Speech Enhancement with Adaptive Wiener Filtering Approach

Ngu War Nyan Win, Ei Mon Mon Swe
University of Computer Studies, Yangon.
nguwar.gwar@gmail.com,eimonmonswe@gmail.com

Abstract

Presence of noise in speech signals deteriorates the performance of automatic speech recognition systems. Noise reduction is very important key-point of the speech enhancement in speech recognition system. The adaptive wiener filtering is a method used to enhance a speech signal which has been degraded by the noise. This approach is implemented in time domain and adaptation of the filter transfer function. The adaptive wiener filtering approach is improving the SNR in the enhanced speech signal. This approach emphasized to reduce the stationary noise. The spectrogram has been performed for speech signal corrupted by white noise, vehicle noise and babble noise. The results demonstrate that this method can perform better for white noise than the vehicle noise and babble noise.

1. Introduction

Many systems rely on Automatic Speech Recognition (ASR) to carry out their necessary tasks. In much speech communication, the presence of background noise causes the quality and intelligibility of speech to degrade, especially when the Signal-to-Noise Ratio (SNR) is low. Speech enhancement algorithms are of great interest because they have many applications such as speech recognition, voice communication and hearing aids, etc. Several techniques have been proposed for this purpose like the spectral subtraction approach, the iteration Wiener filter [7], Wiener filtering [3] and adaptive Wiener filtering. The performances of these techniques depend on quality and intelligibility of the proposed speech signal.

Spectral subtraction is one of the most used techniques for noise reduction [1]. Spectral subtraction is implemented to estimate the short-time spectral magnitude of speech by subtracting noise estimation from the noisy speech [4]. There are a few limitations to the operation of the method [5]. Firstly, the resultant magnitude cannot fall below zero (rectification). Secondly, the technique tends to generate short duration narrow bands energy which sound like “musical” tones and which need to be eliminated. In this system, the attenuation characteristics change with the length of the analysis window. The main problem for using spectral subtraction, it generates unpleasant artificial sound so that it is not suitable for speech coding. And it remains usually a level of residual, unnatural background noise called musical noise.

Wiener filter method has been proposed by J.S.Lim [6], and the method designs the optimal filter minimizing the mean squared error (MSE) in the frequency domain. The musical noise is reduced by the Wiener filter method than the SS method. If accurate power spectrum for clean speech and accurate power spectrum of additive noise can be estimated, the Wiener filter can be designed accurately. However, the power spectrum of clean speech cannot be observed directly. The drawback of the Wiener filter is the fixed frequency response at all frequency response at all frequencies and the requirement to estimate the power spectral density of the clean signal and noise prior to filtering.

Other methods used to reduce the amount of noise in speech signals include: Iterative Wiener filter (IWF) [7]. This method has been adopted as the speech enhancement that estimates speech and noise power spectra using LPC analysis iteratively. The complex speech analysis can estimate more accurate spectrum in low frequencies, thus it is expected that it can perform better for the IWF especially for babble noise or car internal noise that contains much energy in low frequencies.

In this paper, we use the adaptive Wiener filtering by adapting the Wiener filtering in frequency domain [2]. The adaptive Wiener filtering is implemented in time domain and adaptation of the filter transfer function from sample to sample based on the speech signal statistics (local mean and variance).

2. Sound and Noise

Sound is produced by vibrating objects and reaches the listener’s ears as waves in the air or other media. When an object vibrates, it causes slight changes in air pressure. These air pressure changes travel as waves through the air and produce sound. When speaking of noise in relation to sound, what is commonly meant is meaningless sound of greater than usual volume.

Thus, a loud activity may be referred to as noisy. However, conversations of other people may be called noise for people not involved in any of them, and noise can be any unwanted sound such as the noise of dogs barking, neighbors playing with loud music, road traffic sounds, or aircraft, etc. Now noise is presented in many situations of daily life .In many speech communication environments occurs this situation in which the speech signal is superposed by background noise. Some applications are very import to reduce the level of background noise.

