

A PROBABILISTIC SPEECH ENHANCEMENT FILTER UTILIZING THE CONSTRUCTIVE AND DESTRUCTIVE INTERFERENCE OF NOISE

Taufiq Hasan and Md. Kamrul Hasan

Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
e-mail: taufiq.hasan@gmail.com, khasan@eee.buet.ac.bd

ABSTRACT

A new probabilistic speech enhancement filter is presented in this paper considering the three state possibilities of discrete cosine transform (DCT) coefficients of noisy speech: speech absence, speech and noise are constructive, and destructive. The conditional probabilities of the three events are calculated using Gaussian approximations. Unlike conventional fixed values, the speech presence or absence probability in different spectral coefficients is experimentally calculated. A novel set of gain functions is proposed for accommodation of the aforesaid three possibilities, and merged into one, called the expected gain. It is used on the noisy speech component for enhancement. Experimental results are presented to show the effectiveness of the proposed denoising filter.

1. INTRODUCTION

Speech enhancement has been a challenging task for signal processing researchers for decades. Different researchers proposed different methods and ideas for suppression of the unwanted noise, corrupting the speech in practical conditions. Development and widespread deployment of digital communication systems during the last twenty years have brought increased attention to the role of speech enhancement in speech processing problems [1]-[6].

There are a variety of approaches for retrieving speech signal from noisy observations such as the traditional Wiener filtering [6], spectral subtraction rules [1]-[3], power spectral estimation, coefficient thresholding [5], Kalman filtering and perceptual filtering. Among them the Wiener and spectral subtraction type algorithms are widely used because of their low computational complexity and impressive performance. In general, using the family of spectral subtraction type algorithms the enhanced speech spectrum is obtained by subtracting an average noise spectrum from the noisy speech spectrum or by multiplying the noisy spectrum with a gain function [4]. The phase of the noisy speech is kept unchanged since it is assumed that the phase distortion is not perceived by human ear. The main shortcoming of this method, however, is that it introduces musical noise in the enhanced speech.

The reduction of musical noise using only an attenuating filter gain is a difficult task. In effect all of the mentioned approaches can be generalized as a amplitude reduction type algorithm, except [10] where a low distortion dual gain Wiener filter has been reported based on the constructive and destructive interference of speech and noise in the frequency domain. It is experimentally observed that in most spectral components, the speech and noise are additive or a

noisy component is noise dominant, which is the reason behind the overall success of the attenuative gain type speech enhancement techniques. In this paper the idea presented in [10] of a dual gain Wiener filter has been taken a step further by introducing three different gains with a rigorous statistical analysis. Furthermore, the limitations of the dual gain Wiener filter is discussed and a new idea of probabilistic filter gain is presented.

We begin the analysis in a similar approach as [8] to formulate a soft decision Wiener filter considering speech presence and absence probability. The assumption of speech presence and absence probability being equal has been modified by performing extensive experiments on average human speech spectrum. Next we extend the concept by determining the conditional probability of speech and noise being constructive or destructive, given speech is present in that particular spectral component. A set of gains is proposed for use in the three different cases and the expected value of these gains is used for speech enhancement.

2. PROPOSED METHOD

Let $x(t)$, $d(t)$ and $y(t)$ denote the clean speech, noise and noisy speech samples, respectively for the t -th sample in time domain. If it is assumed that the noise is additive, $y(t)$ can be expressed as

$$y(t) = x(t) + d(t). \quad (1)$$

The discrete cosine transform (DCT) domain representation of (1) in the n -th frame and k -th frequency index is

$$Y_{n,k} = X_{n,k} + D_{n,k} \quad (2)$$

where $X_{n,k}$, $D_{n,k}$ and $Y_{n,k}$ are the DCT coefficients of clean speech, noise and noisy speech, respectively. Since DCT coefficients are real, $X_{n,k}$ and $D_{n,k}$ can be constructive or destructive, depending on their polarity. Again, speech can be absent in the sample $Y_{n,k}$, that is speech amplitude can be negligible comparing to the noise amplitude. Thus we define three mutually exclusive events that can occur for a noisy component Y_k :

$$\begin{aligned} H_0 &: \text{Speech is absent} && : Y_k = D_k, \\ H_+ &: \text{Speech and noise are constructive} && : X_k D_k > 0, \\ H_- &: \text{Speech and noise are destructive} && : X_k D_k < 0. \end{aligned}$$

