# Spectral Subtraction for Speech Enhancement and Compression Using LPC

Dr.D.Ambika*, A.Deepa, P.Sumathi
Assistant Professor,Dept of MCA,
V.L.B.Janakiammal College of Arts & Science,Coimbatore
Email* :ambi.ambika123@gmail.com

**Abstract— Pre-Processing of Speech Signal is very crucial in the applications where silence or background noise is completely undesirable. The degradation of speech due to the presence of noise causes severe difficulties in various communication environments. Therefore, in this paper it is needed to first apply a noise removal and silence removal methods, in order to detect" clean" speech segments. Then using that clean speech the compression process was performed using the most powerful compression technique such as Linear Predictive Coding (LPC). Here different samples of spoken words are collected from different speakers and are used for implementation. The samples which is denoised, silence removed and compressed were compared with the samples which is not denoised but silence removed and compressed using LPC. Finally the results were evaluated in terms of compressed ratio (CR), Peak signal-to-noise ratio (PSNR) and Normalized root-mean square error (NRMSE).Finally the result show that the samples which is denoised, silence removed and compressed gives better result than the other samples.**

**Keywords-** Speech Enhancement, Speech Compression, CR, DWT, PSNR, NRMSE

## I. INTRODUCTION

Pre-processing of Speech Signal is used in any speech processing applications such as Noise Removal, Endpoint Detection, Pre-emphasis, Framing, Windowing, Echo Canceling etc. A speech signal contains two main parts, such as: one carries the speech information, and the second section includes silent or noise which is present between the utterances, without any verbal information. So therefore extracting the valuable information from a mixture of conflicting information is essential in any signal processing application. The presence of background noise affects the intelligibility and the quality of the signal. So without disturbing the speech signal, the perceptual quality and intelligibility of a speech can be improved by reducing the background noise and the silent section. Many applications like mobile communications, teleconferencing, speech and speaker recognition systems need more effective noise reduction algorithms for enhanced performance. Hence, the signal has to be cleaned up with noise cancellation technique before it is stored, analyzed, transmitted, or processed. The speech signal contains various parts such as silent zone part, voiced part and unvoiced part. In which voiced part includes the maximum speech information and the unvoiced part contains the whispering sounds [1]. But the silent zone does not contain any speech information. These zones will be present in all speech signals of about 30% to 40% of the whole speech. By removing these zones from the speech signal results in no loss of intelligibility but only the naturalness reduces as the speech becomes continuous without any break in between. Processing these parts leads to significant increase in the computational complexity and these can be avoided by removing such zones from the original speech signal. Compression of signal to lower rates with good speech quality not only eliminates the redundancy issue but also provides a lower bandwidth signal, which solves multiple problems in communication and multimedia applications. Commercial systems that rely on efficient speech coding include cellular communication, voice over internet protocol (VOIP), videoconferencing, electronic toys, archiving, and digital simultaneous voice and data (DSVD), as well as numerous PC-based games and multimedia applications.The compression algorithm is used to reduce the bandwidth requirement and also it is used to provide a level of security for the data being transmitted. It is more important in teleconferencing and wireless communication. The main aim of this research work is to perform denoising, silence removal and compression with the speech signals and it is compared with the samples which are not denoised, but silence removed and compressed. The results with their performances are analyzed subjectively to find out the sample which gives better compression ratio. The paper is organized as follows. Section 2 presents the need for speech enhancement, section 3 deals with the noise removal using spectral subtraction, section 4 explains the silence removal, section 5 investigates the Linear Predictive Coding (LPC),section 6 deals with the performance evaluation and finally, the conclusion is summarized in section 7.

