ADAPTIVE WIENER FILTERING APPROACH FOR SPEECH ENHANCEMENT

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
E-mails: * maro_zizo2010@yahoo.com , # fathi_sayed@yahoo.com

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 spectral subtraction methods and the results reveal its superiority.

Keywords: Speech Enhancement, Spectral Subtraction, Adaptive Wiener Filter

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 quality and intelligibility of the processed speech signal. The improvement of 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 a signal subspace approach [3]. It is used for enhancing a speech signal 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 an eigenvalue or singular value decomposition or by applying the Karhunen-Loeve transform (KLT)[8]. Mi. et. al. have proposed the signal / noise KLT based approach for colored noise removal[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 signal-to-noise ratio in the enhanced speech signal by using an adaptive implementation of the Wiener filter. This implementation is performed in time domain to accommodate for the varying nature of the signal.

The paper is organized as follows: in section II, a review of the spectral subtraction technique is presented. In section III, the traditional Wiener filter in frequency domain is revisited. Section IV, proposes the adaptive Wiener filtering approach for speech enhancement. In section V, 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

Spectral subtraction can be categorized as a non-parametric approach, which simply needs an estimate of the noise spectrum. It is assume that there is an estimate of the noise spectrum that is typically estimated 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 (the noise-free) signal, and $v(n)$ is the white gaussian noise. Assume that the noise and the clean signals are uncorrelated. By applying the spectral subtraction approach that estimates the short term magnitude spectrum of the noise-free signal $|S(\omega)|$ by subtraction of the estimated noise magnitude spectrum $|\hat{V}(\omega)|$ 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,[10]:

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

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

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

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

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

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

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, equation (3) can be rewritten as:

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

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

3 WIENER FILTER IN FREQUNCY 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 a clean signal from that corrupted by additive noise. It is required estimate an optimal filter for the noisy input speech 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 is given by[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 signal v as uncorrelated and stationary signals. The signal-to-noise ratio is defined by[13]:

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

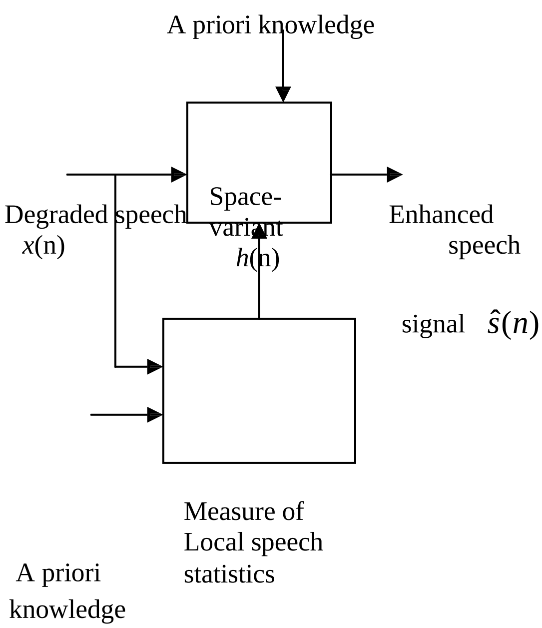
This definition can be incorporated to the Wiener filter equation as follows:

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

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.

4 THE PROPOSED ADAPTIVE WIENER FILTER

This section presents and 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.



$$P_v(\omega) = \sigma_v^2 \qquad (10)$$

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:

$$x(n) = m_x + \sigma_x w(n) \qquad (11)$$

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:

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

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

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

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

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

$$= m_x + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}\,(x(n) - m_x) \qquad (14)$$

If it is assumed that m_x and σ_s are updated at each sample, we can say:

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

Figure 1: Typical adaptive speech enhancement system for additive noise reduction

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:

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 $\hat{s}(n)$ is assumed to 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 signal statistics in the system of Figure 1, the algorithm developed uses the signal variance σ_s^2. The specific method used to designing the space-variant *h(n)* is given by (17.b).

Since $\sigma_x^2 = \sigma_s^2 + \sigma_v^2$ may be estimated from *x*(n) by:

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

Where

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

By this proposed method, we guarantee that the filter transfer function is adapted from sample to sample based on the speech signal statistics.