The probability that $Y_{n,k}$ is in state H_0 is obtained accord-

ing to Bayes rule,

$$\begin{aligned} p(H_0|Y_{n,k}) &= \frac{p(Y_{n,k}|H_0)p(H_0)}{\sum_i p(Y_{n,k}|H_i)p(H_i)} \\ &= \frac{p(H_0)}{p(H_0) + p(H_+) \frac{p(Y_{n,k}|H_+)}{p(Y_{n,k}|H_0)} + p(H_-) \frac{p(Y_{n,k}|H_-)}{p(Y_{n,k}|H_0)}} \end{aligned} \quad (3)$$

Now assuming $X_{n,k}$ and $D_{n,k}$ as Gaussian we can write,

$$\begin{aligned} p(Y_{n,k}|H_0) &= \frac{1}{\sqrt{2\pi E\{Y_{n,k}^2|H_0\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_0\}}\right) \\ &= \frac{1}{\sqrt{2\pi E\{D_{n,k}^2\}}} \exp\left(-\frac{D_{n,k}^2}{2E\{D_{n,k}^2\}}\right) \end{aligned} \quad (4)$$

Also for the events H_+ and H_- ,

$$p(Y_{n,k}|H_+) = \frac{1}{\sqrt{2\pi E\{Y_{n,k}^2|H_+\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_+\}}\right) \quad (5)$$

$$p(Y_{n,k}|H_-) = \frac{1}{\sqrt{2\pi E\{Y_{n,k}^2|H_-\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_-\}}\right) \quad (6)$$

The assumption of normal distribution for events H_+ and H_- is loosely valid, and is used only for simplification. Now we determine the unknown terms in (3). Firstly,

$$\frac{p(Y_{n,k}|H_+)}{p(Y_{n,k}|H_0)} = \sqrt{\frac{E\{Y_{n,k}^2|H_0\}}{E\{Y_{n,k}^2|H_+\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_+\}} + \frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_0\}}\right) \quad (7)$$

and secondly,

$$\frac{p(Y_{n,k}|H_-)}{p(Y_{n,k}|H_0)} = \sqrt{\frac{E\{Y_{n,k}^2|H_0\}}{E\{Y_{n,k}^2|H_-\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_-\}} + \frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_0\}}\right) \quad (8)$$

Now we need to determine the quantities,

$$E\{Y_{n,k}^2|H_+\} = E\{X_{n,k}^2\} + E\{D_{n,k}^2\} + 2E\{X_{n,k}D_{n,k}|H_+\} \quad (9)$$

$$E\{Y_{n,k}^2|H_-\} = E\{X_{n,k}^2\} + E\{D_{n,k}^2\} + 2E\{X_{n,k}D_{n,k}|H_-\} \quad (10)$$

It can be showed that for a Gaussian random variable, $E\{|X|\} = \sqrt{\frac{2}{\pi}}\sigma_X$. Thus,

$$E\{X_{n,k}D_{n,k}|H_+\} = E\{|X_{n,k}||D_{n,k}\} = \frac{2}{\pi}\sigma_X(k)\sigma_D(k) \quad (11)$$

and

$$E\{X_{n,k}D_{n,k}|H_-\} = -E\{|X_{n,k}||D_{n,k}\} = -\frac{2}{\pi}\sigma_X(k)\sigma_D(k) \quad (12)$$

Using these quantities in (9) and (10), substituting the values of $E\{Y_{n,k}^2|H_+\}$ and $E\{Y_{n,k}^2|H_-\}$ into (7) and (8), and upon simplification we have,

$$\frac{p(Y_{n,k}|H_+)}{p(Y_{n,k}|H_0)} = \frac{1}{\sqrt{\xi_{n,k} + 1 + \frac{4}{\pi}\sqrt{\xi_{n,k}}}} \exp\left(\frac{\frac{\gamma_{n,k}}{2}(\xi_{n,k} + \frac{4}{\pi}\sqrt{\xi_{n,k}})}{\xi_{n,k} + 1 + \frac{4}{\pi}\sqrt{\xi_{n,k}}}\right) \quad (13)$$