## II. NEED FOR SPEECH ENHANCEMENT

Speech enhancement is engaged in many applications like speech recognition systems, enhancement of signals in telecommunications, military, teleconferencing, and cellular environments etc. It is one of the most important topics in speech signal processing to improve the performance of the systems in noisy conditions. Many techniques are available for this purpose namely spectral subtraction, extended and iterative wiener filtering, adaptive filtering, kalman filtering, fuzzy algorithms, HMM-based algorithms, signal subspace approach and kernel based approaches etc. Although there are many techniques available, improvement in the (SNR) signal-to-noise ratio is the target of most techniques and also their performances depend on the quality and intelligibility of the processed speech signal. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility and degree of listener fatigue. By improving quality and intelligibility of speech signals it reduces listener's fatigue; improve the performance of hearing aids, cockpit communication, videoconferencing, speech coders and many other speech systems.

Silence removal is also a known techniques adopted for many years which is used for dimensionality reduction in speech that facilitates the system to be computationally more efficient. This type of classification of speech into voiced or silence/unvoiced [2] sounds finds other applications mainly in Fundamental Frequency Estimation, Formant Extraction or Syllable Marking, Stop Consonant Identification and End Point Detection for isolated utterances. The block diagram of the proposed work is shown is figure1.

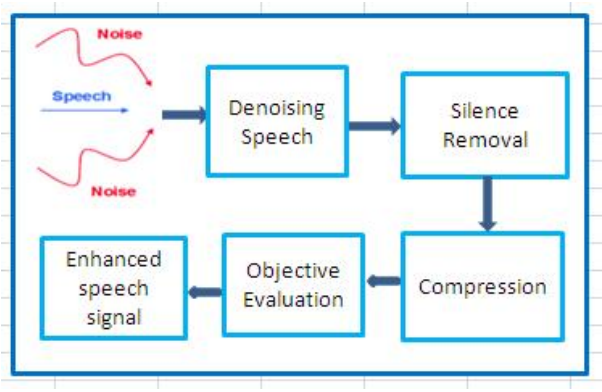


Figure 1.Block diagram of the proposed work

## III. NOISE REMOVAL USING SPECTRAL SUBTRACTION

There have been a lot of research involving speech enhancement that are specifically designed to recover the clean speech. The spectral subtraction algorithm is historically one of the first algorithms proposed for noise reduction [5], and it is based on a simple principle. The subtraction process needs to be done carefully to avoid any speech distortion. If too much is subtracted, then some speech information might be removed, while if too little is subtracted then much of the interfering noise remains. The basic assumption of this algorithm is that the noise is additive and its spectrum does not change with time. This means noise is stationary or it is slowly time varying signal whose spectrum does not change significantly between the updating periods [6]. The noise spectrum can be estimated, and updated; during the periods when the signal is absent or when only noise is present [7]. The goal of spectral subtraction is to suppress the noise from the degraded signal. It can be represented as

$$y(n) = x(n) + d(n) \quad (1)$$

Where y(n) is the noisy speech which is composed of the clean speech signal s(n) and the additive noise signal d(n). It operates by making an estimate of the spectral magnitude during periods of no speech and subtracting this spectral estimate of the noise from the subsequent speech spectral magnitude.

These algorithms attempt to be an omnipresent solution for all types of noise environments. However, the serious drawback of this method is that the enhanced speech is accompanied by unpleasant musical noise artifact which is characterized by tones with random frequencies. Although many solutions have been proposed to reduce this musical noise, results performed with these algorithms illustrate that there is a need for further improvement Taking the short-time Fourier transform of y(n), we get

$$Y(\omega_k) = X(\omega_k) + D(\omega_k) \quad (2)$$

for $\omega_k = 2\pi k/N$ and k = 0,1,2, . . . ,N - 1, where N is the frame length in samples. The enhanced signal is obtained by computing the inverse discrete Fourier transform of the estimated signal spectrum using the phase of the noisy signal. The derivation of the spectral subtraction equations is based on the assumption that the cross terms involving the phase difference between the clean and noise signals are zero. The cross terms are assumed

to be zero because the speech signal is uncorrelated with the interfering noise. While this assumption is generally valid since the speech signal and noise are statistically uncorrelated, it does not hold when applying the spectral subtraction algorithm over short-time (20–30 ms) intervals. Consequently, the resulting equations derived from spectral subtraction are not exact but approximations. The basic block diagram of spectral subtraction method is shown in the "Fig 2".

