Contemporary Engineering Sciences, Vol. 2, 2009, no. 1, 1 - 15

Speech Signals Enhancement Using LPC Analysis based on Inverse Fourier Methods

Mostafa Hydari, Mohammad Reza Karami

Department of Computer Engineering, Faculty of Engineering Noshirvani Institute of Technology P.O. Box: 844, Babol, Iran
baradaran5@yahoo.com, mkarami@nit.ac.ir

Ehsan Nadernejad

Department of Computer Engineering, Faculty of Engineering Mazandaran Institute of Technology, P.O. Box: 744, Babol, Iran
ehsan_nader@yahoo.com

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*), *VAD* analysis, and Voice/ Unvoice (V/UV) detector for noise estimation and extraction, then we compare the proposed method with the previous ones and are able to recover the speech signal much better than the previous methods. Also, good results have been achieved in the auditory tests.

Keywords: Speech signals, Speech enhancement, Spectral subtraction, multi-band spectral subtraction, Inverse Fourier Spectral Subtraction, *LPC* analysis; *VAD*; V/UV detector

I. 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 adds 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, and 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 spectral subtraction, multi-band spectral subtraction, and inverse Fourier spectral subtraction methods and we see a noticeable improvement in these methods.

II. Power Spectral Subtraction (PSS)

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

$$x(n) \quad s(n) \quad n(n) \tag{1}$$

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]

$$R_n(\ ) \quad D_0 \ (\ ) \tag{2}$$

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

Where D_0 is a constant, $R_n(\ )$ is the auto-correlation of the random noise signal, and $R_{s,n}(\ )$ 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:

$$_x(\ ) \quad _s(\ ) \quad _n(\ ) \tag{4}$$

Where $_x$, $_s$, $_n$ are Power Density Spectrum (PDS) of *x*, *s*, *n*, respectively. Following equation (4) by estimating the PDF of the random noise signal, the PDF of the clean speech signal can be estimated as expressed below:

$$\hat{}_s(\ ) \quad _x(\ ) \quad \hat{}_n(\ ) \tag{5}$$

Where $\hat{}_s(\ )$, $\hat{}_n(\ )$ are Estimations of the $_s(\ )$, $_n(\ )$ and 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:

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

$$R_n(\ ;m) \quad D_0\ (\ ) \tag{7}$$

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

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:

$$_s(\ ;m) \quad _x(\ ;m) \quad _n(\ ;m) \tag{9}$$

Also we know that [1]:

$$_x(\ ;m) \quad \frac{X(\ ;m)X^*(\ ;m)}{\mathrm{N}^2} \quad \frac{|X(\ ;m)|^2}{\mathrm{N}^2} \tag{10}$$

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

$$_x(\ ;m) \quad |x(\ ;m)|^2 \tag{11}$$

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

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

Where $|X(\ ;m)|$ is the magnitude of Fourier transform for the windowed *x(n)*; $|S(\ ;m)|$ and $|N(\ ;m)|$ are the magnitude of the Fourier transform for the windowed clean speech signal and windowed noise signal, respectively.
As can be seen from the equation (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].

$$S(\ ;m) \quad X(\ ;m) \tag{13}$$

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:

$$S(\ ;m)\quad \left|S(\ ;m)\right|\exp\ i\ _{S(\ ;m)} \qquad \left|X(\ ;m)\right|^2\quad \left|N(\ ;m)\right|^2\ \ ^{1/2}\exp\ i\ _{S(\ ;m)} \tag{14}$$

Where $S(\ ;m)$, $N(\ ;m)$ are the Fourier transform of the estimated clean signal and Fourier transform of the estimated noise signal, respectively.

