Progress In Electromagnetics Research M, Vol. 4, 167–184, 2008

SPEECH ENHANCEMENT USING AN ADAPTIVE WIENER FILTERING APPROACH

M. A. Abd El-Fattah, M. I. Dessouky, S. M. Diab and F. E. Abd El-samie

Department of Electronics and Electrical communications
Faculty of Electronic Engineering
Menoufia University
Menouf, Egypt

Abstract—This paper proposes the application of the Wiener filter in an adaptive manner in speech enhancement. The proposed adaptive Wiener filter depends on the adaptation of the filter transfer function from sample to sample based on the speech signal statistics (mean and variance). The adaptive Wiener filter is implemented in time domain rather than in frequency domain to accommodate for the varying nature of the speech signal. The proposed method is compared to the traditional Wiener filter and the spectral subtraction methods and the results reveal its superiority.

1. INTRODUCTION

Speech enhancement is one of the most important topics in speech signal processing. Several techniques have been proposed for this purpose like the spectral subtraction approach, the signal subspace approach, adaptive noise canceling and the iterative Wiener filter [1–5]. The performances of these techniques depend on the quality and intelligibility of the processed speech signal. The improvement in the speech signal-to-noise ratio (SNR) is the target of most techniques.

Spectral subtraction is the earliest method for enhancing speech degraded by additive noise [1]. This technique estimates the spectrum of the clean (noise-free) signal by the subtraction of the estimated noise magnitude spectrum from the noisy signal magnitude spectrum while keeping the phase spectrum of the noisy signal. The drawback of this technique is the residual noise.

Another technique is the signal subspace approach [3]. It is used for enhancing speech signals degraded by uncorrelated additive noise or colored noise [6, 7]. The idea of this algorithm is based on the fact that the vector space of the noisy signal can be decomposed into a signal plus noise subspace and an orthogonal noise subspace. Processing is performed on the vectors in the signal plus noise subspace only, while the noise subspace is removed first. Decomposition of the vector space of the noisy signal is performed by applying the singular value decomposition or the Karhunen-Loeve transform (KLT) on the speech signal[8]. Mi et al. have proposed the signal/noise KLT based approach for the removal of colored noise [9]. The idea of this approach is that noisy speech frames are classified into speech-dominated frames and noise-dominated frames. In the speech-dominated frames, the signal KLT matrix is used and in the noise-dominated frames, the noise KLT matrix is used.

In this paper, we present a new technique to improve the SNR in the enhanced speech signal by using an adaptive implementation of the Wiener filter. This implementation is performed in the time domain to accommodate for the varying nature of the signal.

The paper is organized as follows. In Section 2, a review of the spectral subtraction technique is presented. In Section 3, the traditional Wiener filter in frequency domain is revisited. Section 4 proposes the adaptive Wiener filter approach for speech enhancement. In Section 5, a comparative study between the proposed adaptive Wiener filter, the Wiener filter in frequency domain and the spectral subtraction approach is presented.

2. SPECTRAL SUBTRACTION

The spectral subtraction approach can be categorized as a non-parametric approach, which simply needs an estimate of the noise spectrum. It is assumed that there is an estimate of the noise spectrum which is obtained during periods of speaker silence. Let $x(n)$ be a noisy speech signal:

$$x(n) = s(n) + v(n) \tag{1}$$

where $s(n)$ is the clean signal, and $v(n)$ is the white gaussian noise. In this case, the noise and the clean signal can be assumed uncorrelated. So, the spectral subtraction approach can be used to estimate the short term magnitude spectrum of the clean signal $|S(\omega)|$ by the subtraction of the estimated noise magnitude spectrum $\left|\hat{V}(\omega)\right|$ from the noisy signal magnitude spectrum $|X(\omega)|$. It is sufficient to use the noisy signal phase spectrum as an estimate of the clean speech phase spectrum as follows [10]:

$$\hat{S}(\omega) = (|X(\omega)| - \left|\hat{N}(\omega)\right|) \exp(j\angle X(\omega)) \tag{2}$$

The estimated time-domain speech signal is obtained as the inverse Fourier transform of $\hat{S}(\omega)$.

Another way to recover the clean signal $s(n)$ from the noisy signal $x(n)$ using the spectral subtraction approach is performed by assuming that there is an estimate of the power spectrum of the noise $P_v(\omega)$, which is obtained by averaging over multiple frames of a known noise segment. An estimate of the short-time squared magnitude spectrum of the clean signal using this method can be obtained as follows [8]:

$$\left|\hat{S}(\omega)\right|^2 = \begin{cases} |X(\omega)|^2 - \hat{P}_v(\omega), & \text{if } |X(\omega)|^2 - \hat{P}_v(\omega) \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

It is possible to combine this magnitude spectrum estimate with the phase of the noisy signal and then get the Short Time Fourier Transform (STFT) estimate of the clean signal as follows::

$$\hat{S}(\omega) = \left|\hat{S}(\omega)\right| e^{j\angle X(\omega)} \tag{4}$$

A noise-free signal estimate can then be obtained with the inverse Fourier transform. This noise reduction method is a specific case of the general technique given by Weiss et al. and extended by Berouti et al. [2, 12].