## IV. SILENCE REMOVAL

There are several ways of classifying events in speech such as silence (S), where no speech is produced. And in unvoiced (U) part, the vocal cords [4] do not vibrate, so the resulting speech waveform is a periodic or random in nature and where in the silence part no speech is produced. Unvoiced speech sections are generated by forcing air through a constriction formed at a point in the vocal tract (usually toward the mouth end), thus producing turbulence and voiced (V), in which the vocal chords are tensed and therefore vibrate periodically when air flows from the lungs, so the resulting waveform is quasi-periodic. Voiced speech consists mainly of vowel sounds. It is produced by forcing air through the glottis, proper adjustment of the tension of the vocal cords results in opening and closing of the cords, and a production of almost periodic pulses of air. These pulses excite the vocal tract. The speech segments are detected based on the signal energy and spectral centroid. As the signal energy and spectral centroid feature sequences are computed, a simple threshold-based algorithm is applied, in order to extract the speech segments. In general, the following steps are executed:

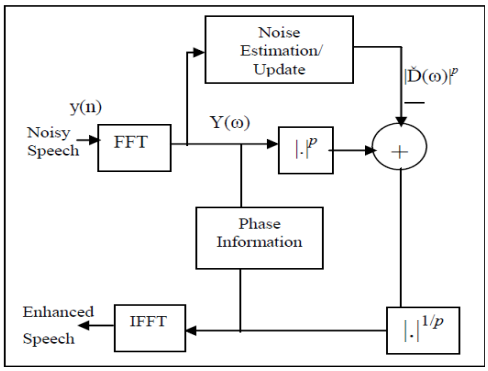


Figure 2 Basic block diagram of Spectral Subtraction Method

1) Two feature sequences are extracted from the whole audio signal.

2) For each sequence two thresholds are dynamically estimated.

3) A simple thresholding criterion is applied on the sequences.

4) Speech segments are detected based on the above criterion and finally a simple post-processing stage is applied.

In the Signal Energy, let $x_i(n)$, $n = 1\ldots\ldots N$ the audio samples of the $i$ - th frame, of length $N$. Then, for each frame $i$ the energy is calculated according to the equation:

$$E(i) = \frac{1}{N}\sum_{n=1}^{N}|x_i(n)|^2 \quad (3)$$

This simple feature can be used for detecting silent periods in audio signals, but also for discriminating between audio classes. Similarly in the Spectral centroid, $C_i$, of the $i$-th frame is defined as the center of "gravity" of its spectrum which is

This simple feature can be used for detecting silent periods in audio signals, but also for discriminating between audio classes. Similarly in the Spectral centroid, $C_i$, of the $i$-th frame is defined as the center of "gravity" of its spectrum which is

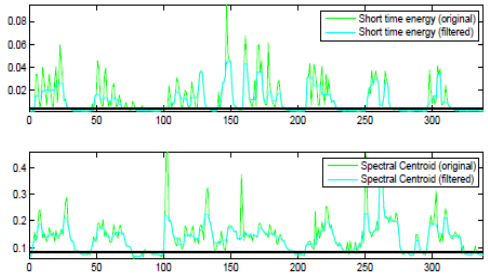


Figure 3 Sequence of the Signal's Energy and the Spectral Centroid

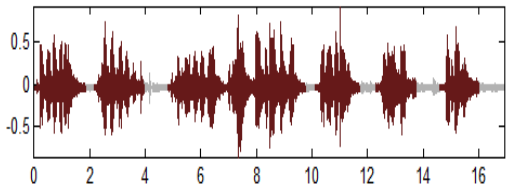


Figure 4 Sample whole speech signals in which Red color represents the Detected Voiced Segments

$$C_i = \frac{\sum_{k=1}^{N}(k+1)X_i(k)}{\sum_{k=1}^{n} X_i(k)} \quad (4)$$

*Xi*(*k*), *k* = 1. . . . . *N*, is the Discrete Fourier Transform (DFT) coefficients of the *i*-th short-term frame, where *N* is the frame length. This feature is a measure of the spectral position, with high values corresponding to "brighter" sounds. In order to extract the feature sequences, the signal is first broken into non-overlapping short-term-windows (frames) of 50 mseconds length. Then for each frame, the two features are calculated, leading to two feature sequences for the whole audio signal. Towards this end, the following process is carried out, for each feature sequence:

1) Compute the histogram of the feature sequence's values.