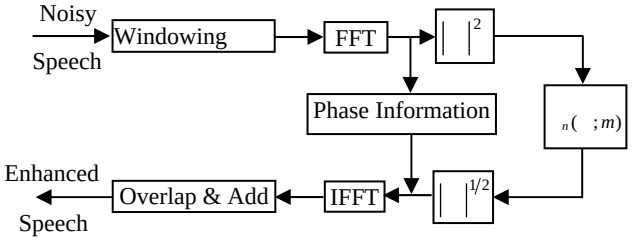

Fig1. diagram block of Power Spectral Subtraction (PSS)

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.

$$S(\ ;m)\quad \left|X(\ ;m)\right|^a\quad \left|N(\ ;m)\right|^a\ \ ^{1/a}\exp\ i\ _{S(\ ;m)} \tag{15}$$

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 :

$$\left|S(\ ;m)\right|\quad \begin{array}{ll}\left|S(\ ;m)\right| & \textit{if}\ \left|S(\ ;m)\right|\quad 0\\ 0 & \textit{elsewhere}\end{array} \tag{16}$$

b) Full-wave correction :

$$\left|S(\quad;m)\right| \quad abc \quad \left|S(\quad;m)\right| \tag{17}$$

III. 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 un-correlation.

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

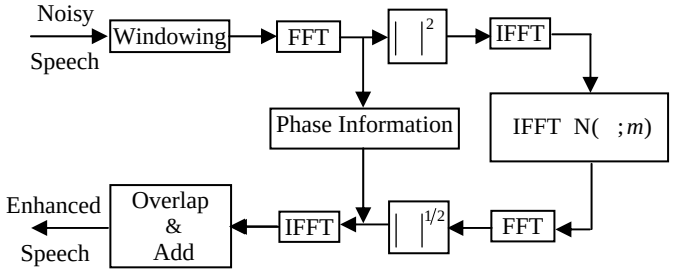

Fig3. diagram block of Inverse Fourier Subtraction

V. 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:

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

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):

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

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

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

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

A: 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]:

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

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

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

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

$$E(z) \quad S(z) \; 1 \quad \sum_{k=1}^{p} a_k z^{k} \quad A(z)S(z) \tag{29}$$

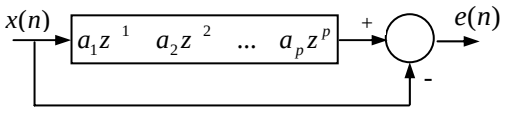

Fig4. diagram block of calculation of error function

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).

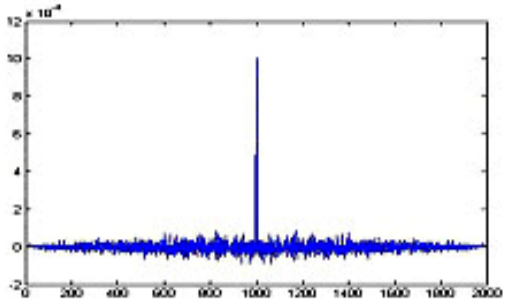

Fig5. auto-correlation function of the error signal.

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.

B: Improving the output of filter A(z) using VAD, V/UV detector algorithms

As mentioned above, in the spectral subtraction and inverse spectral subtraction it's assumed that the noise effect is the same for all of the signal range [5, 6, 7, 8]. But, in practice, this situation sanely aaaa1 happens; Since in addition to the existence of different noise source, there is another fact that is the effect off noise on the speech signal depends on the frequency. This depending leads on the frequency [7]. This dependence leads us to the act that the effect of noise on voice and unvoice signals is not the same. So, we are trying to find a method that by separating the voice and unvoice frames, get a higher accuracy in studying this different effect of noise.
Also, since in our proposed methods, we want to use *LPC* analysis for noise estimation, it's important to know that the estimated noise in the voice frames is nearer to the actual noise compared with the unvoice frames. Fig.6 shows the representation of the amount of this error for voice and unvoice signals versus the number of the filter poles *H(z)*.

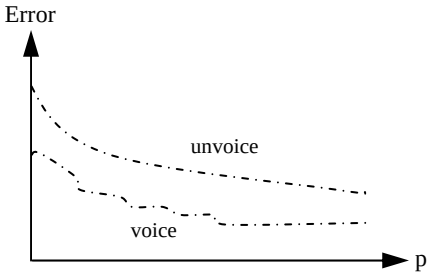

Fig6.

