Adv. Studies Theor. Phys., Vol. 2, 2008, no. 14, 673 - 685

Quality Improvement of Speech Signals Using LPC Analysis

N. Esfandian[1] and E. Nadernejad[1,2]

[1] Ghaemshahr Branch, Islamic Azad University
Ghaemshahr, Iran
Na_esfandian@yahoo.com

[2] Department of Computer Engineering, Faculty of Engineering
Mazandaran Institute of Technology (Non-Governmental)
P.O. Box: 744, Babol, Iran
ehsan_nader@nit.ac.ir

Abstract

This paper, proposes a new methods for speech Signal enhancement based on spectral subtraction, Inverse Fourier Transform. We use the Linear Predicative Coding (*LPC*) analysis for noise estimation and extraction, and then we compare the proposed method with the previous ones and are able to recover the speech signal much better than the previous methods.

Keywords: Speech signals, Noise, Speech enhancement, Spectral subtraction, multi-band spectral subtraction, *LPC* analysis

1. Introduction

Noise reduction is an important issue in speech signal processing systems, like speech signals coding, speech recognition. Thus many methods have been proposed for noise reduction in speech signals, some of which are methods based on spectral subtraction (base & Multi-Band) [5, 6, 7, 8], adaptive filtering [11], Wavelet transform [10, 12, 13, 14, 15].

In the spectral subtraction methods, 3 conditions should be met;

a. Noise must be additive.
b. Noise and the signal must be uncorrelated.
c. A canal must be accessible.

Although the base spectral subtraction method is very simple and efficient, it add a new noise named musical noise. For reducing this noise, the spectral subtraction method applying spectral floor and over-subtraction, can be used [6]. Later, was proposed the multi-band spectral subtraction method in [5]. In this method, the corrupted speech signal is initially divided into several frequency bands, then the spectral subtraction method is applied to each band.

As mentioned before, in this approach it is supposed that the signal and noise are uncorrelated, but actually thus assumption really happens in speech signals. Hence, the inverse Fourier spectral subtraction method has been present, which is the same as the spectral subtraction method, but here, the subtraction, is applied to the inverse Fourier transform. In this method, the problem of the correlation between the signal and noise is solved to some extent. Also, there are some other methods like cepestral subtraction, wavelet transform [10, 12, 13, 14, 15] for noise reduction or elimination (de-noising).

In this paper, first, we describe the base spectral subtraction, multi-band spectral subtraction, and inverse Fourier spectral subtraction methods, then we estimated the noise from the speech signal using the LPC analysis [1, 2, 3, 4]. At last, we apply the estimated noise to the multi-band spectral subtraction, and inverse Fourier spectral subtraction methods and we see a noticeable improvement in these methods.

2. Power Spectral Subtraction (PSS)

We suppose that the signal and noise are additive, so a corrupted speech signal can be expressed as bellow:

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

where $x(n)$ is the corrupted speech signal, $s(n)$ is the clean spectral signal, and $n(n)$ is a random noise signal.

According to the second assumption, the signal and noise are uncorrelated, so we can write: [1]

$$R_n(\tau) = D_0\delta(\tau) \tag{2}$$

$$R_{s,n}(\tau) = 0 \tag{3}$$

where D_0 is a constant, $R_n(\tau)$ is the auto-correlation of the random noise signal, and $R_{s,n}(\tau)$ is the cross-correlation function of the *s* and *n* signals. According to the relations above and by supposing that the s and n signals are stationary, we can write:

$$\Gamma_x(\omega) = \Gamma_s(\omega) + \Gamma_n(\omega) \tag{4}$$

Where Γ_x, Γ_s , Γ_n are Power Density Spectrum (PDS) of *x*, *s*, *n*, respectively. Following eq.(4) by estimating the PDF of the random noise signal, the PDF of the clean speech signal can be estimated as expressed below:

$$\hat{\Gamma}_s(\omega) = \Gamma_x(\omega) - \hat{\Gamma}_n(\omega) \tag{5}$$

where $\hat{\Gamma}_s(\omega)$, $\hat{\Gamma}_n(\omega)$ are Estimations of the $\Gamma_s(\omega)$, $\Gamma_n(\omega)$. Equations (4) and (5) are true only when the clean speech signal and the noise are stationary, but actually this is not always true. Since the clean speech signals are locally stationary in short-time frames, and additionally the assumption that noise is stationary is more acceptable in short time intervals, windowing is applied to the corrupted speech signal. Then the spectral subtraction is applied to each frame by considering *m* as the window number, we have:

$$x(n;m) = s(n;m) + n(n;m) \tag{6}$$

$$R_n(\tau;m) = D_0\delta(\tau) \tag{7}$$

$$R_{s,n}(\tau;m) = 0 \tag{8}$$

Where $x(n;m)$ is the windowed signal of the speech signal $x(n)$. By calculating the PDF for both sides of eq. (6) we have:

$$\Gamma_s(\omega;m) = \Gamma_x(\omega;m) - \Gamma_n(\omega;m) \tag{9}$$

Also we know that [1]:

$$\Gamma_x(\omega;m) = \frac{X(\omega;m)X^*(\omega;m)}{\mathrm{N}^2} = \frac{|X(\omega;m)|^2}{\mathrm{N}^2} \tag{10}$$

where N is the window length (size) and X is the speech signal. The factor $1/N^2$ can be simply neglected, since $|X(\omega;m)|^2$ is the more bigger than the denominator:

$$\Gamma_x(\omega;m) = |x(\omega;m)|^2 \tag{11}$$

The following relation can be achieved using the relations (11), (9):

$$|S(\omega;m)|^2 = |X(\omega;m)|^2 - |N(\omega;m)|^2 \tag{12}$$

where $|X(\omega;m)|$ is the magnitude of Fourier transform for the windowed *x(n)*; $|S(\omega;m)|$ and $|N(\omega;m)|$ are the magnitude of the Fourier transform for the windowed clean speech signal and windowed noise signal, respectively.

As can be seen from the eq (12), to computing the magnitude of the Fourier transform of the clean signal, we need the magnitude of the random noise, hence the random noise signal is estimated from the silence.

There is no speech signal in the silence part.

Now, for achieve the clean speech signal in the time domain, it is necessary to calculate the magnitude of the Fourier transform as well as it is phase, and by short time fast Fourier transform (st.FFT) get the speech signal in time domain.

In all practical applications, the phase of the clean speech sign can be considered equal to the phase of the corrupted speech signal [8].

$$\varphi_{\hat{S}(\omega;m)} = \varphi_{X(\omega;m)} \tag{13}$$

This means that the effect of noise on the phase of the speech signals is not sensible for human ear.

According to the equations (12), (13), the clean speech signal can be estimated as below:

$$\hat{S}(\omega;m) = \left|\hat{S}(\omega;m)\right| \exp\left\{i\,\varphi_{\hat{S}(\omega;m)}\right\}$$
$$= \left[\left|X(\omega;m)\right|^2 - \left|\hat{N}(\omega;m)\right|^2\right]^{1/2} \exp\left\{i\,\varphi_{\hat{S}(\omega;m)}\right\} \tag{14}$$

Where $\hat{S}(\omega;m)$, $\hat{N}(\omega;m)$ are the fourier transform of the estimated clean signal and fourier transform of the estimated noise signal, respectively.

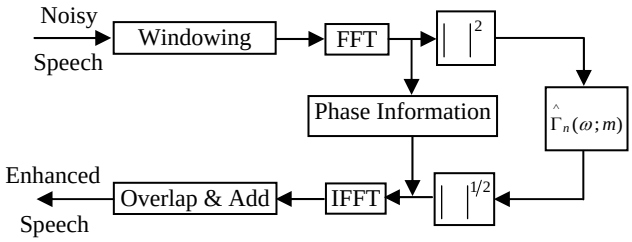


Fig1. diagram block of Power Spectral Subtraction (PSS)