The spectral subtraction approach can be viewed as a filtering operation where high SNR regions of the measured spectrum are attenuated less than low SNR regions. This formulation can be given in terms of the SNR defined as:

$$SNR = \frac{|X(\omega)|^2}{\hat{P}_v(\omega)} \tag{5}$$

Thus, Eq. (3) can be rewritten as:

$$\left|\hat{S}(\omega)\right|^2 = |X(\omega)|^2 - \hat{P}_v(\omega) \approx |X(\omega)|^2 \left[1 + \frac{1}{SNR}\right]^{-1} \tag{6}$$

An important property of noise suppression using the spectral subtraction approach is that the attenuation characteristics change with the length of the analysis window. A common problem of using the spectral subtraction approach is the musicality that results from the rapid coming and going of waves over successive frames [13].

3. WIENER FILTER IN FREQUENCY DOMAIN

The Wiener filter is a popular technique that has been used in many signal enhancement methods. The basic principle of the Wiener filter is to obtain an estimate of the clean signal from that corrupted by additive noise. This estimate is obtained by minimizing the Mean Square Error (MSE) between the desired signal $s(n)$ and the estimated signal $\hat{s}(n)$. The frequency domain solution to this optimization problem gives the following filter transfer function [13]:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_v(\omega)} \tag{7}$$

where $P_s(\omega)$ and $P_v(\omega)$ are the power spectral densities of the clean and the noise signals, respectively. This formula can be derived considering the signal s and the noise v as uncorrelated and stationary signals. The SNR is defined by [13]:

$$SNR = \frac{P_s(\omega)}{\hat{P}_v(\omega)} \tag{8}$$

This definition can be incorporated to the Wiener filter equation as follows:

$$H(\omega) = \left[1 + \frac{1}{SNR}\right]^{-1} \tag{9}$$

The drawback of the Wiener filter is the fixed frequency response at all frequencies and the requirement to estimate the power spectral density of the clean signal and noise prior to filtering.

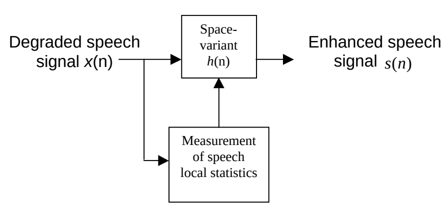

Figure 1. Adaptive Wiener filtering approach for speech enhancement.

4. THE PROPOSED ADAPTIVE WIENER FILTER

This section presents an adaptive implementation of the Wiener filter which benefits from the varying local statistics of the speech signal. A block diagram of the proposed approach is illustrated in Fig. 1. In this approach, the estimated speech signal mean m_x and variance σ_x^2 are exploited.

It is assumed that the additive noise $v(n)$ is of zero mean and has a white nature with variance of σ_v^2. Thus, the power spectrum $P_v(\omega)$ can be approximated by:

$$P_v(\omega) = \sigma_v^2 \tag{10}$$

Consider a small segment of the speech signal, in which the signal $x(n)$ is assumed to be stationary, The signal $x(n)$ can be modeled by:

$$x(n) = m_x + \sigma_x w(n) \tag{11}$$

where m_x and σ_x are the local mean and standard deviation of $x(n)$. $w(n)$ is a unit variance noise.

Within this small segment of speech, the Wiener filter transfer function can be approximated by:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_v(\omega)} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2} \tag{12}$$

From Eq. (12), because $H(\omega)$ is constant over this small segment of speech, the impulse response of the Wiener filter can be obtained by:

$$h(n) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}\delta(n) \tag{13}$$

From Eq. (13), the enhanced speech signal $\hat{s}(n)$ in this local segment can be expressed as:

$$\hat{s}(n) = m_x + (x(n) - m_x) * \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}\delta(n) = m_x + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}(x(n) - m_x) \tag{14}$$

If m_x and σ_s are updated at each sample, we can say:

$$\hat{s}(n) = m_x(n) + \frac{\sigma_s^2(n)}{\sigma_s^2(n) + \sigma_v^2}(x(n) - m_x(n)) \tag{15}$$

In Eq. (15), the local mean $m_x(n)$ and $(x(n) - m_x(n))$ are modified separately from segment to segment and then the results are combined. If σ_s^2 is much larger than σ_v^2 the output signal $\hat{s}(n)$ will be primarily due to $x(n)$ and the input signal $x(n)$ is not attenuated. If σ_s^2 is smaller than σ_v^2, the filtering effect is performed.

Notice that m_x is identical to m_s when m_v is zero. So, we can estimate $m_x(n)$ in Eq. (15) from $x(n)$ by:

$$\hat{m}_s(n) = \hat{m}_x(n) = \frac{1}{(2M+1)} \sum_{k=n-M}^{n+M} x(k) \tag{16}$$

where $(2M+1)$ is the number of samples in the short segment used in the estimation.

To measure the local statistics of the speech signal, we need to estimate the signal variance σ_s^2. Since $\sigma_x^2 = \sigma_s^2 + \sigma_v^2$, then $\sigma_s^2(n)$ may be estimated from $x(n)$ as follows:

$$\hat{\sigma}_s^2(n) = \begin{cases} \hat{\sigma}_x^2(n) - \hat{\sigma}_v^2, & \text{if } \hat{\sigma}_x^2(n) > \hat{\sigma}_v^2 \\ 0, & \text{otherwise} \end{cases} \tag{17a}$$

where

$$\hat{\sigma}_x^2(n) = \frac{1}{(2M+1)} \sum_{k=n-M}^{n+M} (x(k) - \hat{m}_x(n))^2 \tag{17b}$$