Background noise can be stationary or non-stationary noise. In the case of stationary noise, the noise characteristic remains the same with respect to time and spectrum. Some examples for stationary noise are idle engine sound, vehicle and tire noise in a running vehicle, machinery (white), wind, sea wave and babble. White noise is a sound that contains every frequency within the range of human hearing (generally from 20 hertz to 20 kHz) in equal amounts. White noise draws its name from white light in which the power spectral density of the light is distributed over the visible band in such a way that the eye's three color receptors are rather equally stimulated. White noise is a type of noise that is produced by combining sounds of all different frequencies together [8]. Because it contains all frequencies white noise can drown or mask other sounds which may contain significant information needed for input into an ASR system. If a reasonable estimate of white noise contained in a given speech signal can be obtained and removed from a signal, then we should see an improvement in the quality of the speech and efficiency for most ASR systems. And in the case of non-stationary noise, the noise characteristic varies with time and/or the spectrum of non- stationary noise. Some examples for non-stationary noise are the rubbing of a hand, footsteps, passing vehicles, sirens, horns, coughs and sneezes. In this system, we emphasized to reduce the stationary noise for the enhance speech signal.

3. Adaptive Wiener Filtering

Reducing the level of background noise is very important in many communication systems. Several techniques such as the spectral subtraction and Wiener filter are used to reduce the noise and enhance the speech signal

This approach is implemented in time domain and 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 implementation of the adaptive Wiener filtering is shown in figure 1.

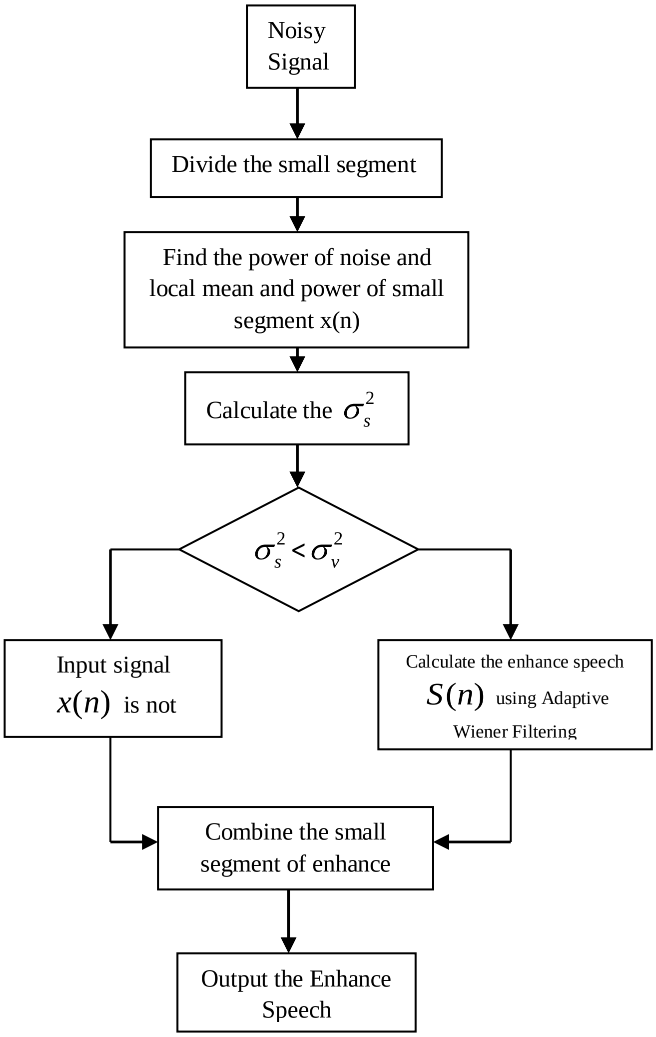


Figure 1: The block diagram of the system.

This system is implemented by the following steps:

Step 1: Divide the small segment signal

Consider a small segment of the speech signal in which the noisy signal $x(n)$ is assumed to be stationary. The input noisy signal is digitized at a rate of 16 kHz, and the time series are processed in frames.

The input noisy signal obtained by: $x(n) = s(n) + v(n)$

Where,

$s(n)$ = clean speech signal

$v(n)$ = noise signal

And then, divide a small segment of the speech signal in which the noisy signal $x(n)$ from the input noisy signal. The number of samples considered per segment is 20 msec.