The above method is called the Power Spectral Subtraction (PSS) methods. Fig.1. Because the second order of the magnitude of the fourier transform, which indicates the power of the signal, is being used, usually they use another power factor other than 2, in the spectral subtraction method, the magnitude of which is achieved using optimizing techniques. The method mentioned above is called the general spectral subtraction (GSS) method.

$$\hat{S}(\omega;m) = \left[\left|X(\omega;m)\right|^a - \left|\hat{N}(\omega;m)\right|^a\right]^{1/a}$$
$$\exp\left\{i\,\varphi_{\hat{S}(\omega;m)}\right\} \tag{15}$$

But, important problem in the spectral subtraction method is the negative values of the fourier transform of the clean signal.

In order words, we can't certainly assume the fourier transform of the clean speech signal in each of the relations (14) and (15), as a positive value. There are two methods for correcting these negative values [1]:

a) half-wave correction :

$$\left|\hat{S}(\omega;m)\right| = \begin{cases} \left|\hat{S}(\omega;m)\right| & if\ \left|\hat{S}(\omega;m)\right| > 0 \\ 0 & elsewhere \end{cases} \tag{16}$$

b) Full-wave correction :

$$\left|\hat{S}(\omega;m)\right| = abc\left\{\left|\hat{S}(\omega;m)\right|\right\} \tag{17}$$

3. Inverse Fourier Spectral Subtraction (IFSS)

In the spectral subtraction method [5, 6, 7, 8], it is assumed that the noise and the signal are uncorrelated. This condition can be met by applying the auto-correlation function to both sides of equation (1). Now, if the accuracy of the relation (4) or (9) is reduced, the accuracy of the un correlation between the signal and noise would become less consequently.

In the inverse Fourier spectral subtraction method, subtraction is applied to the inverse Fourier transform of the magnitude of the Fourier transform of the corrupted signal and the estimated noise signal. It can be evidently said that in the inverse Fourier subtraction method, the subtraction is performed in the time domain in which the un correlation between the signal and noise has less accuracy. Because usually noise is added to the signal in the time domain where it's not certainly uncorrelated, but addition in the frequency domain needs uncorrelation.

In this method, the estimated clean speech signal is calculated according to fig 3.

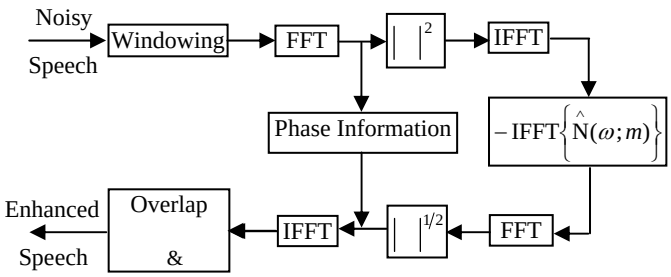


Fig3. diagram block of Inverse Fourier Subtraction

4. Multi-band Spectral Subtraction (MBSS)

As mentioned before, in the base spectral subtraction method, it is supposed that the effect of noise is the same for all the signal rang. However actually this is not actually the case. This is because, in addition to the existing of different noise sources, there is another fact and that is the effect of noise on the speech signal depends on the frequency [7]. This dependence of the noise on frequency, leads us to the fact that the effect of noise on the voice and un-voice letters (signals) is not the same [7]. Fig.2 show the *SNR* in each frame and in four frequency bands for the expression " one four seven three " said by a woman.

Because of the problems mentioned before, *kamath* and *loizou* [5] have proposed the use of spectral subtraction method of *berouti* [6] in several

frequency bands, So that they divided each speech frame into several bands in the frequency domain and then they have applied the relations (15), (19) the windowed signal I each frequency band. In this method, the amount of the musical noise is reduced as well. The important thing in this method is the way the *SNR* is computed in each frame and in each band of that each frame. The amount of *SNR* in each frame and band is calculated as below:

$$SNR_{i,j} = 10\log\frac{\sum_{w=b_i}^{e_i}\left|X(\omega;m_i)\right|^2}{\sum_{w=b_i}^{e_i}\left|\hat{N}(\omega;m_i)\right|^2} \tag{22}$$

Where $SNR_{i,j}$ is the *SNR* in the jth band of the ith frame, $\left|X(\omega;m_i)\right|^2$ is the square of the magnitude of the fourier transform of the corrupted speech signal in the ith frame, $\left|\hat{N}(\omega;m_i)\right|^2$ is the square of the magnitude of the fourier transform of the estimated noise in the ith frame, and b_i and e_i are the initial and the end frequencies of the jth band.