2) Apply a smoothing filter on the histogram.

3) Detect the histogram's local maxima

4) Let M1 and M2 be the positions of the first and second local maxima respectively. The threshold value is computed using the following equation:

$$T = \frac{W.M_1 + M_2}{W + 1} \quad (5)$$

where W is a user-defined parameter. Large values of W lead to threshold values closer to *M*1. The above process is executed for both feature sequences, leading to two thresholds: *T*1 and *T*2, based on the energy sequence and the spectral centroid sequence respectively. As long as the two thresholds have been estimated, the two feature sequences are thresholded, and the segments are formed by successive frames for which the respective feature values for both feature sequences are larger than the computed thresholds. The "fig 3" depicts the sequence of the signal energy and the spectral centroid. The "fig 4" depicts the sample speech sample in which the red color represents the detected voiced segment whereas the grey color depicts the silent region. In this paper only the voiced part (i.e. red ) part is taken thereby the silence can be removed.

## V. LINEAR PREDICTIVE CODING (LPC)

Linear Predictive Coding (LPC) is used to compress the speech signal without losing its audibility. By breaking the speech signal into segments it sends the voiced or unvoiced information, the pitch period and the coefficients for the filter. LPC extracts speech parameters like pitch formants and spectra and it is one of the most powerful methods used in audio and speech signal processing. The principle behind the use of LPC is to minimize the sum of the squared differences between the original speech and estimated speech signal over a finite duration . It allows encoding of good quality speech at a low bit rate and it also provides extremely accurate estimates of speech parameters. The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples.The predictor coefficients are represented by ak and it is normally estimated in every frame with size of 20 ms long. Another important parameter is the gain (G). The transfer function of the time varying digital filter is given by:

$$H(z) = \frac{G}{1 - \sum a^k z^{-k}} \quad (6)$$

The summation is computed starting at k=1 up to p, which will be 10 for the LPC-10 algorithm. The performance of the various samples which is denoised, silence removed and compressed used for the datasets are illustrated in "fig 13".

## VI. PERFORMANCE EVALUATION

Here many speech samples were taken, where some of the signals are denoised, silence removed and at last compression is performed, whereas for some samples only the silence removal and the compression are performed. Here both the sample values are taken and their performance of compression ratio is analyzed in order to find which sample produces best compression ratio. Similarly another set of samples were denoised, silence removed and compressed were compared with the samples which is not silence removed but denoising and compression is done. Here all the samples were compared in order to find which kind of samples produces best compression ratio. There are two ways in which the techniques can be evaluated such as objectively and

subjectively. Objective analysis is done by evaluating the performance of parameters such as Compression Ratio (CR), Peak Signal to Noise ratio (PSNR), and Normalized Root Mean Square Error Rate (NRMSE).Whereas subjective analysis is based on hearing the reconstructed signal and making the judgment which is done by Mean Opinion Score (MOS). For calculating the performance many speech samples are taken from various speakers and each file has a different size with respect to other files. In this paper the objective analysis is done in order to evaluate the parameters and the formulas are given below

*A. Compression Ratio*

$$CR = \frac{Length(x(n))}{Length(r(n))} \tag{7}$$

The compression ratio can be calculated using the above formula, where x(n) is the original signal and r(n) is the reconstructed signal

*B. Peak Signal to Noise Ratio (PSNR)*

$$PSNR = 10\log_{10}\frac{(NX^2)}{\|X - r\|^2} \tag{8}$$

The PSNR can be calculated using the above formula, where N is length of the reconstructed signal, X is the maximum absolute square value of the signal x and $\| x\text{-}r \|^2$ is the energy of the difference between the original and reconstructed signals