5 EXPERIMENTAL RESULTS

For evaluation purposes, we use different speech signals like the handel, laughter and gong signals. White Gaussian noise is added to each speech signal with different SNRs. The different speech enhancement algorithms such as the spectral subtraction method, the Weiner filter in frequency domain and the proposed adaptive Wiener filter are carried out on the noisy speech signals. The peak signal to noise ratio (PSNR) results for each enhancement algorithm are compared.

In the first experiment , all the above-mentioned algorithms are carried out on the Handle signal with different SNRs and the output PSNR results are shown in Fig. (2). The same experiment is repeated for the Laughter and Gong signals and the results are shown in Figs.(3) and (4), respectively.

From these figures, it is clear that the proposed adaptive Wiener filter approach has the best performance for different SNRs. The adaptive Wiener filter approach gives about 3-5 dB improvement at different values of SNR. The non-linearity between input SNR and output PSNR is due to the adaptive nature of the filter.

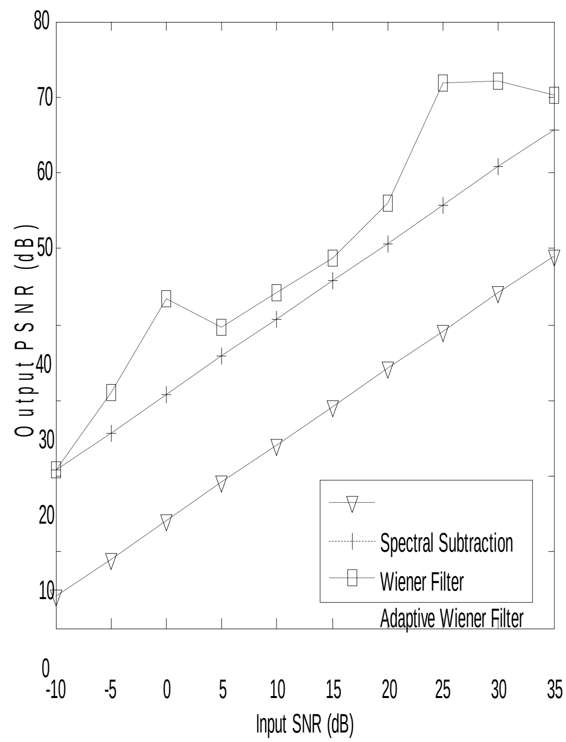


Figure 2: PSNR results for white noise case at-10 dB to +35 dB SNR levels for Handle signal

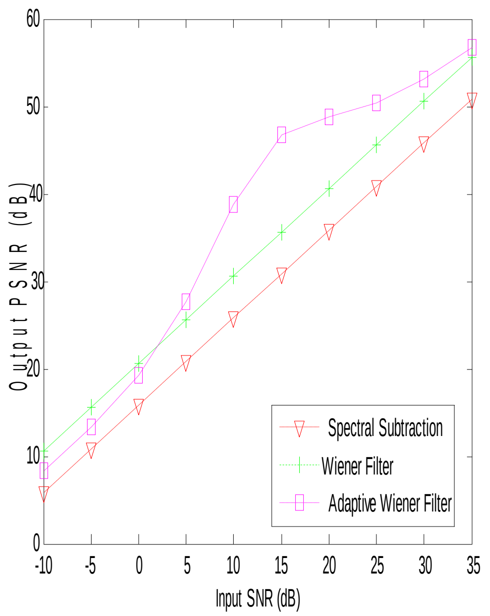


Figure 3: PSNR results for white noise case at -10 dB to +35 dB SNR levels for Laughter signal

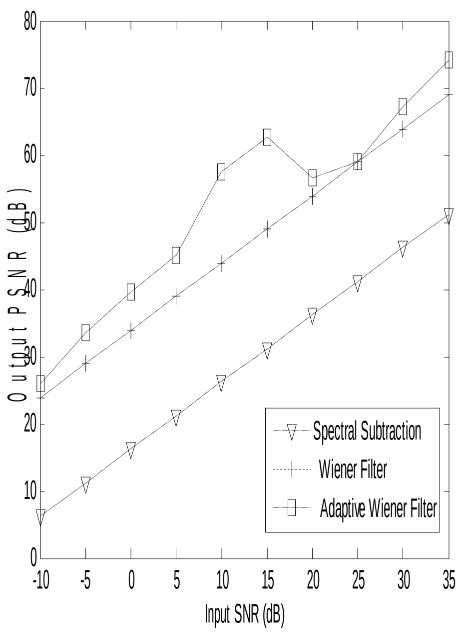


Figure 4: PSNR results for white noise case at -10 dB to +35 dB SNR levels for Gong signal

The results of the different enhancement algorithms for the handle signal with SNRs of 5, 10,15 and 20 dB in the both time and frequency domain are given in Figs. (5) to (12). These results reveal that the best performance is that of the proposed adaptive Wiener filter.