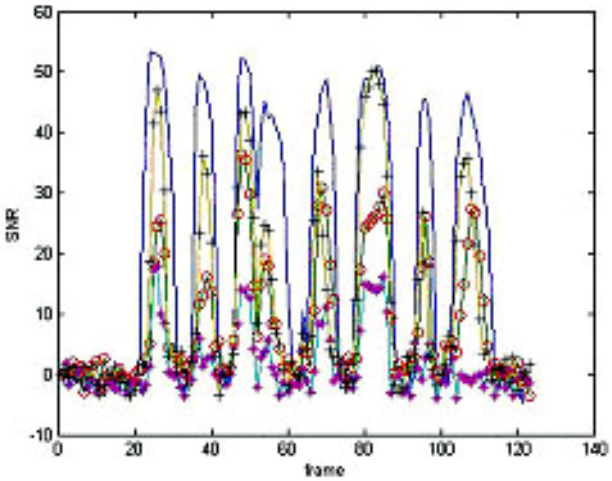


Fig2. *SNR* in four frequency band, in one frame

5. Linear Predication Coefficient (LPC)

Because in our proposed algorithm, LPC analysis [1, 2, 3] is used for noise estimation, in this section we describe this analysis.

The LPC is one of the strongest tools in speech signal processing. The general idea of this analysis is that each sample of the speech sign can be expressed as a linear equation of previous inputs and outputs:

$$s(n) = \sum_{k=1}^{p} a_k\, s(n-k) + \sum_{l=0}^{q} b_l\, u(n-l) \tag{24}$$

Where a_k and b_l are the denominator and nominator of the filter, respectively, and

u(n) is the initial signal which is an impulse burst for voice and is a string of random noise for unvoice [1,2,3,4]. The transform function of the system can be achieved by applying the *Z* transform to the equation (24):

$$H(z) = \frac{S(z)}{U(z)} = \frac{\sum_{l=0}^{q} b_l z^{-l}}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{25}$$

A all pole model is very good estimation for the transform function *H (z)* [1] for speech signals and can be expressed as:

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)} \tag{26}$$

For human's larynx, *P* is an integer number in the range of $10 \le p \le 14$.

The important point in computing the *LPC* is that these coefficients can be directly driven from the speech signal for this reason and because of the dependence of the speech signal on times first, windowing is done the signal then the *LPC* coefficients, are calculated in short frames [2].

4.1. Noise Estimation

According to the discussions above, each sample of the speech signal can computed with a good accuracy just using *P* previous samples of that signal (without using their previous *P* samples) [1]:

$$\hat{S}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{27}$$

the error signal is actually the difference between the main speech signal and the speech signal estimate from *P* previous samples:

$$e(n) = S(n) - \hat{S}(n) = S(n) - \sum_{k=1}^{p} a_k S(n-k \tag{28}$$

If we apply *Z* transform on both sides of the relation above we have:

$$E(z) = S(z)\left[1 - \sum_{k=1}^{p} a_k z^{-k}\right] = A(z)S(z) \tag{29}$$

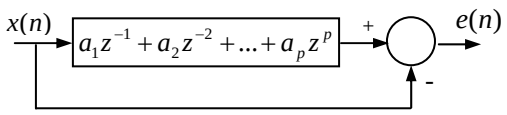


Fig4. diagram block of calculation of error function

where, is the z transform of the error signal and has the $E(z)$ characteristics of a noise, sine a linear filter separates the uncorrelated. Part of the signal the most of which is noise for proving this claim, it's enough to calculate the auto-correlation function of the signal $e(n)$.