*C. Normalized Root Mean Square Error (NRMSE)*

$$NRMSE = \sqrt{\frac{(x(n) - r(n))^2}{(x(n) - \mu x(n))^2}} \tag{9}$$

The NRMSE can be calculated using the above formula, where x (n) is the speech signal, r(n) is the reconstructed signal and $\mu_x(n)$ is the mean of the speech signal. Based on the performance evaluation, the sample which is denoised, silence removed and compressed produces best result than the other samples. It achieves higher PSNR and lower NRMSE than the other samples. In the Compression ratio also these samples produces good quality than the other method. The table1 shows the Performance Analysis for the DWT and LPC based on PSNR, NRMSE and CR. The following "fig 10, 11, 12, and 13" represents the performance of two techniques in the graphical form.

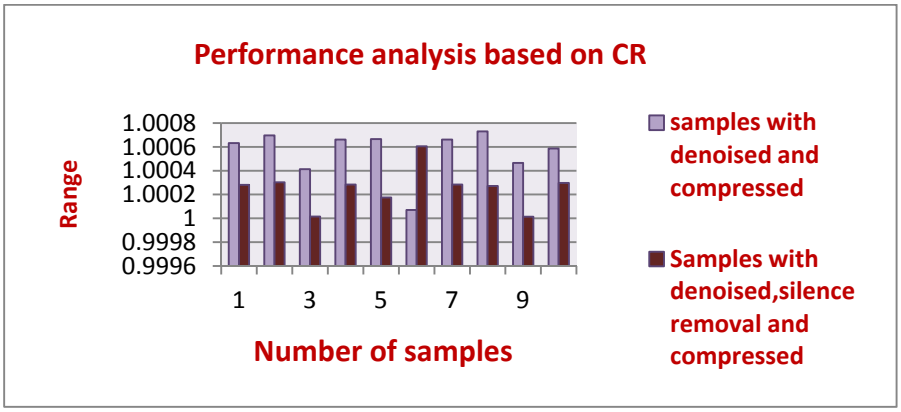


Figure 10 Performance evaluation based on CR

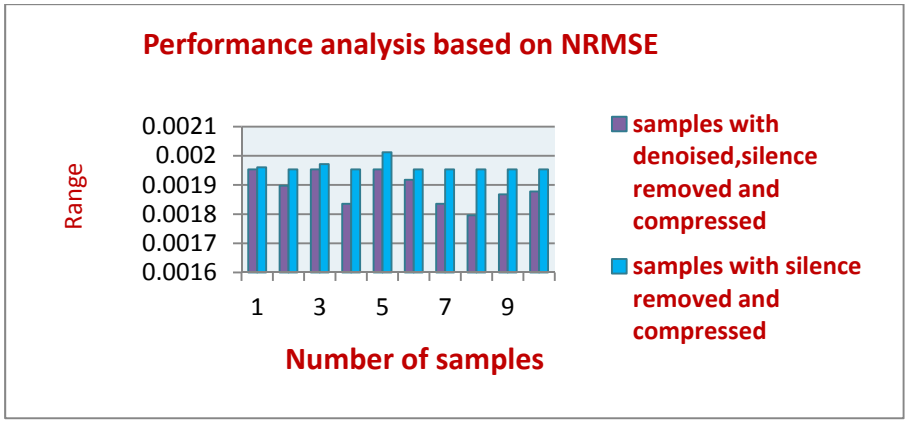


Figure 11 Performance evaluation based on NRMSE

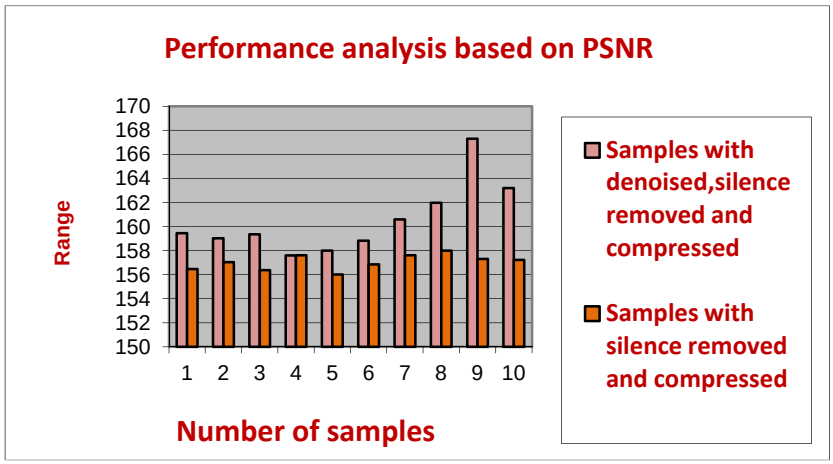


Figure 12 Performance evaluation based on PSNR

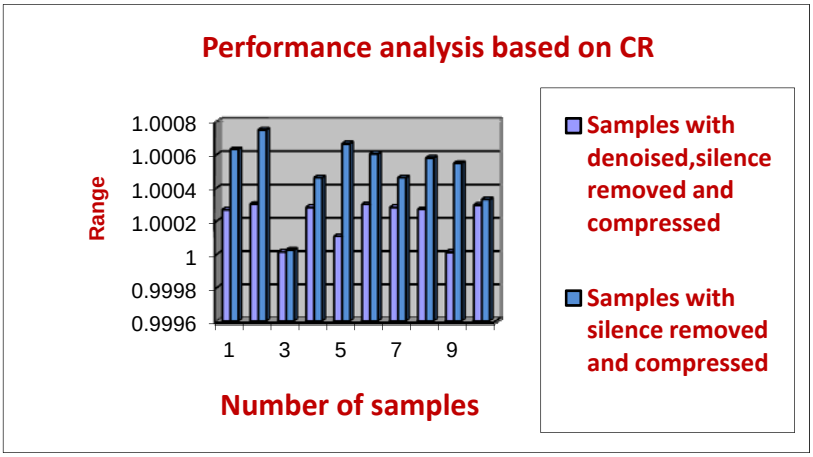


Figure 13 Performance evaluation based on CR

## VII. CONCLUSION

Data compression is the technology of representing information with lowest number of bits. To achieve an optimum clean speech which is affected by silence and noise, the spectral subtraction algorithm is used to remove noise and compression is done using LPC. In this paper as observed from different samples it is clear that the sample which is denoised, silence removed