\right) = \frac{\int_f H_{\alpha,\text{o}}^2(f)\phi_X(f)df}{\int_f H_{\alpha,\text{o}}^2(f)\phi_V(f)df}. \tag{23}$$

The noise reduction factor quantifies the amount of noise rejected by the filtering process [18]. The narrowband and broadband noise reduction factors are defined respectively as [4]

$$\xi_{\text{nr}}\left[H_{\alpha,\text{o}}(f)\right] = \frac{\phi_V(f)}{H_{\alpha,\text{o}}^2(f)\phi_V(f)} = \frac{1}{H_{\alpha,\text{o}}^2(f)}, \tag{24}$$

$$\begin{aligned}\xi_{\text{nr}}\left(H_{\alpha,\text{o}}\right) &= \frac{\int_f \phi_V(f)df}{\int_f H_{\alpha,\text{o}}^2(f)\phi_V(f)df} \\ &= \frac{\int_f \phi_V(f)df}{\int_f \phi_V(f) \times \xi_{\text{nr}}^{-1}\left[H_{\alpha,\text{o}}(f)\right]df}.\end{aligned} \tag{25}$$

We always have $\xi_{\text{nr}}\left[H_{\alpha,\text{o}}(f)\right] \geq 1$ and $\xi_{\text{nr}}\left(H_{\alpha,\text{o}}\right) \geq 1$.

The gain $H_{\alpha,\text{o}}(f)$ adds distortion to the desired signal, $X(f)$. In order to evaluate the level of this distortion, the so-called speech reduction factor was introduced, which is defined as the variance of the desired signal over the variance of the filtered version of the desired signal [4]. For the problem described in this paper, the narrowband and broadband speech reduction factors can be defined as

$$\xi_{\text{sr}}\left[H_{\alpha,\text{o}}(f)\right] = \frac{\phi_X(f)}{H_{\alpha,\text{o}}^2(f)\phi_X(f)} = \frac{1}{H_{\alpha,\text{o}}^2(f)}, \tag{26}$$

$$\begin{aligned}\xi_{\text{sr}}\left(H_{\alpha,\text{o}}\right) &= \frac{\int_f \phi_X(f)df}{\int_f H_{\alpha,\text{o}}^2(f)\phi_X(f)df} \\ &= \frac{\int_f \phi_X(f)df}{\int_f \phi_X(f) \times \xi_{\text{sr}}^{-1}\left[H_{\alpha,\text{o}}(f)\right]df}.\end{aligned} \tag{27}$$

We see that $\xi_{\text{sr}}\left[H_{\alpha,\text{o}}(f)\right] \geq 1$ and $\xi_{\text{sr}}\left(H_{\alpha,\text{o}}\right) \geq 1$.

It is clear that we always have

$$\frac{\text{oSNR}\left(H_{\alpha,\text{o}}\right)}{\text{iSNR}} = \frac{\xi_{\text{nr}}\left(H_{\alpha,\text{o}}\right)}{\xi_{\text{sr}}\left(H_{\alpha,\text{o}}\right)}. \tag{28}$$

Another useful performance measure is the speech distortion index defined as

$$\begin{aligned}\upsilon_{\text{sd}}\left[H_{\alpha,\text{o}}(f)\right] &= \frac{E\left[\left|H_{\alpha,\text{o}}(f)X(f) - X(f)\right|^2\right]}{\phi_X(f)} \\ &= \left[H_{\alpha,\text{o}}(f) - 1\right]^2\end{aligned} \tag{29}$$

in the narrowband case and as

$$\begin{aligned}\upsilon_{\text{sd}}\left(H_{\alpha,\text{o}}\right) &= \frac{\int_f E\left[\left|H_{\alpha,\text{o}}(f)X(f) - X(f)\right|^2\right]df}{\int_f \phi_X(f)df} \\ &= \frac{\int_f \phi_X(f) \times \upsilon_{\text{sd}}\left[H_{\alpha,\text{o}}(f)\right]df}{\int_f \phi_X(f)df}\end{aligned} \tag{30}$$

in the broadband case. It can be checked that $0 \leq \upsilon_{\text{sd}}\left[H_{\alpha,\text{o}}(f)\right] \leq 1$ and $0 \leq \upsilon_{\text{sd}}\left(H_{\alpha,\text{o}}\right) \leq 1$.