In fig.5, the auto-correlation of the error, signal which belongs to a speech signal from the *Timit* database, is plotted. As can be seen, the signal $e(n)$ has the characteristics of noise, since it's auto-correlation signal is the same as the auto-correlation function of the random noise signal.

In our proposed algorithm, we have used this signal for noise estimated that has resulted in a great improvement in the *SNR* of the corrupted speech signals (compared to the exiting methods).

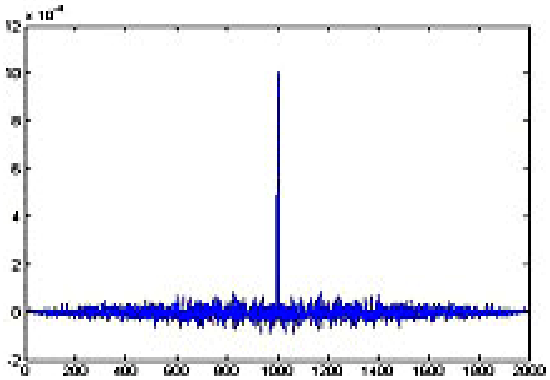


Fig5. Auto-correlation functions of the error signal.

Although the signal $e(n)$ is not the noise signal added to the clean speech signal, it has most it's characteristics. On the other hand, some of the uncorrelated signal related to the speech signal exists in the output of the filter $A(z)$, which is negligible compared to noise.

6. Implementation and comparison of experimental result

In this part we pay to implementation of PSS, IFSS and MBSS spectral subtraction methods. After implementation, with selection of clean speech signal from TIMIT database and noising it with white Gauss Ian noise we produce degraded speech signal. Now, we apply degraded speech signal as input to each of up implementation systems and we consider the output for purposes of signal to noise ratio (SNR), transient state graph and spectrogram.

In after stage, we use from LPC analysis. To this like that we apply obtained outputs from up systems as input to LPC system, that in this condition LPSS, LPIFSS and LPMBSS systems are producing. With considering amount of SNR, transient state graph and spectrogram of new systems (with LPC) we see that results are improved relative to prior state (without LPC).

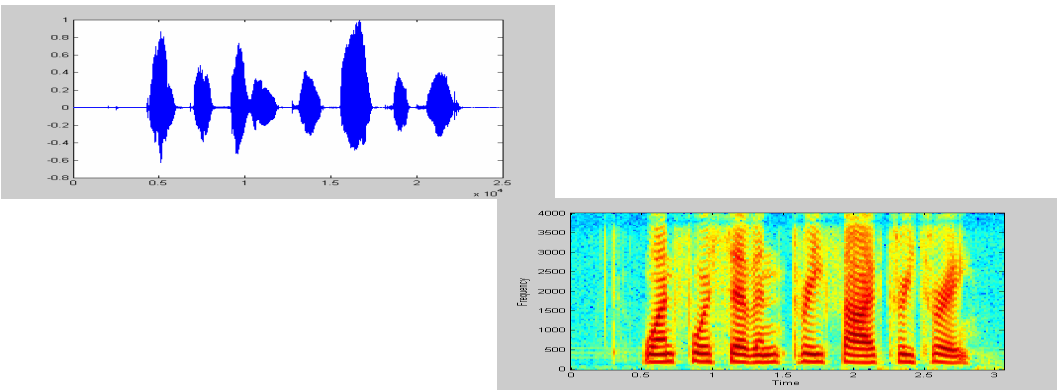
Table (1): Test results without LPC

Measurement type : signal to noise ratio (db) Clean signal :sampled clean speech signal from TIMIT database Added noise to clean signal : White Gaussian noise			
Input SNR **Enhancement Method**	**5 db**	**10 db**	**15 db**
PSS	**10.1222**	**12.4210**	**13.8862**
IFSS	**8.3520**	**13.4906**	**17.6932**
MBSS	**15.6857**	**15.3443**	**15.5970**

