DOI: 10.21917/ijct.2012.0089

EVALUATION OF SPEECH ENHANCEMENT IN NOISY CONDITIONS USING A SPECTRAL SUBTRACTION AND LINEAR PREDICTION COMBINATION

M. Thirumarai Chellapandi[1] and P. Kabilan[2]
Department of Computer Science, Madurai Kamaraj University College, India
E-mail: [1]messagethiru@yahoo.com and [2]spkabi@gmail.com

Abstract
Improving quality and intelligibility of speech signals in mobile devices has been studied with great interest in the past. Speech information in communication channels is usually corrupted by additive acoustic noise, reverberation or channel noise. This paper explores into the possibilities of enhancing corrupted speech using the Spectral Subtraction (SS) and Linear Prediction Coding (LPC) system for mobile applications only in acoustically noisy conditions. A Spectral Subtraction block is cascaded in series with a LPC system. For a pth order LPC system, a Levinson-Durbin based algorithm computes the LPC coefficients. Typically LPC is used as a data reduction system in speech communication but in this work, we try to find an optimum pth order LPC system that could enhance speech quality. We focus on improving speech quality and not speech intelligibility in this paper. The algorithm output will be evaluated objectively with a combined Perceptual Evaluation of Speech Quality (PESQ) and Itakura-Saito (IS) system and will be compared against Mean Opinion Scores (MOS) of various other Speech Enhancement algorithms.

Keywords:
Speech Enhancement, Linear Prediction Coding, Evaluation, Measurements

1. INTRODUCTION

Speech captured by microphones in cell phones or hearing aids are always corrupted by either additive noise or reverberation or both of them simultaneously. Hence, the signal of interest-speech needs to be cleaned off irrelevant contents that cause the speech corruption. However, removing the irrelevant information must not degrade the relevant information (speech). The objective of this paper is to enhance speech quality in order to reduce listener fatigue.

Noise is everywhere around. Even in places that we feel is quiet, will have a noise floor well below the full scale level. In streets, restaurants, theaters, airports, exhibitions, markets and shopping centers noise is prevalent. During conversations on mobiles or hearing aids, these noise contents contaminate the meaningful speech conversation between person A and person B. In concert halls, the direct sound (speech or music) is contaminated by the early and late reflections from the surrounding walls. This paper will explore a way to remove the additive noise from the signal recorded of a single microphone. We cascade a spectral subtraction based noise cancellation algorithm to a Linear prediction algorithm in series and evaluate the output with a PESQ-IS executable.

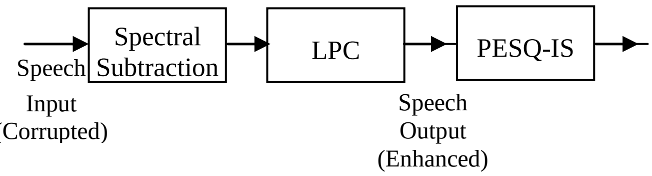

Fig.1. Paper block diagram: SS algorithm cascaded in series with LPC algorithm and Evaluation block

2. LITERATURE REVIEW: SPEECH ENHANCEMENT

From 1970's, several single microphone DSP strategies have been put forward in literature to cancel the noisy speech. Loizou [1] in his book gives a thorough review of these algorithms and broadly groups them into 4 categories:

- Spectral Subtraction algorithms [7], [8]
- Wiener filtering algorithms
- Statistical model-based algorithms
- Subspace algorithms.

The Spectral Subtraction algorithm attempts to estimate the background noise spectrum and subtract it from the noisy speech frequency spectrum. Wiener filtering algorithms searches for an optimum filter that minimizes the mean-squared error (MSE) between output and desired signal. Statistical model-based algorithms deploy statistical strategies to estimate and enhance the speech frequency spectrum. The subspace algorithms decompose the corrupted signal into signal and noise subspaces and subsequently nullify the noise subspace. Hu and Loizou [1] compared the performance of these different algorithms categories. Additionally, there are many more algorithm varieties. For example, in [5] the algorithm exploits the harmonic nature of speech components. There are dual to multi-microphone (microphone array) based active noise cancellation techniques too. In this work, we explore into noise cancellation methods for signal recorded of single microphone only.

3. WHY SPECTRAL SUBTRACTION?

In mobiles, the microphones are usually very close to the speaker's mouth but in case of hearing aids, the target speech is usually far away from microphones. Hence, algorithms that are meant to cancel noise for hearing aids must be able to deal with signals of very low Signal to noise Ratio's (SNR) as compared against algorithms for mobile devices. For mobile application, which is the project interest, it is enough that the algorithm is able to cancel noise for SNR 10 dB. In [3], it is shown that Spectral Subtraction algorithm performs much better at 10 dB

SNR when compared to other algorithms. Hence, we cascade Spectral Subtraction to an LPC system to examine further possibilities of speech enhancement at higher signal to noise ratios.