and compressed gives the better result. Similarly a good reconstructed signal should produce high PSNR and low NRMSE which means the signal have low error and high reliability. By analyzing these samples PSNR and NRMSE values, the samples which is denoised, silence removed and compressed gives the better result.

## REFERENCES


[1] Jan Vanus, "The use of the adaptive noise cancellation for voice Communication with the control system", International Journal of Computer Science and Applications, Technomathematics Research Foundation, Vol. 8, No. 1, pp. 54 – 70, 2011.

[2] A.Martin, D.Charlet, and L.Mauuary, "Robust speech / non-speech detection using LDA applied to MFCC", in Proc.ICASSP2001,vol.1,2001,pp.237-240.

[3] L.R. Rabiner and B. H. Juang, "Fundamentals of speech recognition," 1st Indian Reprint, Pearson Education.

[4] G. Saha, Sandipan Chakroborty, Suman Senapati, " A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications"

[5] Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process. ASSP-27 (2),113–120.

[6] Anuradha R. Fukane, Shashikant,L. Sahare, "Different Approaches of "Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments", International Journal of Scientific & Engineering Research, Volume 2, Issue 5, May-2011,ISSN 2229-5518.

[7] Nicholas W. D. Evans, John S. Mason, Wei M. Liu and Benot Fauve, "On the fundamental limitations of spectral subtraction: An assessment by automatic speech recognition", School of Engineering, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP, UK.