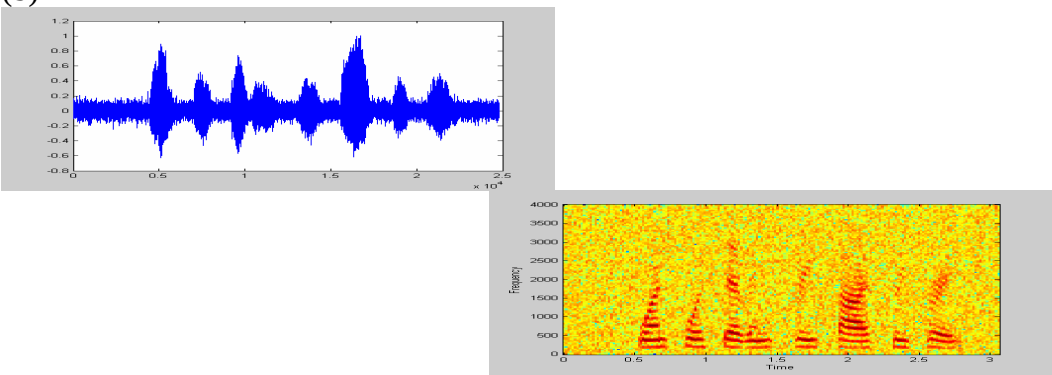
Table (2): Test results with LPC

Measurement type : signal to noise ratio (db) Clean signal :sampled clean speech signal from TIMIT database Added noise to clean signal : White Gaussian noise			
Input SNR **Enhancement Method**	**5 db**	**10 db**	**15 db**
LPSS	**7.6068**	**12.4734**	**17.1987**
LPIFSS	**9.7938**	**14.8062**	**18.7845**
LPMBSS	**15.7268**	**15.6472**	**15.7837**

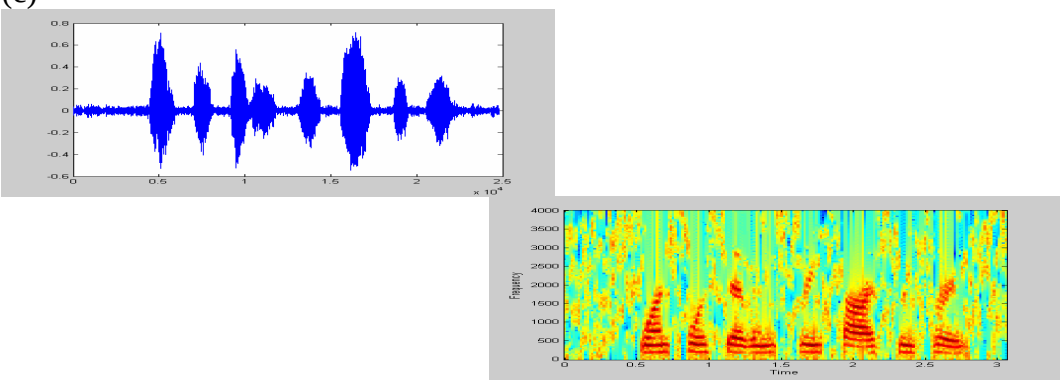
(a)



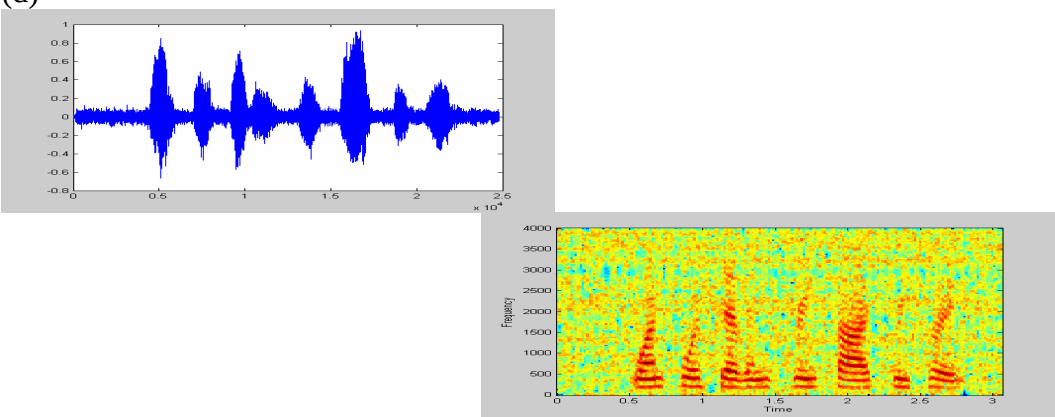
(b)