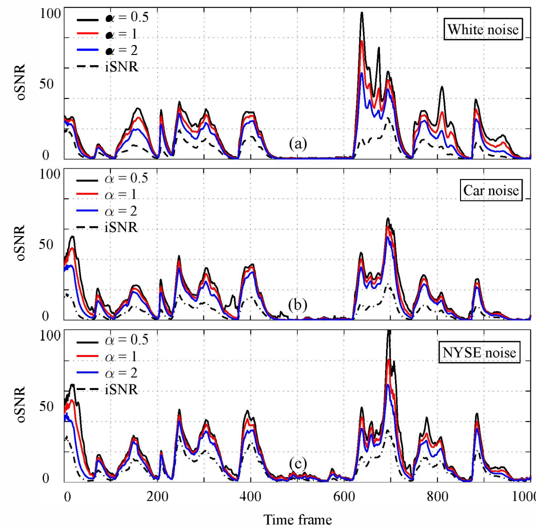

Fig. 1. Broadband input and output SNRs of the parametric gain for different orders over time in different noise cases: (a) white Gaussian noise, (b) car noise, and (c) NYSE noise. The time-domain input SNR is 10 dB.

5. SIMULATIONS

We have formulated the single-channel parametric gain approach to the problem of noise reduction in Section 3. In this section, we study the performance of this approach through experiments. Meanwhile, we attempt to address two fundamental questions: 1) why are we interested in the parametric gain approach instead of using the Wiener gain and 2) why do we want to estimate the noise instead of the desired speech directly?

The clean speech signals are recorded from a female and a male talker in a quiet office room. We consider three types of noise: a white Gaussian random process, a babble noise signal recorded in a New York Stock Exchange (NYSE) room, and a car noise signal recorded in a sedan running 50 MPH on a highway. All the signals are 30 seconds long and the sampling frequency is 8 kHz. The noisy signal is obtained by adding noise into the clean speech with a specified input SNR level.

To implement the parametric gain, the time domain signals are partitioned into overlapping frames (the frame size is 128 and the overlapping factor is 75%) with a Kaiser window, and then transformed into the STFT domain using a 128-point FFT. The noise variance $\phi_{|V|,\alpha}(f)$ and $|V(f)|^\alpha$ are blindly estimated from the noisy speech signal using a variant of the minimum controlled recursive average (MCRA) algorithm [10]. The quantity $\phi_{|Y|,\alpha}(f)$ and $\phi_{|V||Y|,\alpha}(f)$ are computed using $Y(f)$ and the estimated noise spectrum and noise statistics with a recursive method where the forgetting factor is set to 0.6. After enhancement, the signals are transformed into the time domain using the inverse STFT with an overlap add method.

In the first experiment, we examine the performance of the parametric gain method in different noise cases. Figure 1 plot the broadband input and output SNRs in three different noise cases (white Gaussian, car, and NYSE), where $\alpha = \{0.5, 1, 2\}$. The time-domain input SNR, which is calculated using a long term average, is 10 dB. We see that the parametric gain can improve the broadband SNR significantly in all three studied cases. We also see that the output SNR deceases as the value of α increases. So, the value of α plays an important role on the noise reduction performance. To see this more clearly, we carried out a set of experiments. Figure 2 plot the noise

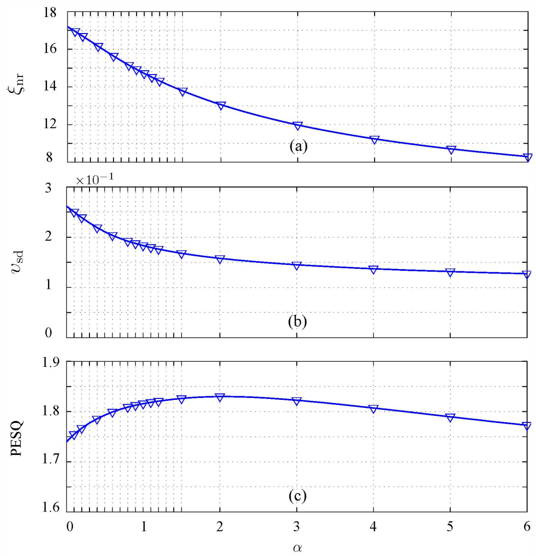

Fig. 2. Performance of the parametric gain as a function of the parameter α in white Gaussian noise with an input SNR of 0 dB: (a) noise reduction factor, (b) speech distortion index, and (c) PESQ score.