and

$$\frac{p(Y_{n,k}|H_-)}{p(Y_{n,k}|H_0)} = \frac{1}{\sqrt{\xi_{n,k} + 1 - \frac{4}{\pi}\sqrt{\xi_{n,k}}}} \exp\left(\frac{\frac{\gamma_{n,k}}{2}(\xi_{n,k} - \frac{4}{\pi}\sqrt{\xi_{n,k}})}{\xi_{n,k} + 1 - \frac{4}{\pi}\sqrt{\xi_{n,k}}}\right) \quad (14)$$

where,

$$\begin{aligned} \gamma_{n,k} &= \text{SNR}_{\text{post}}(n,k) = \frac{Y_{n,k}^2}{E\{D_{n,k}^2\}} \\ \xi_{n,k} &= \text{SNR}_{\text{prior}}(n,k) = \frac{E\{X_{n,k}^2\}}{E\{D_{n,k}^2\}} \end{aligned}$$

Using (13) and (14), we have from (3),

$$p(H_0|Y_{n,k}) = \frac{p(H_0)}{p(H_0) + p(H_+)f_{n,k} + p(H_-)\phi_{n,k}} \quad (15)$$

where,

$$f_{n,k} = \frac{1}{\sqrt{\xi_{n,k} + 1 + \frac{4}{\pi}\sqrt{\xi_{n,k}}}} \exp\left(\frac{\frac{\gamma_{n,k}}{2}(\xi_{n,k} + \frac{4}{\pi}\sqrt{\xi_{n,k}})}{\xi_{n,k} + 1 + \frac{4}{\pi}\sqrt{\xi_{n,k}}}\right) \quad (16)$$

and

$$\phi_{n,k} = \frac{1}{\sqrt{\xi_{n,k} + 1 - \frac{4}{\pi}\sqrt{\xi_{n,k}}}} \exp\left(\frac{\frac{\gamma_{n,k}}{2}(\xi_{n,k} - \frac{4}{\pi}\sqrt{\xi_{n,k}})}{\xi_{n,k} + 1 - \frac{4}{\pi}\sqrt{\xi_{n,k}}}\right) \quad (17)$$

Similarly the probability that $Y_{n,k}$ is in state H_+ is obtained using Bayes rule,

$$\begin{aligned} p(H_+|Y_{n,k}) &= \frac{p(Y_{n,k}|H_+)p(H_+)}{\sum_i p(Y_{n,k}|H_i)p(H_i)} \\ &= \frac{p(H_+)}{p(H_+) + p(H_0) \frac{p(Y_{n,k}|H_0)}{p(Y_{n,k}|H_+)} + p(H_-) \frac{p(Y_{n,k}|H_-)}{p(Y_{n,k}|H_+)}} \end{aligned} \quad (18)$$

Now, approaching as before,

$$\frac{p(Y_{n,k}|H_-)}{p(Y_{n,k}|H_+)} = \sqrt{\frac{E\{Y_{n,k}^2|H_+\}}{E\{Y_{n,k}^2|H_-\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_-\}} + \frac{Y_{n,k}^2}{2E\{Y_{n,k}^2|H_+\}}\right)$$

using the quantities of (9), (10) and using similar simplification, we obtain,

$$\frac{p(Y_{n,k}|H_-)}{p(Y_{n,k}|H_+)} = \sqrt{\frac{\xi_{n,k} + 1 + \frac{4}{\pi}\sqrt{\xi_{n,k}}}{\xi_{n,k} + 1 - \frac{4}{\pi}\sqrt{\xi_{n,k}}}} \exp\left(\frac{-\frac{4\gamma_{n,k}}{\pi}\sqrt{\xi_{n,k}}}{(\xi_{n,k} + 1)^2 - \frac{16}{\pi^2}\xi_{n,k}}\right)$$

Thus,

$$p(H_+|Y_{n,k}) = \frac{p(H_+)}{p(H_+) + p(H_0)f_{n,k}^{-1} + p(H_-)\psi_{n,k}} \quad (19)$$