The other fact is that a silence frame only consists of noise. When this frame passes through the filter A(z), it becomes weaker so it must be multiplied by the gain of an amplifier and for reducing the amount of error in unvoice signals, the estimated noise should the amount of error in unvoice signals, the estimated noise should be weakened.

$$N(\quad ;m) \qquad N(\quad ;m) \tag{30}$$

According to the discussions above, is different for the voice and unvoice frames.

VI. Inverse Fourier Subtraction using LPC, VAD, and V/UV Detector Analysis (LPIFSS)

As can be seen in fig.2, in the inverse Fourier spectral subtraction method, like the spectral subtraction [5,6,7,8], we need an estimation of noise that uses the silence part of the corrupted speech signal. Usually the first frame of the speech signal is considered as the silence part of the signal. In this method it's supposed that:

1- First frame of the corrupted signal with noise belong to the signal.
2- The effect of noise should be the same in all the signal range.

To improve the method above, it's suggested that instead of directly applying the estimated noise of the silence part of the signal to the algorithm, it's better to pass it through A(z) first, and then consider the output of the filter as noise an applying it to the inverse Fourier spectral subtraction algorithm, Because the output of this filter is nearer to noise rather than first estimation of noise. On the other hand (also), for solving problem of the changing effect of noise in the corrupted speech signal rang, we calculate the estimated noise of the each frame from the nearest silence frame.

To have, the most likelihood between the estimated on the actual noise, it's recommended to average the estimated Fourier transform:

$$\left|\overline{N}(\quad;m)\right| \quad \frac{1}{m}\sum_{1}^{m} N(\quad;m) \tag{34}$$

$$\begin{aligned} s(n;m) \quad & IFFT\{\{FFT[IFFT\{|X(\quad;m)|^{a}\} \\ & IFFT\{|\overline{N}(\quad;m)|\}^{a}]^{1/a}\}\exp\{i\underset{s(\quad;m_{kj})}{\quad}\}\} \end{aligned} \tag{35}$$

VI. Experimental Result

In this part, want to compare the propose methods with the previous ones. For this reason, first in chart 1 the *PSS*, *GSS* and *LPSS* methods have been compared with each other *for SNR= 0, 5, 10 db* (noise of the white Gaussian noise type) to show the ability of the propose noise estimation method for improving the *SNR* of the speech signals, in all the *SNR* rang, from low to good.

Chart1, show the power of the LPSS method in noise reduction. As mentioned earlier in the previous methods, usually the first frame is used as the silence part; the weakness of this method becomes obvious when the first frame is not silence. Chart2 compares the LPSS method with PSS and GSS methods for speech signals from the *TIMIT* database and shows the priority of proposed noise estimation method for these kinds of signals.

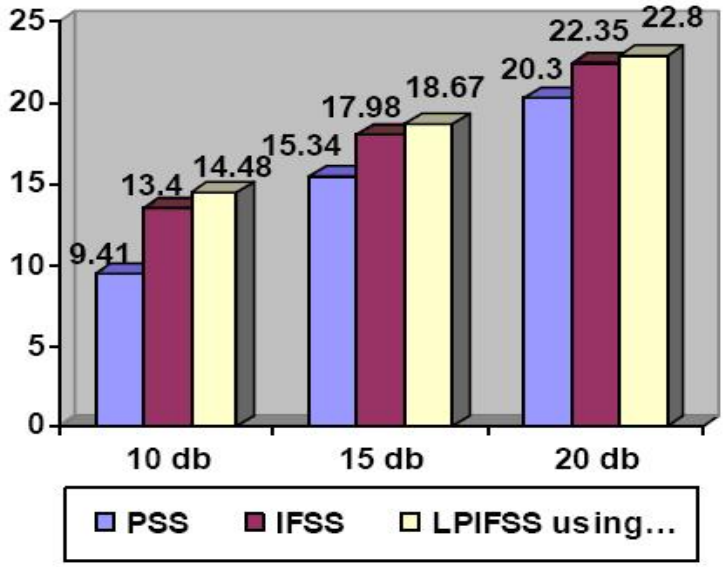

Chart1. Comparison of the *PSS*, *IFSS* and *LPIFSS* with *SNR 0, 5, 10 dB* (noise is *WGN* type)