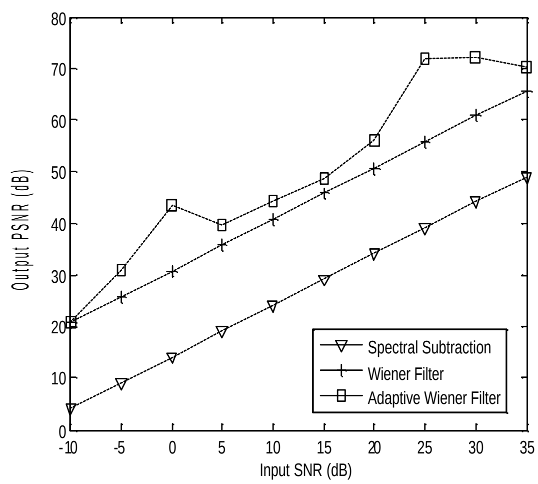

Figure 2. PSNR results for white noise case at -10 dB to $+35$ dB SNR levels for Handle signal.

By this method, we guarantee the adaptation of the filter transfer function from sample to sample based on the local statistics of the speech signal.

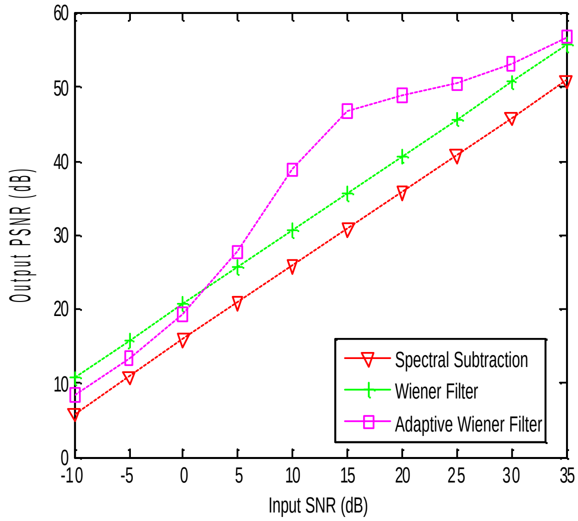

Figure 3. PSNR results for white noise case at -10 dB to $+35$ dB SNR levels for Laughter signal.

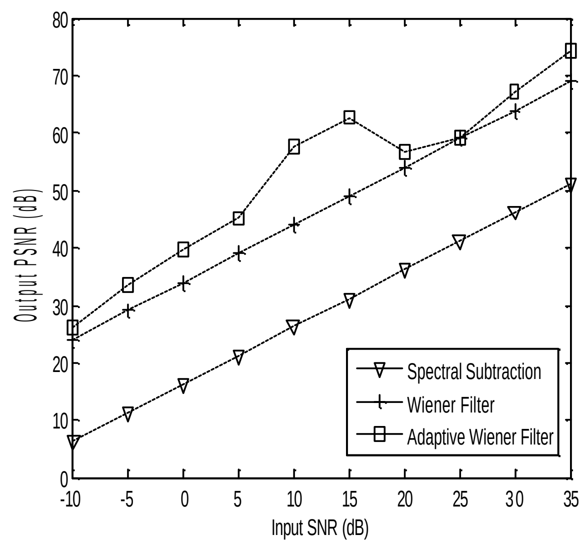

Figure 4. PSNR results for white noise case at -10 dB to $+35$ dB SNR levels for Gong signal.

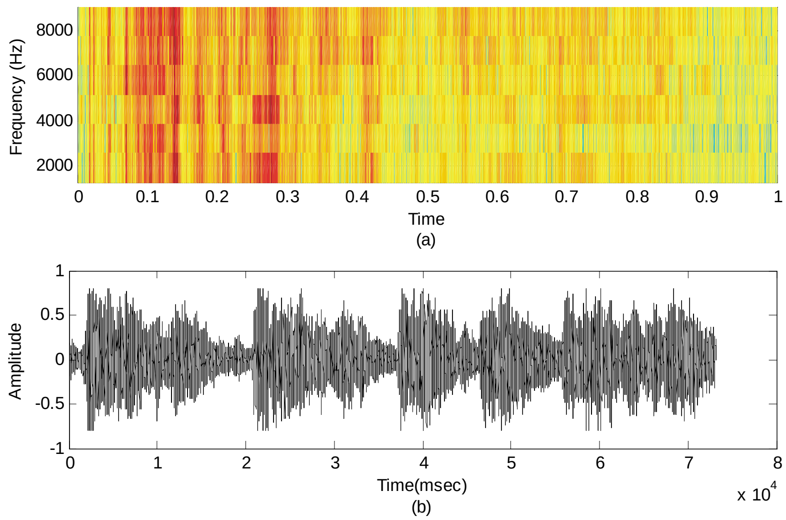

Figure 5. The clean signal. (a) The spectrogram. (b) The time signal.

5. EXPERIMENTAL RESULTS

For evaluation purposes, we use different speech signals; the handle, the laughter and the gong signals. The noisy signal is obtained by adding white gaussian noise to the speech signal with different SNR values. The output peak signal to noise ratio (PSNR) results for the wiener filter, the spectral subtraction approach and the proposed adaptive wiener filter are shown in Figs. 2, 3 and 4. From these figures, it is clear that the proposed adaptive wiener filter has the best performance for different SNR values. The adaptive wiener filter approach gives about 3–5 dB improvement at different values of SNR.