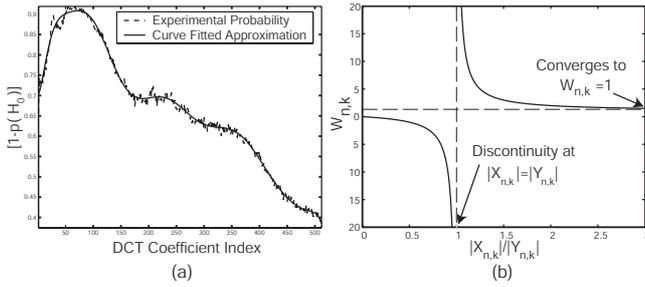


Figure 1: (a) Speech presence probability in different DCT coefficients (experimental (· ·) and fitted curve (—)), (b) Variation of ideal filter gain when signal and noise are destructive.

where,

$$\psi_{n,k} = \sqrt{\frac{\xi_{n,k} + 1 + \frac{4}{\pi} \sqrt{\xi_{n,k}}}{\xi_{n,k} + 1 - \frac{4}{\pi} \sqrt{\xi_{n,k}}}} \exp\left(\frac{-\frac{4\gamma_{n,k}}{\pi} \sqrt{\xi_{n,k}}}{(\xi_{n,k} + 1)^2 - \frac{16}{\pi^2} \xi_{n,k}}\right). \quad (20)$$

Similarly the probability that $Y_{n,k}$ is in state H_- will be given as,

$$p(H_- | Y_{n,k}) = \frac{p(H_-)}{p(H_-) + p(H_0) \frac{p(Y_{n,k}|H_0)}{p(Y_{n,k}|H_-)} + p(H_+) \frac{p(Y_{n,k}|H_+)}{p(Y_{n,k}|H_-)}} \quad (21)$$

Substituting (17) and (20) into (21), we obtain

$$p(H_- | Y_{n,k}) = \frac{p(H_-)}{p(H_-) + p(H_0) \phi_{n,k}^{-1} + p(H_+) \psi_{n,k}^{-1}}. \quad (22)$$

Up to this point we have achieved our desired conditional probability expressions in suitable formulations. The problem now remains is how to estimate $p(H_0)$, $p(H_+)$ and $p(H_-)$.

3. DETERMINATION OF $P(H_0)$, $P(H_+)$ AND $P(H_-)$

It is obvious that speech presence in all frequency index cannot be equally likely. And for noisy speech, noise is expected to be present in all frequency components irrespective of speech presence or absence. It follows that speech presence or absence does not depend on noise presence in anyway.

This fact encouraged us to determine experimentally the real probability of speech presence in the DCT domain using a very large number of speech utterances. The details of the procedure is described in the experiment section. The results presented in Fig. 1 (a) shows the speech presence probability in different DCT coefficients. We can clearly see that the probability of speech presence is very high at the lower mid range frequency and gradually decreases at the higher frequencies. Assuming that the curve can be expressed by a nonlinear equation, we conclude that the probability of speech absence in the k th DCT coefficient will be a function of k itself:

$$p(H_0) = 1 - v(k) \quad (23)$$

where $v(k)$ is the mathematical equation of the curve in Fig. 1(a) expressing speech presence probability for different values of k . In this paper, we have approximated this function using a curve fitting method.

For $p(H_+)$ and $p(H_-)$, since the DCT coefficients of speech and noise are both random in nature, it is equally likely that they will be constructive or destructive. Thus, after determining $p(H_0)$, if $p(H_+)$ and $p(H_-)$ are equal, it follows that,

$$p(H_+) = p(H_-) = 0.5 \times (1 - p(H_0))$$

4. THE IDEAL GAIN FILTER

It is apparent that the relative polarity of speech and noise spectral components is very important in determining an optimum gain. As for the conventional Wiener filter, even though it minimizes the mean squared error in a given sample space, its gain is always less than unity. Which means, it is always an attenuating filter even though some noisy speech components are actually reduced by the interference of noise components. The basic principle of a denoising filter with gain $W_{n,k}$ is expressed as

$$\hat{X}_{n,k} = W_{n,k} Y_{n,k} \quad (24)$$

where $\hat{X}_{n,k}$ denotes the enhanced speech component. Theoretically, the ideal optimum filter gain should be less than unity only when event H_+ has occurred. For event H_- , however there can be three cases. $|X_{n,k}| > |D_{n,k}|$ and $|X_{n,k}| < |D_{n,k}|$ and $|X_{n,k}| = |D_{n,k}|$. If $|X_{n,k}| = |D_{n,k}|$, the ideal filter gain is infinite, because here $Y_{n,k} = 0$.