Step 2: Calculate the power of noise $P_v(w)$

Next, we find the power of noise $P_v(w)$ to calculate the power of speech σ_s^2.It is assumed that the additive noise $v(n)$ is zero mean, so thus, the power spectrum of noise $P_v(w)$ is obtained by:

$$P_v(w) = \sigma_v^2$$

There is an estimate of the power spectrum of the noise $P_v(w)$ which is obtained by averaging over multiple frames of a known noise segment.

Step 3: Calculate the local mean of noisy signal $m_x(n)$

Next, we find a local mean of signal $m_x(n)$ to calculate the enhance speech $S(n)$ .We assumed that m_v is zero, so that we can calculate $m_x(n)$ from *x(n)* by:

$$m_x(n) = \frac{1}{(2M+1)} \sum_{k=n-M}^{n+M} x(k)$$

where, (2M+1) is the number of samples in the small segment.

Step 4: Calculate the power of speech signal σ_s^2 and compare the power of noisy signal $P_v(w)$

The power of noisy signal $\sigma_x 2$ is get from the sum of the power of speech signal σ_s^2 and power of noise signal σ_v^2. So, we can estimate the power of speech signal σ_s^2 from *x (n)* by:

$$\sigma_s^2(n) = \begin{cases} \sigma_x^2(n) - \sigma_v^2, \ if \ \ \sigma_x^2(n) > \sigma_v^2 \\ \\ 0 \qquad\qquad\qquad , \text{ otherwise} \end{cases}$$

The power of noisy signal $\sigma_x{}^2$ can be estimated by:

$$\sigma_x^2(n) = \frac{1}{(2M+1)} \sum_{k=n-M}^{n+M} (x(k) - m_x(n))^2$$

And then, compare the power of noise signal σ_v^2 and the power of speech signal σ_s^2.If σ_s^2 is greater than σ_v^2 the output signal $S(n)$ is assumed to be primarily due to $x(n)$ and signal $S(n)$ is not attenuated. Otherwise, the filtering effect is performed.

For the small segment of speech, the Wiener Filter transfer function can be approximated by:

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

Step5: Calculate the enhance speech $S(n)$

From the above steps we get the enhance speech $S(n)$ within the local segment can be expressed as:

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

On this step, we get the value of local mean $m_x(n)$ and $(x(n) - m_x(n))$ are changed separately from segment to segment. Finally, we get the enhance speech by combing the segments.

4. Experimental Results

The perception of a speech signal is usually measured in terms of its quality and intelligibility. In order to assess the performance of speech signal, objective test, subjective test and spectrogram can perform.

4.1 Objective Evaluation

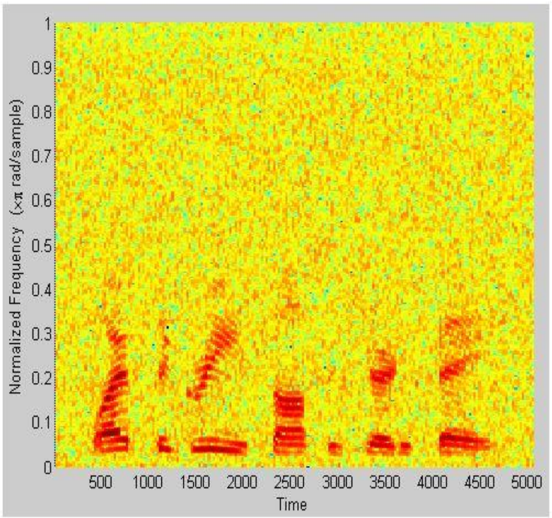
Signal-to-noise ratio (SNR) is defined as the ratio of a signal power to the noise power corrupting the signal. SNR compares the level of a desired signal to the level of background noise. The higher the ratio, the less obtrusive the background noise is.