Some experiments are carried out on the handle signal shown in Fig. 5 with SNR values from of 5 to 20 dB to test all the speech enhancement algorithms mentioned in this paper. In all of these experiments, the spectrogram is used with the time signal to clarify the time and frequency contents of the signal. The noisy handle signal with SNR values from 5 to 20 dB in 5 dB steps are shown in Fig. 6.

The results of the spectral subtraction approach, the Wiener filter approach and the adaptive Wiener filter approach are shown in Figs. 7, 8 and 9, respectively. The PSNR results of all these speech enhancement approaches are tabulated in Table 1. From the figures and the table, it is clear that the best performance is that of the proposed adaptive Wiener filter because it considers the variation of the local statistics of the noisy speech signal.

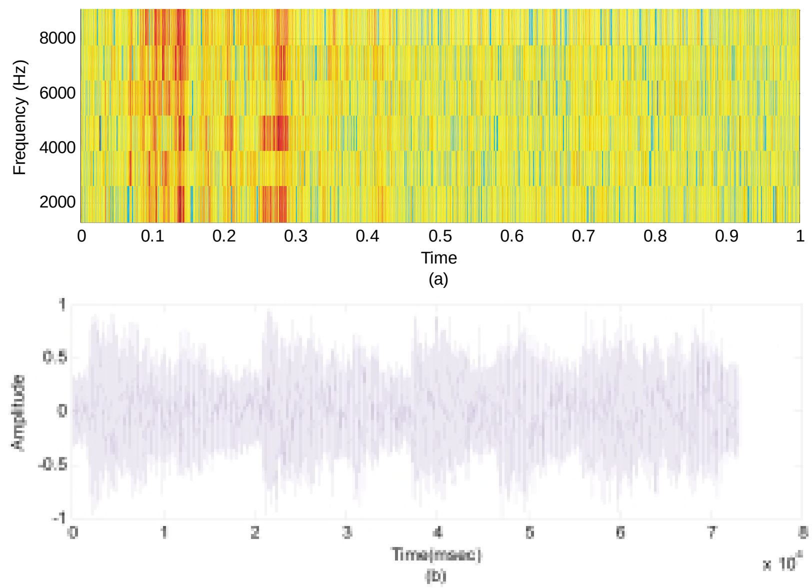

Figure 6a. The noisy signal at SNR = 5 dB (a) The spectrogram, (b) The time signal.

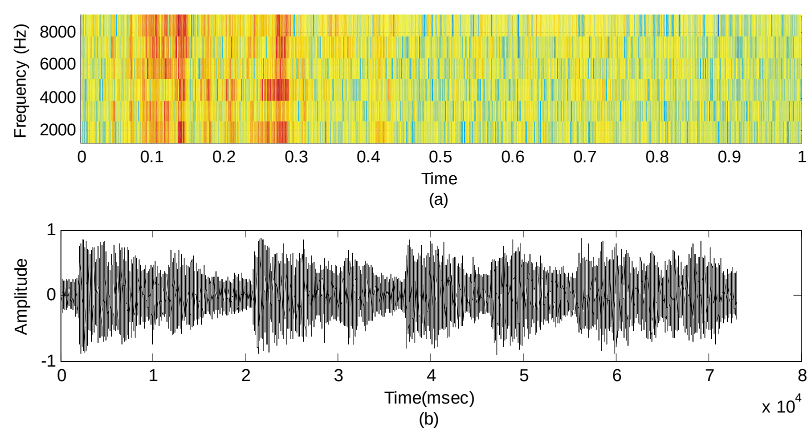

Figure 6b. The noisy signal at SNR = 10 dB (a) the spectrogram, (b) The time signal.

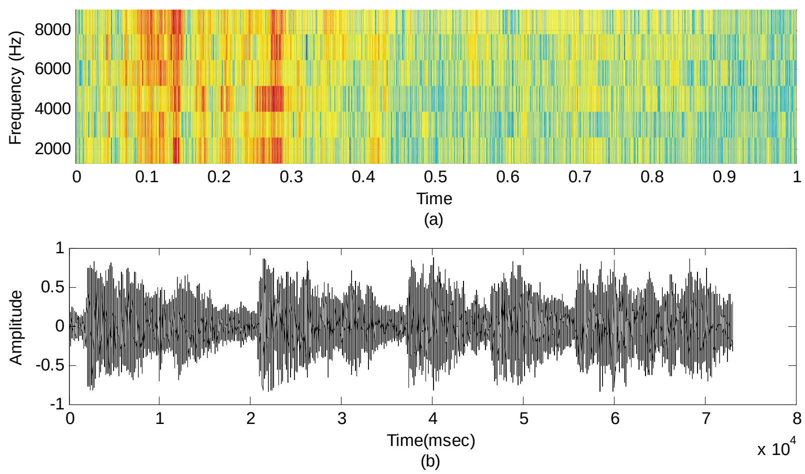

Figure 6c. The noisy signal at SNR = 15 dB (a) The spectrogram, (b) The time signal.