4. WHY LINEAR PREDICTION CODING?

Spectral Subtraction has been justified to be a suitable algorithm to meet the software requirements of this project. To improve the software performance further, we cascade a LPC system [9] with SS.

Formants are the spectral peaks of the sound spectrum of the voice. In speech science and phonetics, formant is also used to mean an acoustic resonance of the human vocal tract. Usually, there are four speech formants in spectral region 1-4 KHz. On a Z-transform plane, we need 4 conjugate pair poles (8 poles) to model the formants in the speech spectrum and at least 2 to 4 poles to model the spectral roll off in the high frequencies. Hence, in general, engineers choose an LPC order (p value) 10 to 12 for successful modeling of vocal resonance that varies in time because of the change in tract volume.

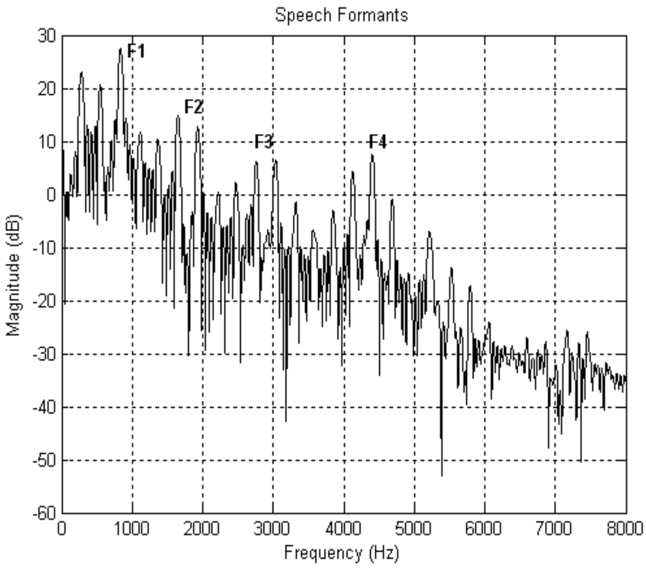

Fig.2. Speech formats la.wav Fs = 16 KHz. F1 to F4 show four formants in the low frequencies

The Input Speech is corrupted. It has spurious peaks in the spectrum caused by additive noise. By modeling the vocal tract response and applying it on corrupt signal, we can get rid of the spurious speech and obtain a speech signal that is closer to pure speech that is created by pumped excitation from lungs into the vocal tract. The vocal tract response varies from person to person and also in time. The tract response is unknown. The higher the LPC filter order, the system will model the spurious noise peaks as model response of the tract. Higher the filter order, the model response will closely follow the actual spectrum shape (Fig.3 and Fig.4). Therefore, we need to find an optimal pth order filter that will only model the formants and skip the spurious noise peaks and valleys in the spectrum. This approach is expected to enhance the speech quality.

Generally, LPC is used for data reduction applications in speech processing. In this work, we use it for enhancing the speech. The error signal generated by the subtraction of estimated signal from original is used directly to excite the resonance formants obtained by the LPC model. There is no need for codebook (CELP) excitation. A combination of mixing the error signal with original signal will also be examined as a source for exciting the resonance cavity response. There is no encoder and decoder in this system because LPC is just used for speech enhancement. A p^{th} order LPC filter will be optimally approximated that enhances the speech effectively.