4.2 Subjective Evaluation

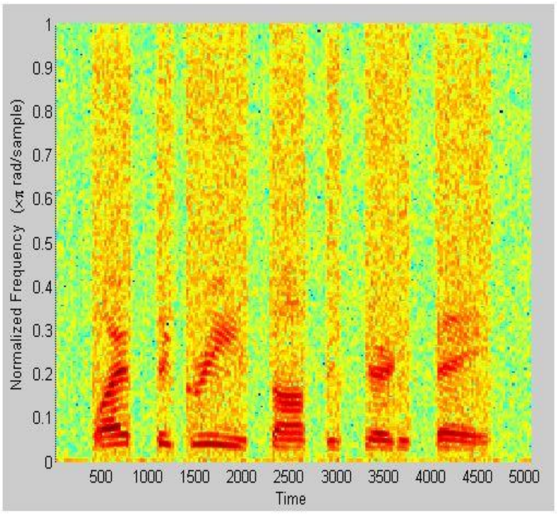
The quality is a subjective measure which reflects on individual preferences of listeners. The standard of the subjective evaluation uses separate rating scales to independently estimate the subjective quality of the speech signal alone, the background noise alone, and overall quality.

4.3 Speech Spectrogram

Objective measured do not given indications about the structure of the residual noise. Speech spectrograms constitute a well-suited tool for observing this structure. Thus, we use the spectrogram to show the performance of the system. We use sentence spoken by English news is corrupted in three stationary noise environments from NOISEX database to test the performances of the simulation programs. And we also test this program by different SNR. But this algorithm does not work for very low SNR. For the higher SNR, the speech signal gets better quality. Figure 2, Figure 3 and Figure 4 show the output results with input noisy speech for SNR levels with 5 dB. From this result, we can see that enhancement is about 2-3dB.

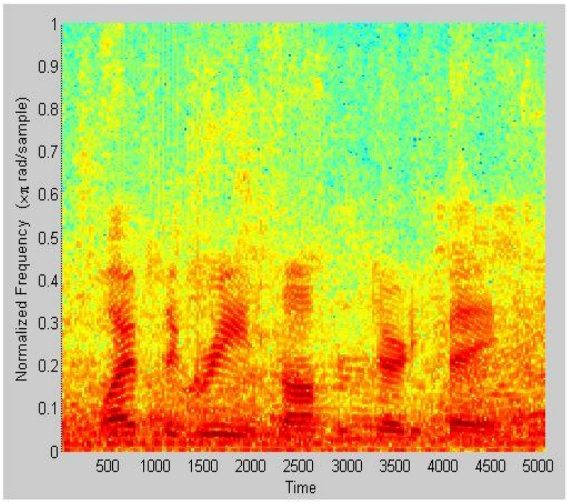


(a)Noisy Signal

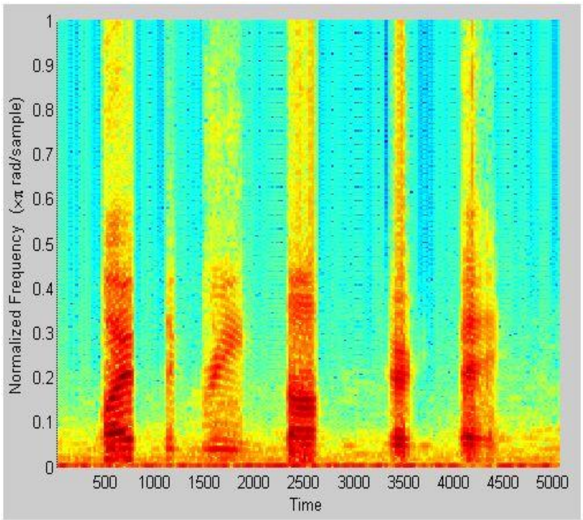


(b) Enhance Speech Signal

Figure 2: The spectrogram of (a) Noisy Signal and (b)Enhance Speech Signal with the White noise at SNR level =5dB. On this figure, the SNR is improve 3dB above the original SNR using white noise.