reduction factor, speech distortion index, and the PESQ score as a function of α in white Gaussian noise with an input SNR of 0 dB. Note that the broadband noise reduction factor and speech distortion index plotted in this figure are the average results over all the frames. The PESQ score is computed as follows. For each speaker (we have a male and female speaker), the PESQ score is computed by comparing the enhanced signal with the original clean speech. The final PESQ score is obtained by mapping these two scores to the PESQ listening quality objective score [20].

As we can see from Fig. 2, both the noise reduction factor and speech distortion index decrease as the value of α increases. The underlying reason can be explained as follows. The parametric gain in (18) is an increasing function of α. It can be checked that $\lim_{\alpha\to 0} H_{\alpha,\rm o}(f) = 0$ and $\lim_{\alpha\to +\infty} H_{\alpha,\rm o}(f) = 1$. Therefore, if $\alpha = 0$, all the noise is removed, and so is the speech signal, which means that both the noise reduction and the speech distortion are maximized. If $\alpha \to +\infty$, there will be neither speech distortion or noise reduction. When the value of α increases, the parametric gain will add less speech distortion, but also reduce less noise. In comparison, the PESQ score is not a monotonic function with respect to α. It first increases and then decreases as the value of α increases. The highest PESQ score is obtained when α is approximately 2 in the condition of white noise with 0-dB input SNR. This answers the question why we are interested in the parametric gain approach. Briefly, the parametric gain approach can yield better noise reduction performance as compared to the Wiener gain if the parameter α is properly chosen.

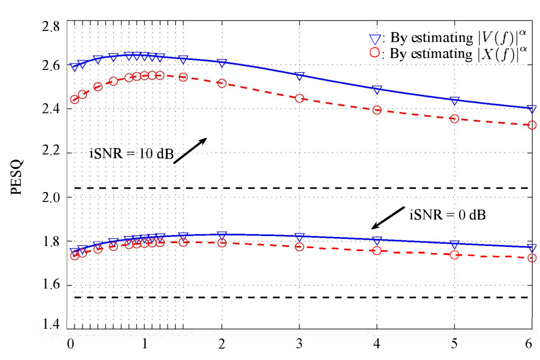

Fig. 3. PESQ score as a function of the parameter α in white Gaussian noise. The two dashed black lines show the PESQ score of the noisy signal in 0 dB and 10 dB, respectively.

In the second experiment, we study the difference between estimating directly the desired signal, $X(f)$, and estimating directly the noise. Following the same principles of Section 3, we can derive the gain to directly estimate $|X(f)|^\alpha$, i.e.,

$$G_{\alpha,\rm o}(f) = \sqrt[\alpha]{\frac{\phi_{|X||Y|,\alpha}(f)}{\phi_{|Y|,\alpha}(f)}}, \tag{31}$$

where

$$\begin{aligned}\phi_{|X||Y|,\alpha}(f) &= E\left[|X(f)|^\alpha\ |Y(f)|^\alpha\right] \\ &= E\left[\left(|Y(f)|^2 + |V(f)|^2 - 2|V(f)Y(f)|\right)^{\frac{\alpha}{2}} |Y(f)|^\alpha\right]. \end{aligned}\tag{32}$$

Since $X(f)$ is not known in practice, it is difficult to directly estimate $\phi_{|X||Y|,\alpha}(f)$. Generally, however, it is assumed that we can estimate $|V(f)|^2$ and $\phi_V(f)$. Substituting these noise statistics into (32), one can compute the gain in an indirect way.