x 10^4
Frequency
Time

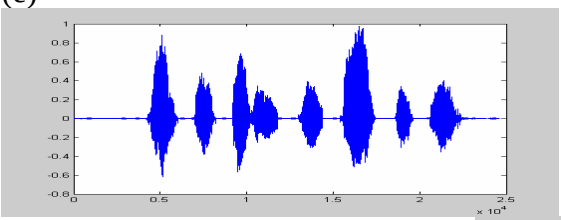
(c)

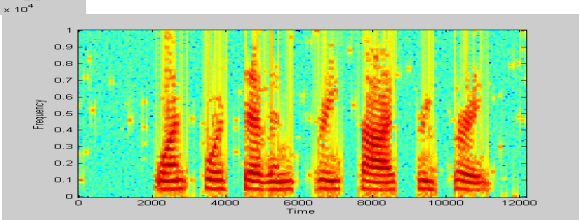

x 10^4
Frequency
Time

(d)

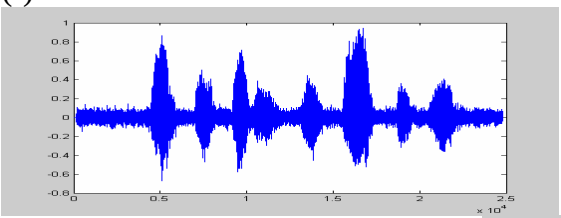

x 10^4
Frequency
Time

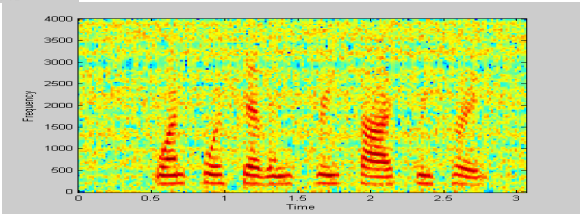
(e)



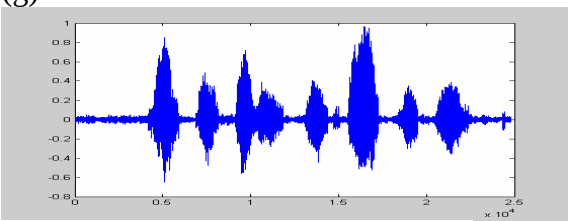


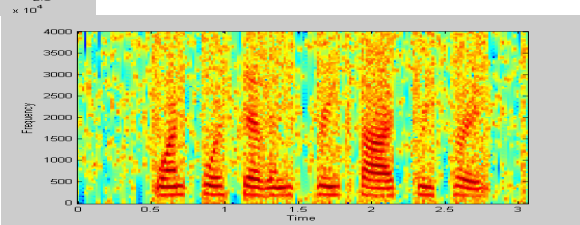
(f)





(g)





Fig(5): Transient state graph
a) clean speech signal
b) degraded speech signal (SNR=5db)
c) Enhanced speech signal with PSS method
d) Enhanced speech signal with IFSS method
e) Enhanced speech signal with MBSS method
f) Enhanced speech signal with LPSS method
g) Enhanced speech signal with LPIFSS method

Fig(6): Spectrogram
a) clean speech signal
b) degraded speech signal (SNR=5db)
c) Enhanced speech signal with PSS method
d) Enhanced speech signal with IFSS method
e) Enhanced speech signal with MBSS method
f) Enhanced speech signal with LPSS method
g) Enhanced speech signal with LPIFSS method

8. Conclusions

In this paper, we proposed a new method for speech signal enhancement based on inverse fourier transform spectral subtraction.