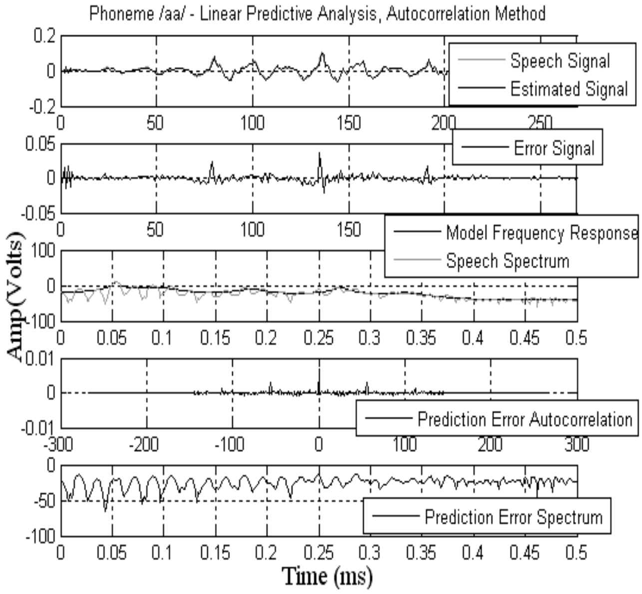

Fig.3. LPC Order 12

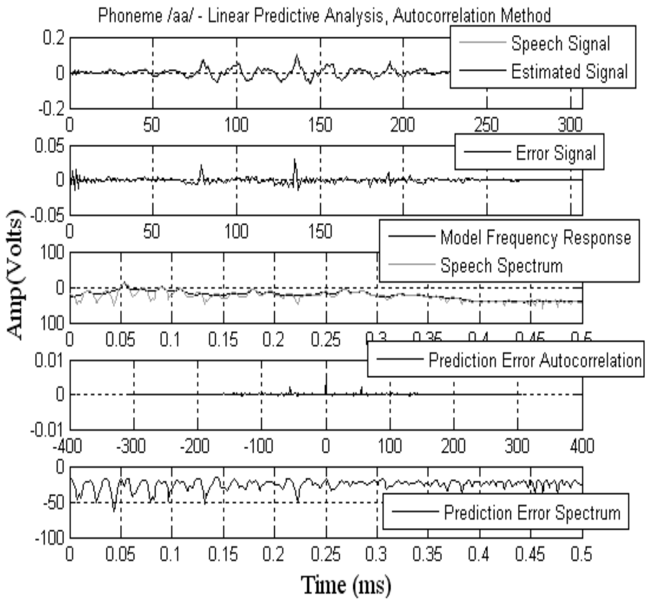

Fig.4. LPC Order 50

5. SPECTRAL SUBTRACTION (SS): ALGORITHM DETAILS

The spectral subtraction algorithm is used widely in speech enhancement [2]. A noise corrupted speech signal $y(n)$ is composed of clean speech signal $x(n)$ and noise $d(n)$. The noise and the clean speech are assumed to be independent and uncorrelated. The spectrum of the noise signal $D(\omega)$ obtained by Fourier transform is subtracted from corrupted speech spectrum to attain clean speech spectrum. The clean speech spectrum is reconstructed to a voltage signal back in the time domain signal using the inverse Fourier transform.

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

$$X(\omega) = Y(\omega) - D(\omega) \tag{2}$$

5.1 PROBLEMS IN SS: MUSICAL NOISE

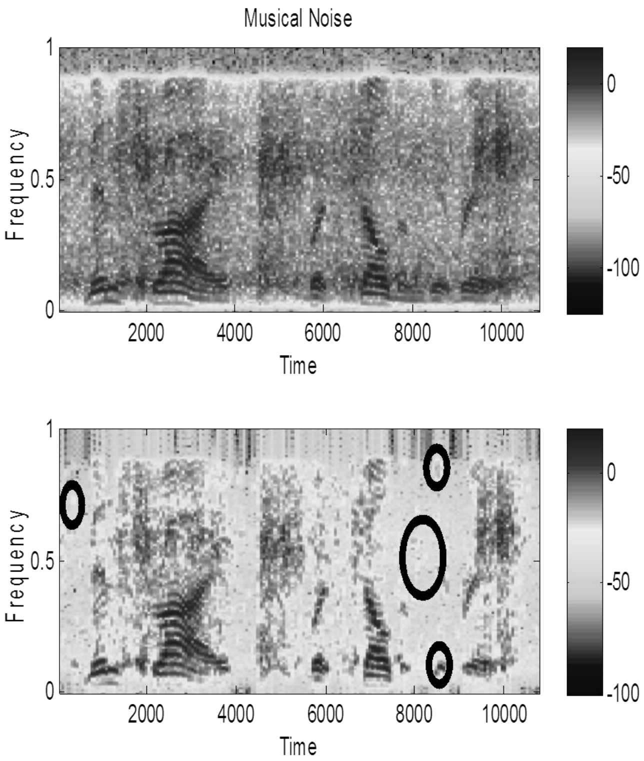

Fig.5. Isolated peaks shown in elliptical circles contribute to musical noise

The spectral subtraction by Boll in [2] uses a STFT (short-time Fourier transform) to calculate magnitude, subtract bias from the noise estimate, and does a half wave rectification to avoid negative magnitude spectrum. A voice activity detector (VAD) is also used to attenuate the noisy signal when speech is absent. Isolated peaks are created during the non-linear processing of negative values during half-wave rectification. After doing inverse Fourier transform, in re-synthesized time domain signal, these peaks sound similar to tones with frequencies that change randomly from frame to frame. This type of noise has a warbling sound along with a tone like quality, and is generally called as '*musical noise*'. Musical noises can be more annoying than the actual background noises like babble noise or restaurant noise.