Figure 3 plot the PESQ scores for enhanced signal using $G_{\alpha,\rm o}(f)$ and $H_{\alpha,\rm o}(f)$. One can see that the PESQ score with the use of $H_{\alpha,\rm o}(f)$ is higher than that of using $G_{\alpha,\rm o}(f)$. This justifies the reason why we need to estimate the noise instead of the desired speech directly. It is observed that the value of α to produce the largest PESQ score changes with the input SNR. So, in practice, it would be better if the value of α is determined adaptively according to the SNR condition, which will be our future work.

6. CONCLUSIONS

In this paper, we studied the single-channel noise reduction problem in the frequency domain with a parametric gain approach. Instead of directly estimating the desired speech signal, we formulated the problem as first to estimate the noise and then the speech estimate is obtained by subtracting the noise estimate from the noisy signal. The optimal parametric gain in the minimum MSE sense was derived to obtain the noise estimate. With experiments, we illustrated the advantage of using the parametric approach over the traditional Wiener gain. It was also demonstrated why we should formulate the problem as estimating the noise first instead of the desired speech.

7. RELATION TO PRIOR WORK

Various noise reduction methods have been developed over the past several decades [1]– [18], among which the gain approach in the short-time Fourier transform (STFT) domain is by far one of the most widely used methods in practice. Traditionally, the gain is formulated to directly estimate the desired clean speech using noisy and noise signal statistics. In this paper, we reformulate the problem as one of estimating the noise first; and then the speech estimate is obtained by subtracting the noise estimate from the noisy signal. We show that this new method is more practical to implement and can produce better noise reduction performance as compared to the traditional parametric Wiener filtering techniques if the order of the parametric gain is not equal to 1. If the order is 1, the parametric gain is similar to the traditional Wiener gain.

8. REFERENCES

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

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

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

[4] J. Benesty, J. Chen, and E. Habets, *Speech Enhancement in the STFT Domain.* Springer Briefs in Electrical and Computer Engineering, 2012.

[5] J. Chen, J. Benesty, Y. Huang, and E. J. Diethorn, "Fundamentals of Noise Reduction," in *Springer Handbook on Speech Processing and Speech Communication*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., Berlin: Springer-Verlag, pp. 843–871, 2007.

[6] P. C. Yong, S. Nordholm, and H. H. Dam, "Trade-off evaluation for speech enhancement algorithms with respect to the a priori SNR estimation," in *Proc. IEEE ICASSP*, 2012, pp. 4657–4660.

[7] J. H. L. Hansen, V. Radhakrishnan, and K. H. Arehart, "Speech enhancement based on generalized minimum mean square error estimators and masking properties of the auditory system," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, pp. 2049–2063, Nov. 2006.

[8] Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms," *J. Acoust. Soc. Am.*, vol. 122, pp. 1777–1786, Sept. 2007.

[9] L. S. Yong, G. Nedelko and N. Sven, "Speech enhancement using multiple soft constrained subband beamformers and non-coherent technique," in *Proc. IEEE ICASSP*, 2003, pp. 489–492.

[10] I. Cohen, "Optimal Speech Enhancement Under Signal Presence Uncertainty Using Log-Spectral Amplitude Estimator," *IEEE Signal Process. Lett.*, vol. 9, pp. 113–126, Apr. 2002.

[11] E. Plourde and B. Champagne, "Generalized Bayesian estimators of the spectral amplitude for speech enhancement," *IEEE Signal Process. Lett.*, vol. 16, pp. 485–488, Jun. 2009.

[12] C. H. You, S. N. Koh, and S. Rahardja, "β-order MMSE spectral amplitude estimation for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 13, pp. 475–486, Jul. 2005.

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

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

[15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 1109–1121, Dec. 1984.

[16] C. Breithaupt, M. Krawczyk, and R. Martin, "Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech," in *Proc. IEEE ICASSP*, 2008, pp. 4037–4040.

[17] Y. Huang and J. Benesty, "A multi-frame approach to the frequency-domain single-channel noise reduction problem," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, pp. 1256–1269, 2012.

[18] 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, pp. 1218–1234, Jul. 2006.

[19] ITU-T P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, ITU-T Recommendation P.862.

[20] Mapping Function for Transforming Raw Result Scores to MOS-LQO, ITU-T Rec. P. 862.1, 2003.