When $|X_{n,k}| > |D_{n,k}|$, $Y_{n,k}$ is less than $X_{n,k}$ in magnitude, and of the same sign. Thus we need a gain that is greater than one in this region.

A very interesting case arises when $|X_{n,k}| < |D_{n,k}|$. Here in $Y_{n,k}$, noise is greater than the signal and thus it has reduced the signal so much that it is now of an opposite polarity of $X_{n,k}$. Which means, the ideal filter should have a gain that is negative and of appropriate magnitude, so that (24) is satisfied. Which means it will actually modify the phase of the noisy signal appropriately to reconstruct the clean signal. The variation of the ideal $W_{n,k}$ with the relative magnitude of $|X_{n,k}|$ and $|D_{n,k}|$ is shown in Fig. 1 (b). As expected, the gain is discontinuous at $|X_{n,k}| = |D_{n,k}|$, and at lower values of $|X_{n,k}|$, it is negative. It approaches 1 for very high SNRs and reduces to zero at very low SNRs. If we are to use different gains for constructive and destructive interference between speech and noise, we must derive a gain that follows approximately the ideal gain curve, except for the discontinuity.

5. THE ATTENUATING AND THE AMPLIFYING GAIN

Though the dual gain Wiener filter presented in [10] is very unique in nature and the concept of using different gains for constructive and destructive interference is interesting, the gain expressions fail to meet the criteria for the ideal filter mentioned in the previous section. The destructive gain proposed in [10] cannot guarantee a value that is greater than one in the range where signal is greater than noise and destructive interference has occurred. Also it fails to give a negative gain in the region when noise is stronger than the signal. Therefore, the dual gain algorithm proposed in [10] requires further investigation. Here we attempt to derive an

expression for an exact gain that closely follows the ideal filter characteristics.

It is obvious from (24) that we want the gain $W_{n,k}$ to be,

$$W_{n,k} = \frac{X_{n,k}}{Y_{n,k}}.$$

Substituting $Y_{n,k}$ from (2),

$$W_{n,k} = \frac{X_{n,k}}{X_{n,k} + D_{n,k}}.$$

Multiplying the numerator and denominator by $X_{n,k}$,

$$W_{n,k} = \frac{X_{n,k}^2}{X_{n,k}^2 + X_{n,k}D_{n,k}}$$

This equation can be written in the following form for the constructive and destructive cases,

$$W_{n,k} = \frac{X_{n,k}^2}{X_{n,k}^2 \pm |X_{n,k}||D_{n,k}|} \quad (25)$$

where the $+$ sign is for constructive and $-$ sign for destructive noise interference. It is clear that the values of (25) cannot be obtained directly, which is the reason why the earlier approaches were always aimed to minimize the mean squared error. Since we cannot determine the instantaneous terms $|X_{n,k}||D_{n,k}|$ and $X_{n,k}^2$, we replace them by their expected values:

$$W_{n,k} = \frac{E\{X_{n,k}^2\}}{E\{X_{n,k}^2\} \pm E\{|X_{n,k}||D_{n,k}|\}}$$

Dividing the numerator and denominator by $E\{D_{n,k}^2\}$ and using the relations (11) and (12) we have,

$$W_{n,k} = \frac{\xi_{n,k}}{\xi_{n,k} \pm \frac{2}{\pi} \sqrt{\xi_{n,k}}} \quad (26)$$

where $+$ and $-$ signs will be used for constructive and destructive interferences, respectively. We denote these gains as the attenuating and the amplifying gain.

It is easily notable that the gain in (26) is always less than unity for event H_+ , for all $\xi_{n,k} > 0$. However, for the destructive case, this gain has a discontinuity at $\xi_{n,k} = \frac{4}{\pi^2}$, which is similar to our ideal filter gain. This discontinuity is obvious but impractical, since we cannot predict for which component the noise has exactly canceled the signal.