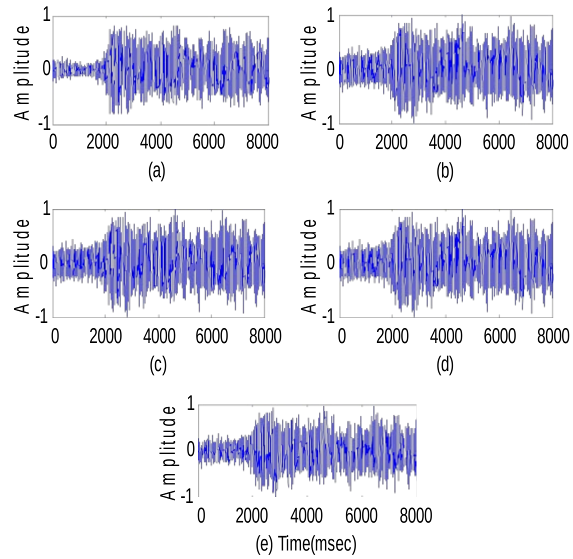


Figure 5: Time domain results of the Handel sig. at SNR = +5dB (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

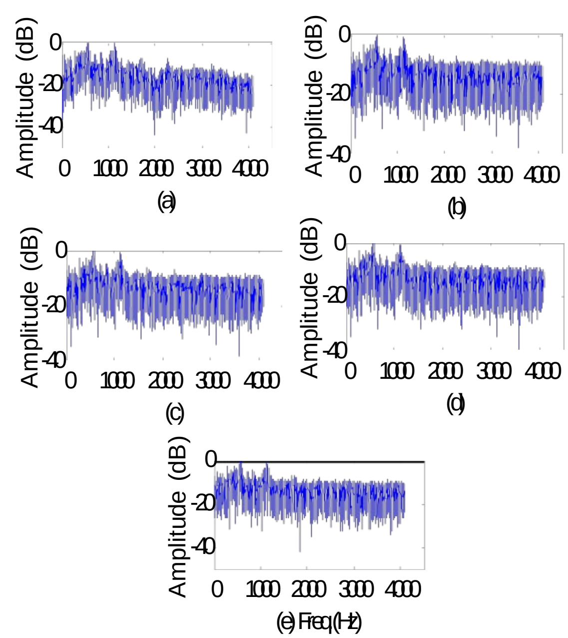


Figure 6: The spectrum of the Handel sig. in Fig.(5) (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

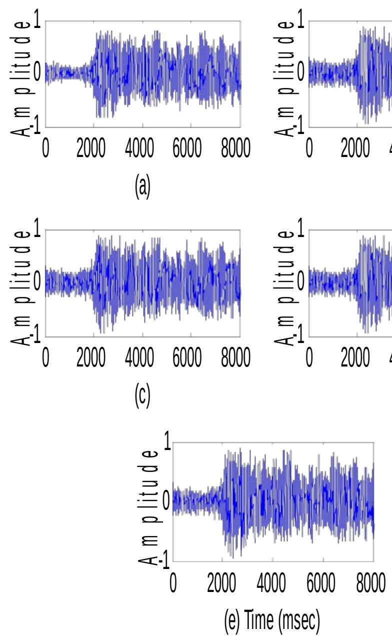


Figure 7: Time domain results of the Handel sig. at SNR = 10 dB (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