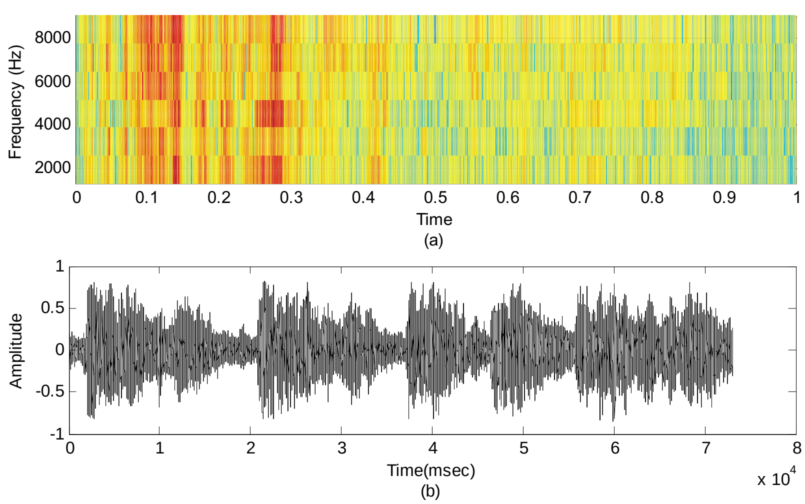

Figure 6d. The noisy signal at SNR = 20 dB (a) the spectrogram, (b) The time signal.

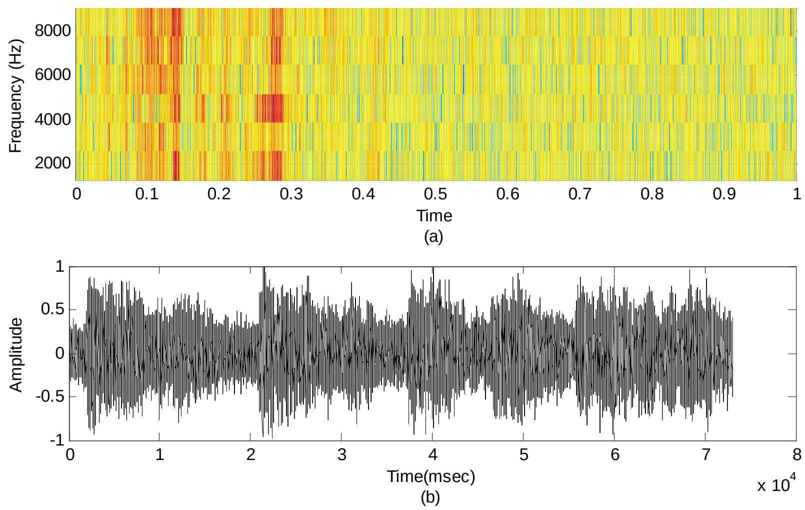

Figure 7a. The spectral subtraction technique at SNR = 5 dB (a) the spectrogram, (b) The time signal.

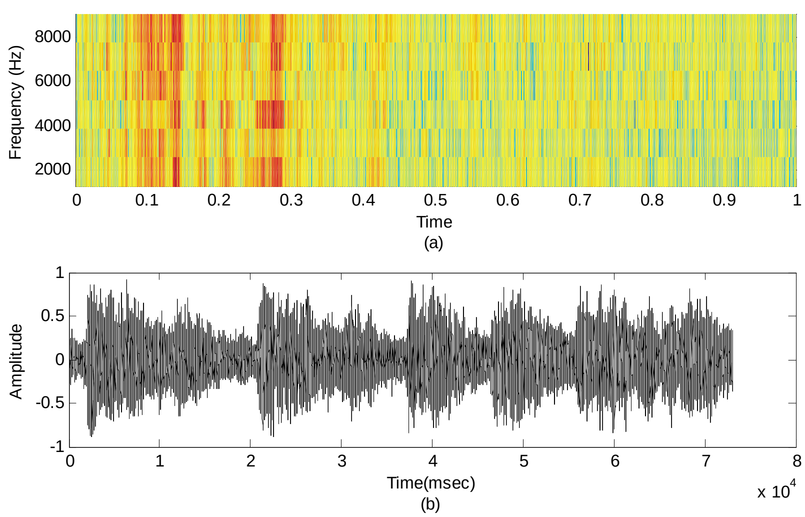

Figure 7b. The spectral subtraction at SNR = 10 dB (a) The spectrogram, (b) The time signal.

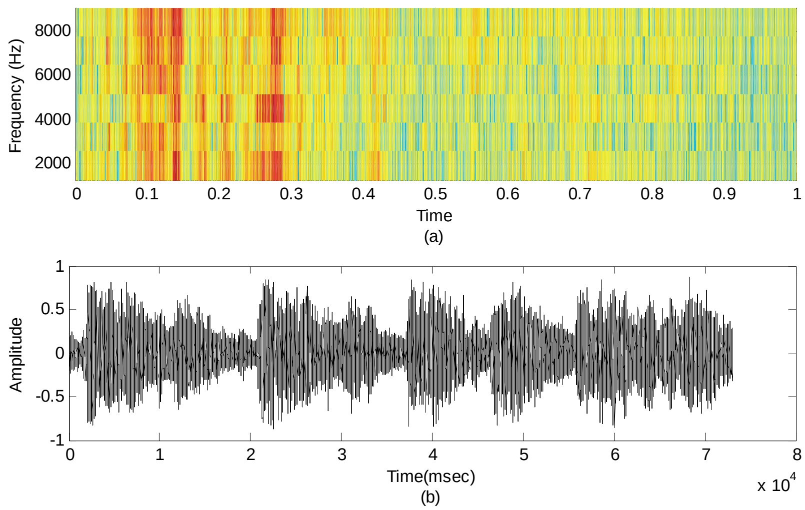

Figure 7c. The spectral subtraction technique at SNR = 15 dB (a) The spectrogram, (b) The time signal.