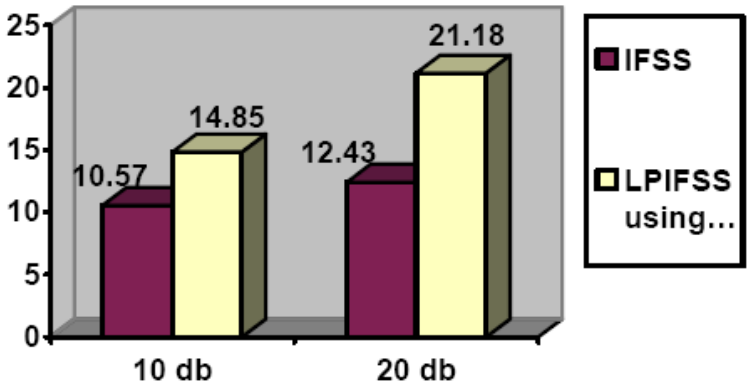

Chart2, compare LPSS method with IFSS methods when the first frame is not silent

Next, we have applied, *PSS*, IF*SS* and *LPIFSS* on (to) 50 corrupted speech signal (noise of the *WGN* type) from the *TIMIT* database with the initial *SNR* of 10 db and the average output *SNRs* have been showed in chart (3). As can be seen, applying the proposed noise estimation method on each of the methods, results in an enhanced *SNR* of output.

For a deeper study on the proposed methods and comparing them with the previous ones, the clean and corrupted speech signals with *SNR = 10 db* in cooperation with their improved signals using the *GSS*, *IFSS*, *LPSS* and *LPIFSS* methods have been shown in fig (8) and their spectrum have been plotted in fig (9).

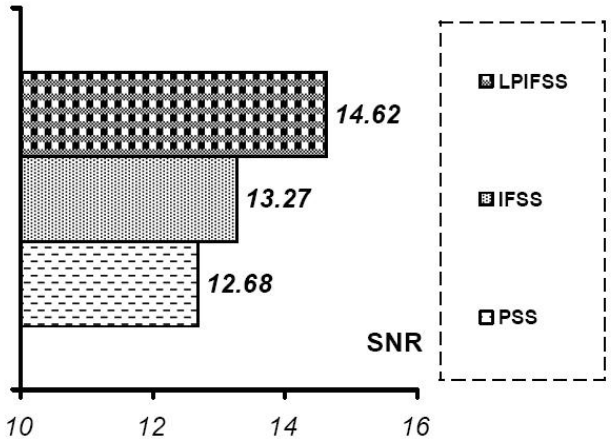

Chart3. compare of the output *SNR* in proposed and existing methods with the initial *SNR* of 10 db