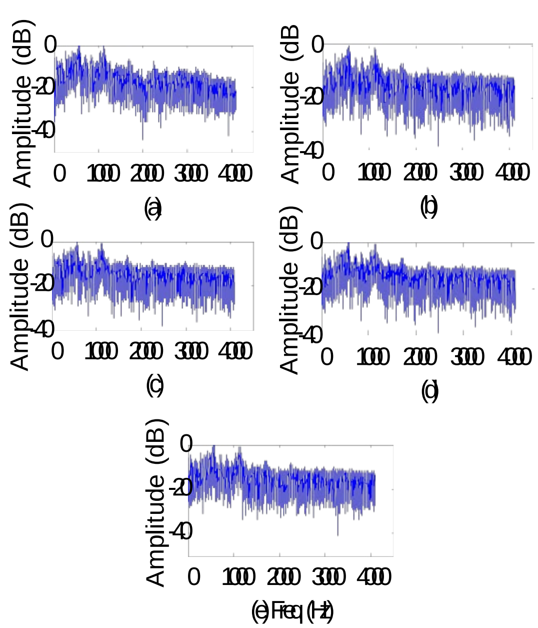


Figure 8: The spectrum of the Handel sig. in Fig.(7) (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

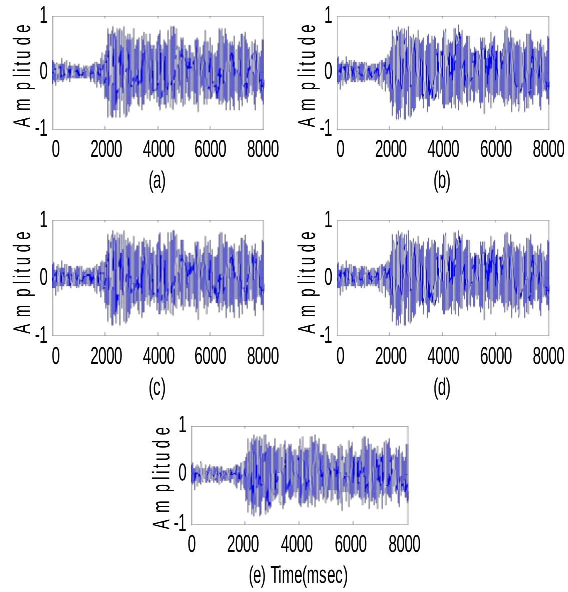


Figure 9: Time domain results of the Handel sig. at SNR = 15 dB (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

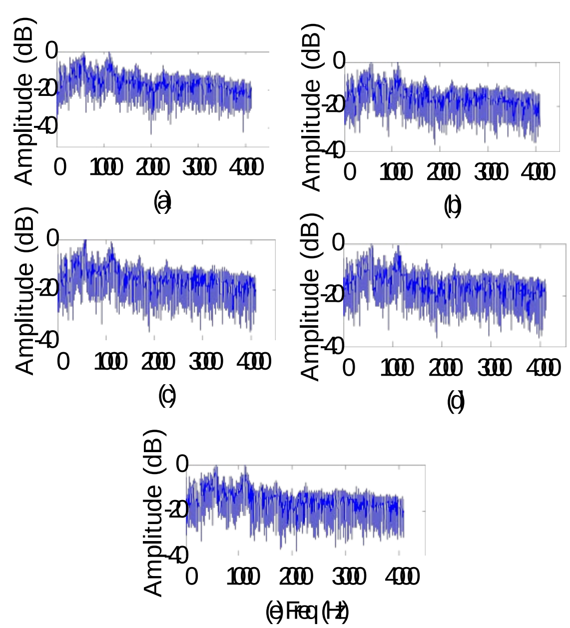


Figure 10: The spectrum of the Handel sig. in Fig.(9) (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

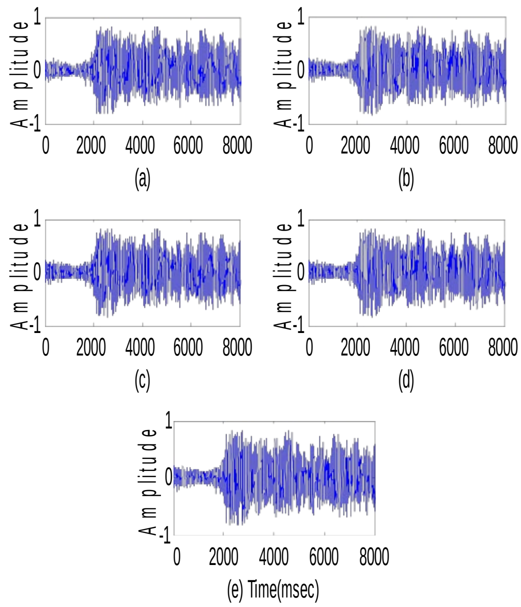


Figure 11: Time domain results of the Handel sig. at SNR = 20 dB (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