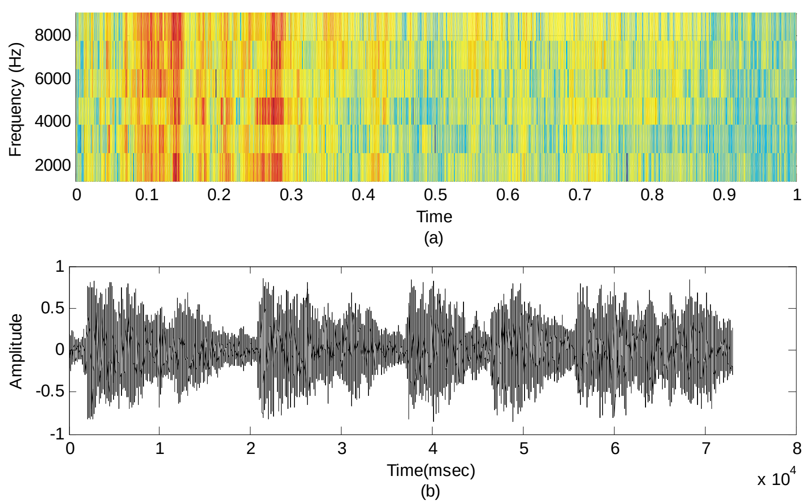

Figure 7d. The spectral subtraction technique at SNR = 20 dB (a) the spectrogram, (b) The time signal.

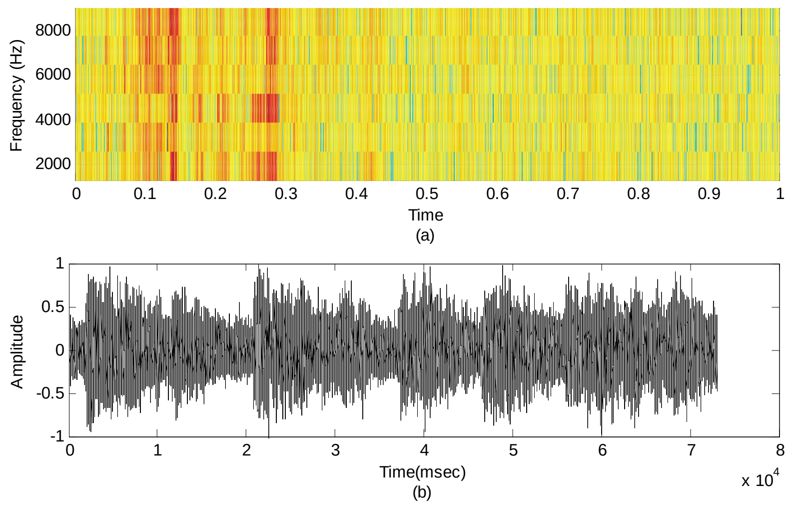

Figure 8a. The Wiener filter technique At SNR = 5 dB (a) The spectrogram, (b) The time signal.

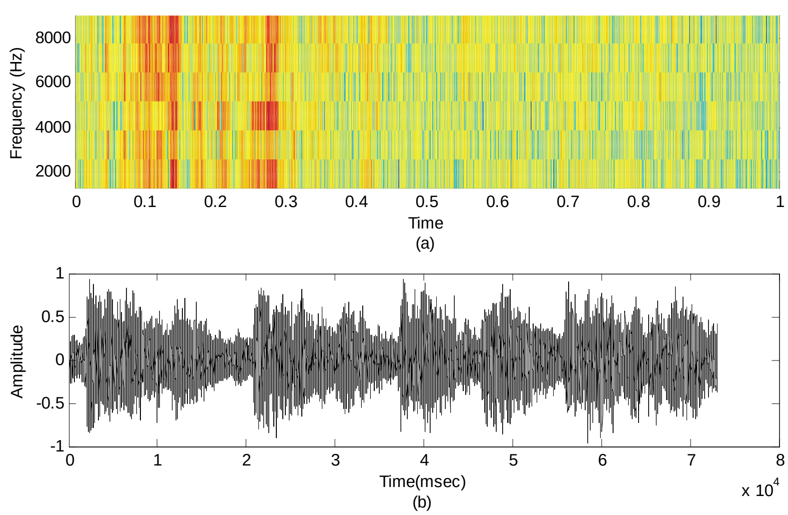

Figure 8b. The Wiener filter technique At SNR = 10 dB (a) The spectrogram, (b) The time signal.

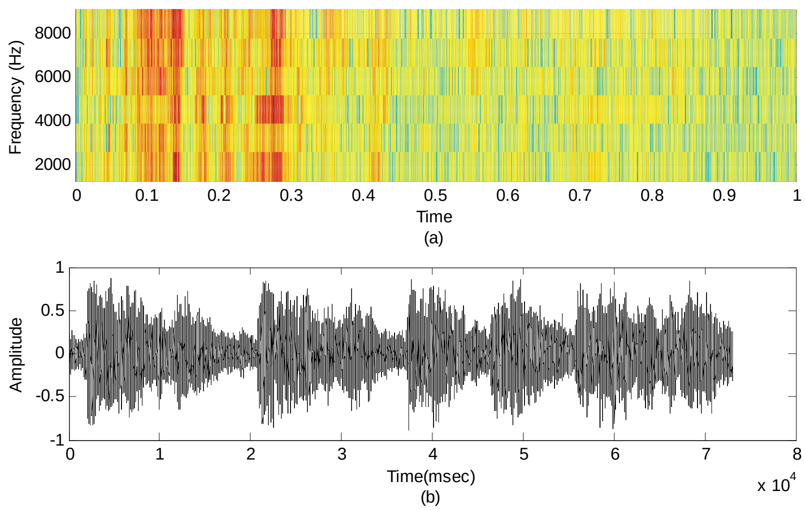

Figure 8c. The Wiener filter technique At SNR = 15 dB (a) The spectrogram, (b) The time signal.