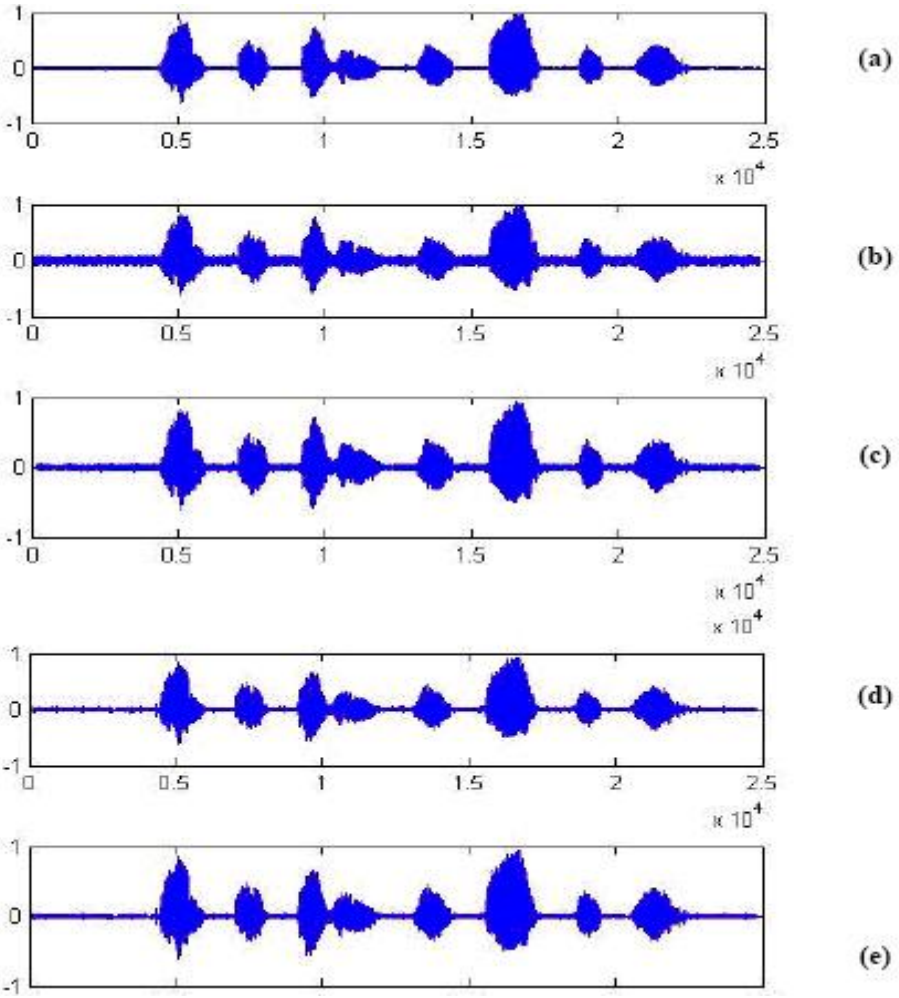

Fig8. enhancement results for speech corrupted by WGN a) Clean speech signal b) Noisy Speech (SNR=10) c) enhancement speech by PSS (SNR=12.68) d) enhance speech by IFFS (SNR=13.27db) e) enhanced speech by LPIFSS (SNR = 14.62)