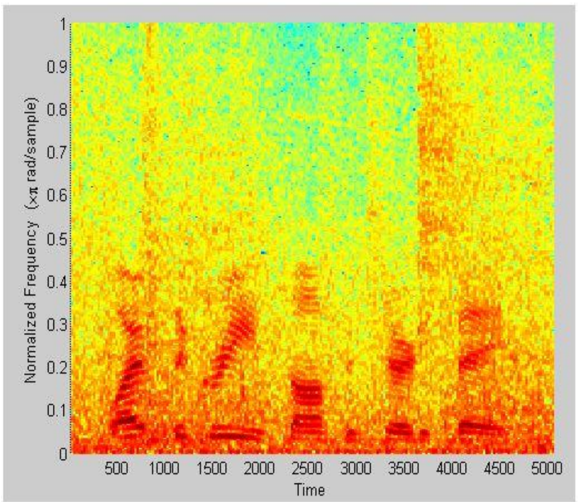


(a) Noisy Signal

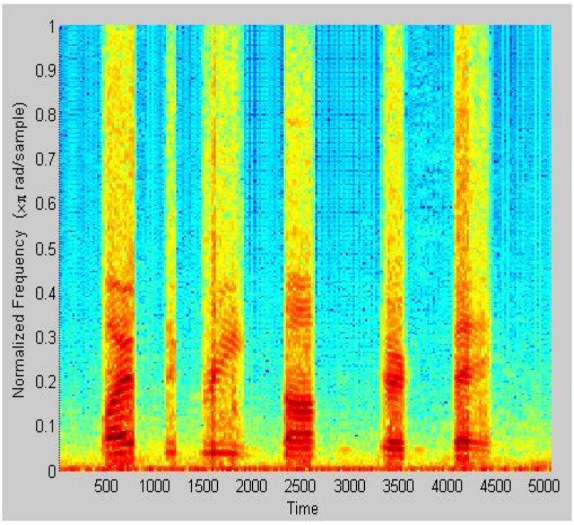


(b) Enhance Speech Signal

Figure 3: The spectrogram of (a) Noisy Signal and (b)Enhance Speech Signal with the babble noise at SNR level = 5dB. On this figure, the SNR is improve 2dB above the original SNR using babble noise.



(a) Noisy Signal



(b) Enhance Speech Signal

Figure 4: The spectrogram of (a) Noisy Signal and (b)Enhance Speech Signal with the Vehicle noise at SNR level = 5dB. On this figure, the SNR is improve 2dB above the original SNR using vehicle noise.

The speech signal of Figure 2.(a) is corrupted by white noise. The speech signal of Figure 2.(b) is enhanced by using the Adaptive Wiener Filtering. On this process, the speech signal is improve the SNR level 3 dB. And the Figure 3(a) and 4(a) is corrupted by babble noise and vehicle noise respectively. But on this noise, the SNR level is imporve 2 dB. The speech signal of Figure 3(b) and 4(b) is enhanced by using the Adaptive Wiener Filtering respectively.So that this method can perform better for the white noise than babble noise and vehicle noise.

5. Conclusions

This approach is based on using the wiener transfer function from sample to sample of the speech signal statistics (local mean and variance).This approach can treat the musical noise than the spectral subtraction approach. It provides the best SNR improvement. It provides the best noise reduction performance; the output signal-to-noise ratio for each speech sample approximates closely to the maximum possible value for that sample. Experimental results have shown that the proposed technique is very effective for enhancing the speech signal. And from this result have shown that this method can perform better for the white noise than babble noise and vehicle noise as the stationary noise.

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] M. A. Abd El-Fattah, M. I. Dessouky, S. M. Diab and F. E. Abd El-samie,”Adaptive wiener filtering approach for speech enhancement.”, Progress In Electromagnetic Research M, Vol. 4, 167–184, 2008.

[3] S. Chirtmay and M. Tahernezhadi ,” Speech Enhancement using Wiener Filtering “,Department os Electrical Engineering, Northern Illinois.

[4] Jeffery J. Faneuff , D. Richard Brown III , “Noise Reduction and Increased VAD accuracy using Spectral Subtraction”, ISPC ’03, March 31-April 3, 2003 , Dallas,TX.

[5] A.G. Maher, R.W.King, ” A comparison of noise reduction techniques for speech recognition in telecommunications Environments”, The institution of Engineers Australia Communications Conference Sydney, October 20-22 1992.

[6] J.S.Lim and A.V.Oppenheim, “All-pole modeling of degraded speech,” IEEE Trans., ASSP-26, pp.197-210, 1978.

[7] F. Keiichi, “Speech enhancement based on Iterative Wiener Filter using Complex Speech Analysis”, Computing & Networking Center, Univ. of the Ryukyus Senbaru 1, Nishihara, Okinawa, 903-0213, Japan.

[8] “Howstuffworks.” *What is White Noise?* 1998 – 2005.August 5, 2005. <http://home.howstuffworks. com / question47.htm>