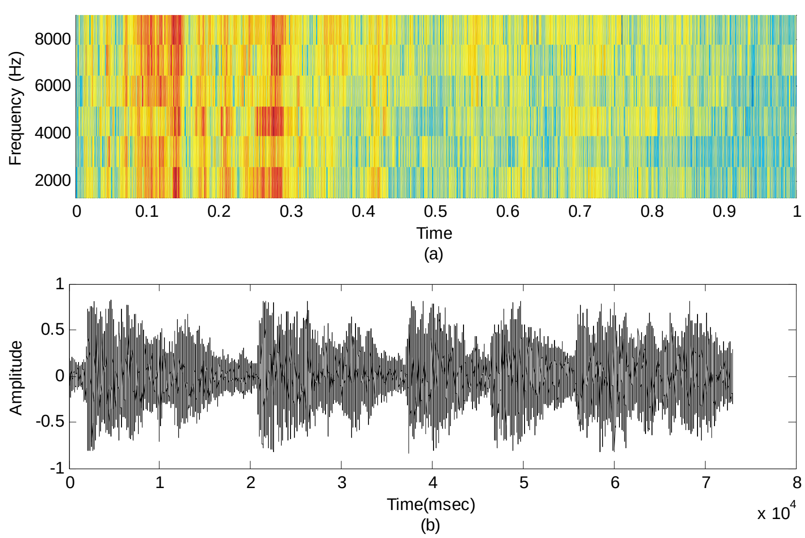

Figure 8d. The Wiener filter technique At SNR = 20 dB (a) The spectrogram, (b) The time signal.

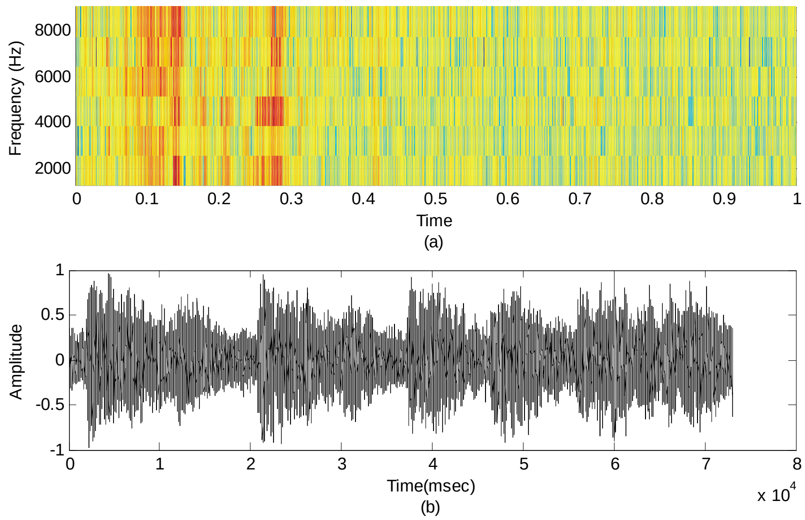

Figure 9a. The adaptive Wiener filter technique At SNR $= 5\,\mathrm{dB}$ (a) The spectrogram, (b) The time signal.

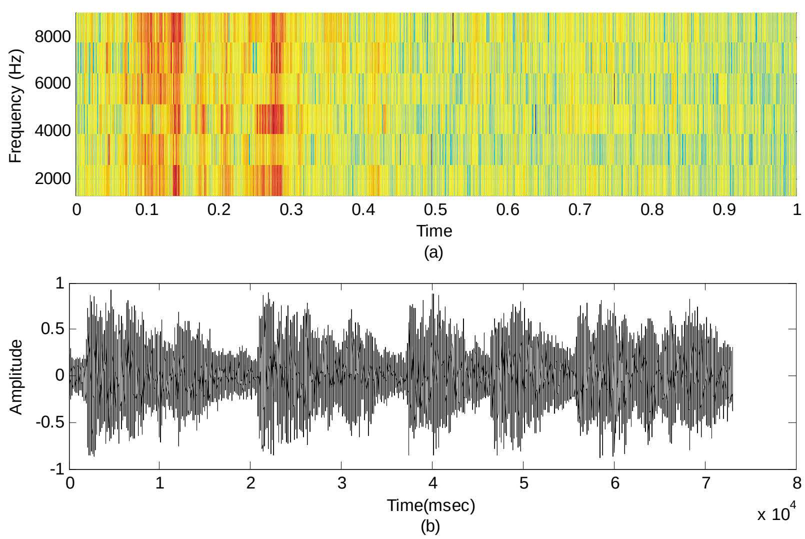

Figure 9b. The adaptive Wiener filter technique At SNR $= 10\,\mathrm{dB}$ (a) The spectrogram, (b) The time signal.