By studying these speech improving methods, we have shown their weakness in the accurate estimation of noise. For solving this problem, the idea of using the *LPC* analysis for noise estimation was proposed. By studying these methods, we understood that in all the methods, there's a need for noise estimation or some of its parameters. As a result, we tried to find a method which is able to give a better and more accurate estimation of noise.

In the LPC analysis, we are looking for a filter and a model for the larynx that has all the larynx characteristics and by applying noise to it's input, we get speech signal at it output. So, if we apply the speech signal to the inverse model (filter), we must get noise signal at it's output.

Since, the un correlate part of the noise speech signal appears at the filter output that because of the filter linearity, most of it is noise.

We have applied this noise in the speech signal enhancement and used it to improve the method above.

By comparing the proposed and the exiting methods, we have seen that the proposed methods improved the *SNR* of the enhanced signals

References

[1] J. R. Deller, J. H. L. Hansen, J.G. proakis, “Discrete-time processing of speech signals”. 2nd edition, IEEE press, 2000.

[2] L. R. Rabiner, R. W. Schafer. “Digital processing of speech signals”. Prentice Hall, 1978.

[3] J. Tierney “A study of LPC analysis of speech in additive noise”, IEEE trans. Acoust. Speech and signal process, ASSP-28,4, pp:389-379 (Aug.1980).

[4] M.r. Sambur, N.s. Jayant “LPC analysis/synthesis from speech inputs containing guantizing noise or additive white noise”, IEEE Trans. Acoust. Speech and signal process. ASSP-24, 6, pp:488-494 (Dec.1976).

[5] S. Kamath P. Loizou, “A Multi-band spectral subtraction method for Enhancing speech corrupted by colored noise”, proceedings of ICASSP-2002, Orlando, FL, May 2002.

[6] M. Berouti, R. Schwartz, J. Makhoul, “Enhancement of speech corrupted by acoustic noise”, proc. IEEE ICASSP , Washington DC, April 1979, 208-211.

[7] Y. Ghanbari, M. R. Karami, B. Amelifard, "Improved multi-band spectral subtraction method for speech enhancement", Proceedings of the 6th ISTED International conference SIGNAL AND IMAGE PROCESSING, pp:225-230, Agust 23-25, 2004, Honolulu, Hawaii, USA.

[8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. On Acoust. Speech & signal processing, vol. ASSP-27, April 1979, pp:113-120.

[9] P. S. Whitehead, D.V. Andeson, M. A. Clements, "Adaptive acoustic noise suppression for speech enhancement", IEEE International Conference on Multimedia & Expo., July 2003.

[10] D. L. Donoho, "De-noising by soft-thresholding, IEEE Transactions on Information Theory", vol. 41, No. 3, May 1995, pp:613-627.

[11] K. Y. Lee, B. G. Lee, S. Ann, "Adaptive filtering for speech enhancement in colored noise", IEEE Trans. On Signal Processing Letters, Vol. 4, October 1997, pp:277-279.

[12] Ing Yann Soon, Soo Ngee Koh, Chai Liat Yeo, "Wavelet for Speech Denoising", TENCON 97, Brisbane, Australia, 1997, pp: 479-482.

[13] H.Sheikhzadeh, H. R. Abutalebi, "An Improved Wavelet-Based Speech Enhancement System", in proc. 7th European Conference on Speech Communication and Technology (EuroSpeech), Aalborg, Denmark, Sep. 2001.

[14] Y. Ghanbari, M.R. Karami, " Spectral Subtraction in the Wavelet Domain for Speech Enhancement", IJSIT, vol.1, No.1, August 2004

[15] Y.Ghanbari, M.R. Karami-Mollaee 'A new approach for speech enhancement based on adaptive thresholding of wavelet packets' Speech communication 48 (2006) 927-940.

[16] H. Sameti, H. Sheikhzadeh, Li Deng, R. L. Brennan, "HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise", IEEE Transactions on Speech and Audio Processing, Vol. 6, No. 5, September 1998.

Received: February 19, 2008