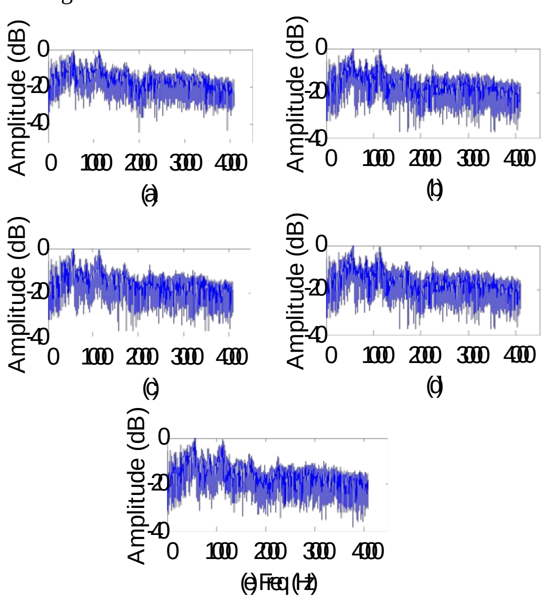


Figure 12: The spectrum of the Handel sig. in Fig.(11) (a) original sig. (b) noisy sig. (c) spectral subtraction. (d) Wiener filtering. (e) adaptive Wiener filtering.

6 CONCLUSION

An adaptive Wiener filter approach for speech enhancement is proposed in this papaper. This approach depends on the adaptation of the filter transfer function from sample to sample based on the speech signal statistics(mean and variance). This results indicates that the proposed approach provides the best SNR improvement among the spectral subtraction approach and the traditional Wiener filter approach in frequency domain. The results also indicate that the proposed approach can treat musical noise better than the spectral subtraction approach and it can avoid the drawbacks of Wiener filter in frequency domain .

REFERENCES

[1] S. F. Boll: Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27,. pp. 113-120 (1979).

[2] M. Berouti, R. Schwartz, and J. Makhoul: Enhancement of speech corrupted by acoustic noise, Proc. IEEE Int. Conf. Acoust., Speech Signal Processing, pp. 208-211 (1979).

[3] Y. Ephriam and H. L. Van Trees: A signal subspace approach for speech enhancement, in Proc. International Conference on Acoustic, Speech and Signal Processing, vol. II, Detroit, MI, U.S.A., pp. 355-358, May (1993).

[4] Simon Haykin: Adaptive Filter Theory, Prentice-Hall, ISBN 0-13-322760-X, (1996).

[5] J. S. Lim and A. V. Oppenheim.: All-pole Modelling of Degraded Speech, IEEE Trans. Acoust., Speech, Signal Processing, ASSP-26, June (1978).

[6] Y. Ephraim and H. L. Van Trees, A spectrally-based signal subspace approach for speech enhancement, in IEEE ICASSP, pp. 804-807 (1995).

[7] Y. Hu and P. Loizou: A subspace approach for enhancing speech corrupted by colored noise, in Proc. International Conference on Acoustics, Speech and Signal Processing, vol. I, Orlando, FL, U.S.A., pp. 573-576, May (2002).

[8] A. Rezayee and S. Gazor: An adaptive KLT approach for speech enhancement, IEEE Trans. Speech Audio Processing, vol. 9, pp. 87-95 Feb. (2001).

[9] U. Mittal and N. Phamdo: Signal/noise KLT based approach for enhancing speech degraded by colored noise, IEEE Trans. Speech Audio Processing, vol. 8, NO. 2, pp. 159-167,(2000).

[10] John R. Deller, John G. Proakis, and John H. L. Hansen. Discrete- Time Processing of Speech

Signals. Prentice-Hall, ISBN 0-02-328301-7 (1997).
[11] S. F. Boll: Suppression of Acoustic Noise in Speech Using Spectral Sub- traction. IEEE Trans. Acoustics, Speech, and Signal Processing. vol. ASSP-29. no. 2, pp. 113-120, April (1979).
[12] M. R. Weiss, E. Aschkenasy, and T. W. Parsons: Processing Speech Signal to Attenuate Interference, in Proc. IEEE Symp. Speech Recognition, pp. 292-293, April (1974).
[13] J. S. Lim and A. V. Oppenheim: Enhancement and band width compression of Noisy speech, Proc. of the IEEE, vol. 67, No..12, pp. 1586-1604, Dec. (1979).