To handle this discontinuity, we modify (26) for the amplifying gain as

$$W_{n,k}^- = \frac{\xi_{n,k}}{\xi_{n,k} - \frac{2}{\pi} \sqrt{\xi_{n,k}} + \lambda}$$

where λ is a positive constant. Using this modification, we have actually discarded the negative property of the gain. This is done because it is safer to reduce the component in magnitude rather than to use a high gain with a negative polarity for the destructive case. That would introduce a very prominent distortion in case of a wrong decision of polarity.

Now that we have distinct gain expressions for the constructive and destructive events, we now use our probabilities to utilize our gains. We now summarize our proposed gains in three different events, which are in fact the only possible events, as follows.

$$\begin{aligned} W_{n,k}^0 &= 0 && \text{when speech is in state } H_0 \\ W_{n,k}^+ &= \frac{\xi_{n,k}}{\xi_{n,k} + \frac{2}{\pi} \sqrt{\xi_{n,k}}} && \text{when speech is in state } H_+ \\ W_{n,k}^- &= \frac{\xi_{n,k}}{\xi_{n,k} - \frac{2}{\pi} \sqrt{\xi_{n,k}} + \lambda} && \text{when speech is in state } H_- \end{aligned} \quad (27)$$

Thus the probabilistic filter gain will be the expected value of the gains given by

$$\bar{W}_{n,k} = W_{n,k}^0 p(H_0|Y_{n,k}) + W_{n,k}^+ p(H_+|Y_{n,k}) + W_{n,k}^- p(H_-|Y_{n,k}). \quad (28)$$

The concept of expected gain is necessary because we do not have certain knowledge of the events H_+ and H_- . An optimum expression is still to be developed for handling these two cases including the special case of sign reversal of a noisy component.

6. EXPERIMENTS AND DISCUSSION

We have performed two different experiments that were required for the implementation of our algorithm. First, the experimental speech presence/absence probability is determined and approximated. Second, our proposed probabilistic gain is tested and a comparative performance analysis is presented.

6.1 Determination of speech absence probability

In this experiment 1000 utterances were used from the TIMIT database, having almost equal number of male and female speakers. The sampling frequency was 8 kHz. A frame size of 512 samples (64 ms) was taken with 50% overlap and the 512 point DCT was calculated in each frame. In total, 92332 speech frames were processed. The event of speech absence H_0 is defined as,

$$H_0 : |X_k| < \frac{1}{10} \sqrt{E\{X_k^2\}}$$

where X_k is the k th DCT index of a clean speech frame. Stated otherwise, speech components lower than one-tenth of the standard deviation of that frame, has been considered to be negligible. Thus, the probability of speech presence given the DCT index is k ,

$$p(H_0) = \frac{n(H_0|k)}{N}$$

where $n(H_0|k)$ is the number of occurrence of event H_0 for the k th DCT index and $N = 92332$. Fig. 1 (a) is the plot of $p(H_0)$ vs. k . A curve was fitted to this envelope using a 15th order polynomial and thus the function $v(k)$ in (23) was approximated.

6.2 Implementation of the expected gain

The probabilistic filter gain using the expected value of gains in different conditions presented in this paper, has been tested using 5 male and 5 female utterances randomly taken from the TIMIT database. All of the utterances were corrupted

with white noise of SNR ranging from -10dB to +25dB, taken from the 'NOISEX' database. The sampling frequency was 8 KHz. A frame size of 32 ms (256 samples) was used for framing and the overlap-add method with 50% overlap was used for signal decomposition.

The value of the a priori SNR was calculated using the variable averaging parameter α proposed in [4]. We have used $\lambda = 1$ in (27) in the destructive case. The results obtained from the conventional Wiener filter, parametric method (PARA) [11] and the MPE algorithm [7] which incorporates speech presence and absence probability in its gain functions are also presented for comparison. The averaged results of the 10 utterances are plotted in Figs. 2 and 3.

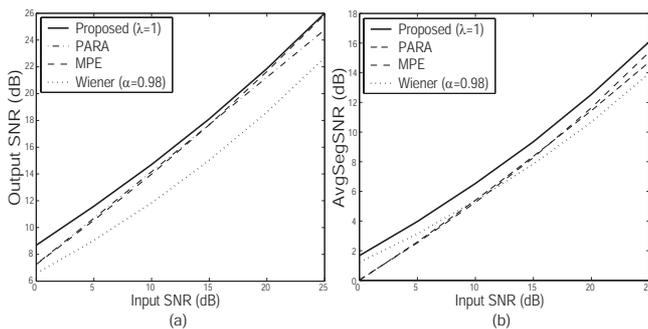


Figure 2: (a) Variation of overall SNR with input SNR, (b) Variation of average segmental SNR with input SNR.