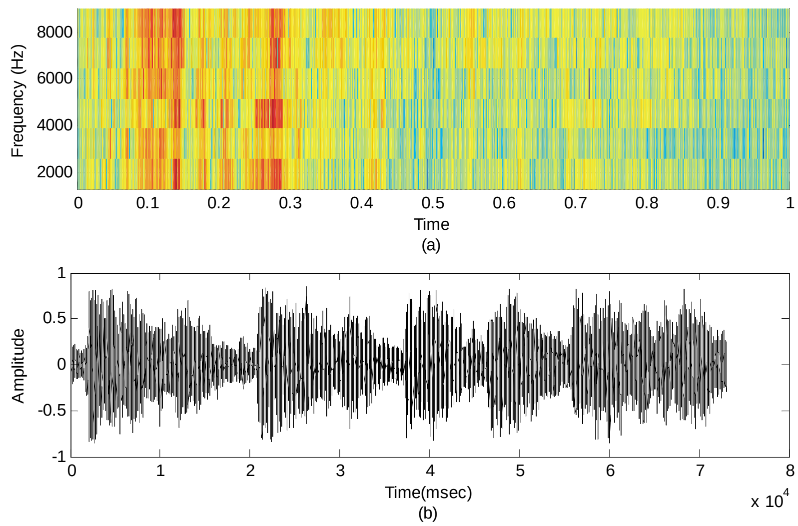

Figure 9c. The adaptive Wiener filter technique At SNR = 15 dB (a) The spectrogram, (b) The time signal.

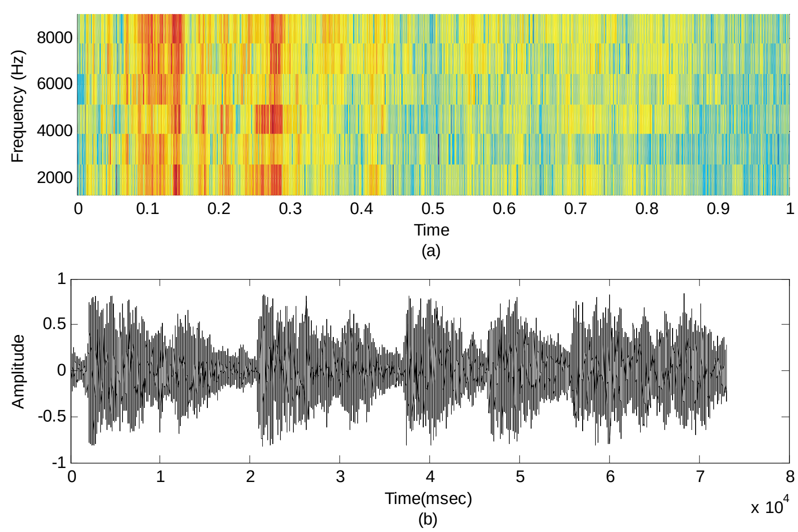

Figure 9d. The adaptive Wiener filter technique At SNR = 20 dB (a) The spectrogram, (b) The time signal.

Table 1. PSNR results in dB for the speech enhancement approaches applied to the handle signal at different SNR values.

SNR	Noisy Signal	Spectral subtraction	Frequency domain Wiener filter	Adaptive Wiener Filter
5 dB	19.1383	19.1407	22.7568	28.6086
10 dB	24.1217	24.1228	28.9876	32.9784
15 dB	29.1543	29.1547	32.0856	37.3434
20 dB	34.1144	34.1146	37.5216	40.1843

6. CONCLUSION

An adaptive Wiener filter approach for speech enhancement has been proposed in this paper. A mathematical derivation of the filter transfer function has been introduced. This filter is applied by the adaptation of its transfer function from sample to sample based on the speech signal statistics (mean and variance). The experimental results indicate that the proposed filter provides the best PSNR improvement among the spectral subtraction approach and the traditional Wiener filter approach which is implemented in the frequency domain.

REFERENCES

1. Boll, S. F., "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-27, 113–120, 1979.
2. Berouti, M., R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proc. IEEE Int. Conf. Acoust., Speech Signal Processing*, 208–211, 1979.
3. Ephriam, Y. and H. L. van Trees, "A signal subspace approach for speech enhancement," *Proc. International Conference on Acoustic, Speech and Signal Processing*, Vol. 2, 355–358, Detroit, MI, USA., May 1993.
4. Haykin, S., *Adaptive Filter Theory*, Prentice-Hall, ISBN 0-13-322760-X, 1996.
5. Lim, J. S. and A. V. Oppenheim, "All-pole modelling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-26, June 1978.
6. Ephraim, Y. and H. L. van Trees, "A spectrally-based signal

subspace approach for speech enhancement," *IEEE ICASSP*, 804–807, 1995.

7. Hu, Y. and P. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," *Proc. International Conference on Acoustics, Speech and Signal Processing*, Vol. 1, 573–576, Orlando, FL, USA, May 2002.
8. Rezayee, A. and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, Vol. 9, 87–95, Feb. 2001.
9. Mittal, U. and N. Phamdo, "Signal/noise KLT based approach for enhancing speech degraded by colored noise," *IEEE Trans. Speech Audio Processing*, Vol. 8, No. 2, 159–167, 2000.
10. Deller, J. R., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Prentice-Hall, 1997.
11. Boll, S. F., "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-29, No. 2, 113–120, April 1979.
12. Weiss, M. R., E. Aschkenasy, and T. W. Parsons, "Processing speech signal to attenuate interference," *Proc. IEEE Symp. Speech Recognition*, 292–293, April 1974.
13. Lim, J. S. and A. V. Oppenheim, "Enhancement and band width compression of noisy speech," *Proc. of the IEEE*, Vol. 67, No. 12, 1586–1604, Dec. 1979.