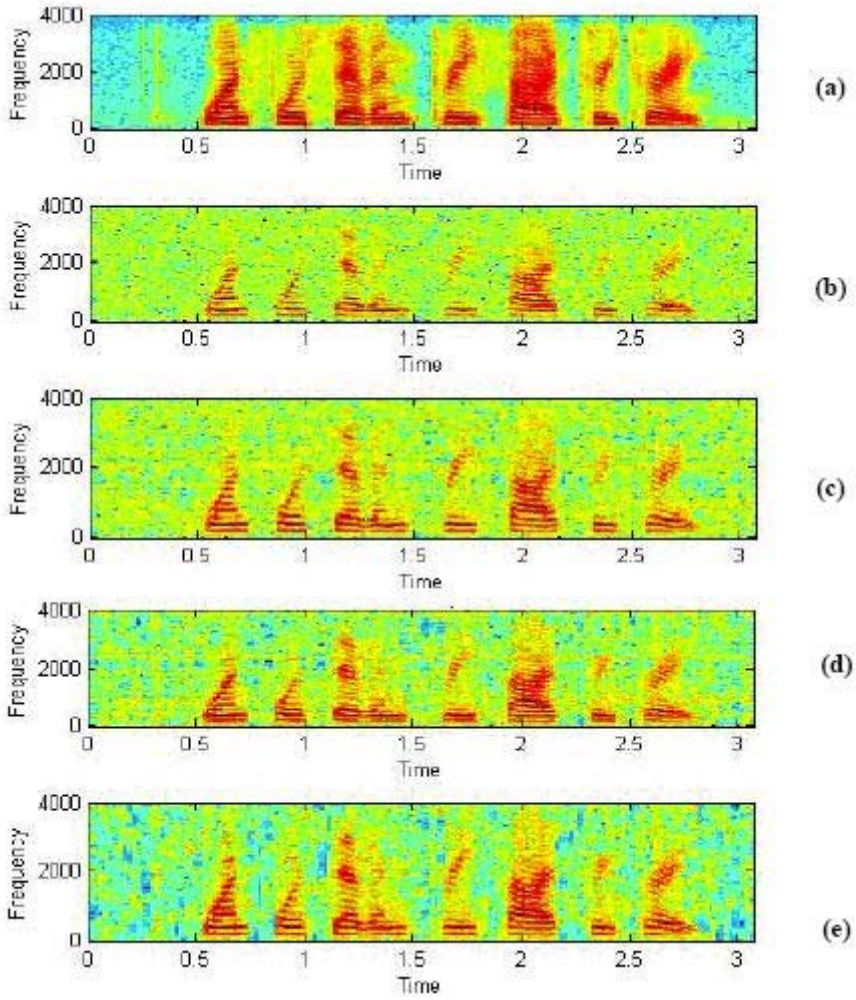

Fig9. Spectrogram for speech corrupted by WGN a) Clean Speech signal b) Noisy Speech (SNR=10) c) Enhanced speech by PSS (SNR=12.68) d) enhance speech by IFSS (SNR 13.27) e) enhanced speech by LPIFSS (SNR =14.62)

VII. Mean Opinion Score (MOS) Auditory Test

Up to now (has been), the *SNR* of the enhanced speech signal used for the comparison between the proposed methods and the previous ones. Now we wanted to compare them qualitatively and thus we use the auditory test [14].

We have applied 10 speech signals from the *TIMITS* database with the initial *SNR* of 10 db, on the conventional and proposed algorithm, supposing that the noise is of WGN type, and have asked 6 people (3 women and 3 men in a wide age rang) To score the enhanced signals. The results are shown in table 1 [14]. The average of their scores for these 10 speech signals (60 tries for each initial) *SNR* as shown in chart 4.

Table1. MOS Auditory Test, Five-point scales for quality and impairment, and associated scores [16]

Score	Impairment
5(Excellent)	Imperceptible
4 (Good)	(Just) Perceptible but not annoying
3 (Fair)	Perceptible and slightly annoying
2 (Poor)	Annoying but not objectionable
1 (Bad)	Very annoying (Objectionable)

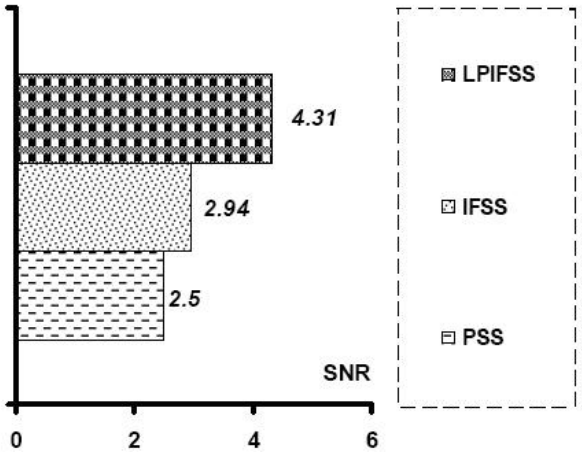

Chart 4. compare proposed and existing methods by MOS test

VIII. 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 its 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.

Next, we tried to improve this estimated noise and we have used VAD and V/UV detector algorithms. After improving the previous methods, we presented a method for speech signals that doesn't need and estimation of noise or it's parameters.

By comparing the proposed and the exiting methods, we have seen that the proposed methods improved the *SNR* of the enhanced signals as well as showing a better response in the MOS test.

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: April 1, 2008