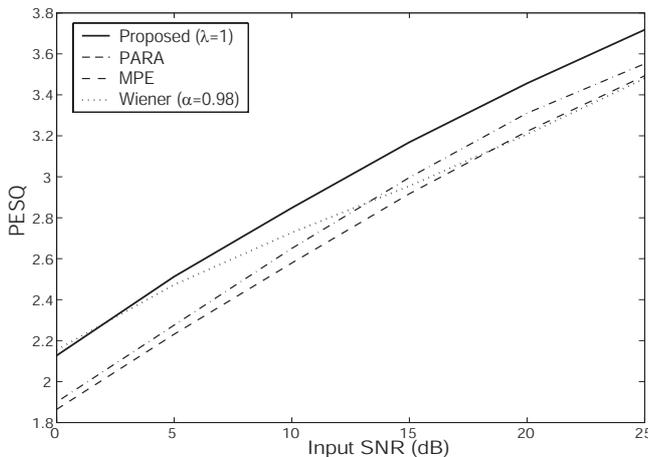


Figure 3: Variation of PESQ with input SNR.

From Fig. 2 (a), we can see that the proposed method using $\lambda = 1$ shows significant SNR improvement compared to parametric method (PARA) [11], modified power estimation (MPE), and conventional Wiener filter.

In Fig. 2 (b), we clearly observe the superiority of the proposed method in terms of the average segmental SNR (AvgSegSNR). This quality index, which is highly correlated to human listening, is higher than other methods for almost the entire input SNR range. The better listening quality is also certified by the PESQ (Perceptual Evaluation of Speech Quality) scores [9] shown in Fig. 3. Nevertheless, it is the listening quality we are most concerned about. It is observed that using the optimum averaging parameter pro-

posed in [4], even though the overall SNR and other quality indices are significantly improved, the listening quality is degraded. Still, the variable averaging parameter is the key to better estimation of $\xi_{n,k}$, which is an important parameter that determines the accuracy of the probabilities we have determined. The proposed method, uses this optimum α . This is the basic advantage of our probabilistic soft-thresholding idea. Better estimation of the a priori SNR and an optimum gain in constructive and destructive cases will definitely take this probabilistic method a leap forward.

7. CONCLUSION

A novel probabilistic filter gain considering constructive and destructive interference of noise has been proposed for speech enhancement. Simulation results presented have demonstrated superiority of the this filter over some popular speech enhancement algorithms.

REFERENCES

- [1] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. of IEEE*, vol. 67, no. 12, pp. 1586-1604, 1979.
- [2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, pp. 113-120, 1979.
- [3] S. Salahuddin, S. Z. Al Islam, M. K. Hasan, M. R. Khan, "Soft thresholding for DCT speech enhancement," *Electron. lett.*, vol. 38, no.24, pp. 1605-1607, 2002.
- [4] M. K. Hasan, S. Salahuddin and M. R. Khan, "A modified a priori SNR for speech enhancement using spectral subtraction rules," *IEEE Signal Processing Letters*, vol. 11, no. 4, pp. 450-453, April 2004.
- [5] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, Vol. 41, pp. 613-627, May 1995.
- [6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Speech Audio Processing*, vol. ASSP-32, pp. 1109-1121, 1984.
- [7] P. Scalart and J. Vieira-Filho, "Speech enhancement based on a priori signal to noise estimation," *Proc. ICASSP*, pp. 629-632, 1996.
- [8] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Ac. Speech and Signal Processing*, vol. ASSP-28, No. 2, pp. 137-145, April 1980.
- [9] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assesment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 2, Salt Lake City, UT, 2001, pp. 749-752.
- [10] I. Y. Soon, and S. N. Koh, "Low distortion speech enhancement," *IEE Proc.-Vis. Image Signal Process.*, vol. 147, pp.247-253, 2000.
- [11] B. L. Sim, Y. C. Tong, J. S. Chang and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 328-337, 1998.