Fig.5 shows a noisy spectrogram and the bottom picture shows a processed clean spectrogram with isolated spectral peaks in the spectrum that contribute to the musical noise phenomenon.

6. LINEAR PREDICTION CODING (LPC): ALGORITHM DETAILS

A simple LPC system is shown in Fig.6. Speech analysis and synthesis with Linear Predictive Coding (LPC) exploit the predictable nature of speech signals. Cross-correlation, autocorrelation, and auto covariance provide the mathematical tools to determine this predictability. If we know the autocorrelation of the speech sequence, we can use the Levinson-Durbin algorithm to and an efficient solution to the least mean-square modeling problem and use the solution to compress or re-synthesize the speech [4].

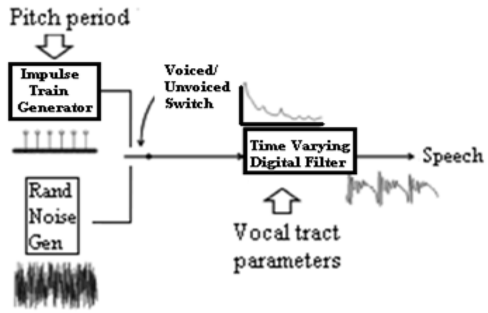

Fig.6. A simple LPC system

The linear prediction problem can be stated as finding the coefficients which result in the best prediction of the speech sample in terms of the past samples. Linear prediction models the human vocal tract as an infinite impulse response (IIR) system that produces the speech signal. For vowel sounds and other voiced regions of speech, which have a resonant structure and high degree of similarity over-time shifts that are multiples of their pitch period, this modeling produces an efficient representation of the sound [4]. The general linear system transfer function gives rise to three different types of linear model, dependent on the form of the transfer function H(z):

- When the numerator of the transfer function is constant, an all-pole or autoregressive (AR) model is defined.
- The all-zero or moving average model assumes that the denominator of the transfer function is a constant.
- The third and most general case is the mixed pole-zero model or autoregressive moving average (ARMA) model, where nothing is assumed about the transfer function.

The all-pole model for linear prediction is the most widely studied and implemented of these three approaches [4].

7. HUMAN SPEECH PRODUCTION, ANATOMY AND FUNCTION

The lungs initiate the speech process by acting as the bellows that expels air up into the other regions of the system. The air that leaves the lungs then enters into the remaining regions of the speech production system via the trachea.

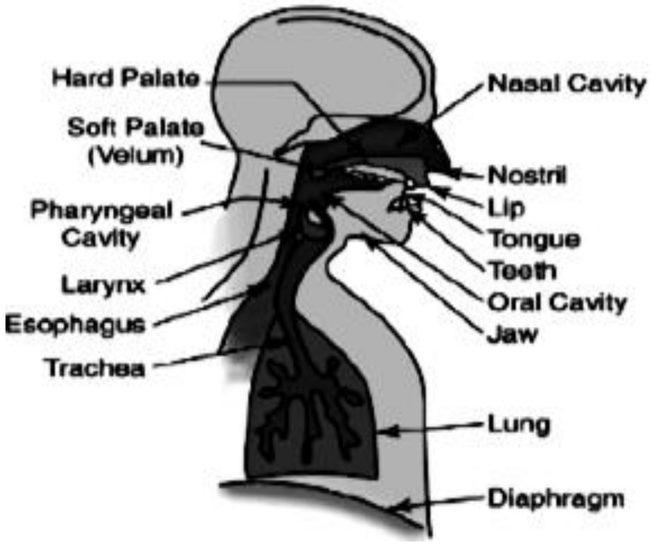

Fig.7. Human speech production system: Courtesy [4]

The turbulent air stream is driven up the trachea into the larynx. The larynx is a box-like apparatus that consists of muscles and cartilage. Two membranes, known as the vocal folds, span the structure, supported at the front by the thyroid cartilage and at the back by the Arytenoid cartilages. The arytenoids are attached to muscles which enable them to approximate and separate the vocal folds. The space between the vocal folds is called the glottis. A speech sound is classified as voiced or voiceless depending on the glottal behavior as air passes through it [4].

As air rushes through the glottis, the suction phenomenon known as the Bernoulli Effect is observed. This effect due to decreased pressure across the constriction aperture adducts the folds back together. The interplay between these forces results in vocal fold vibration, producing a voiced sound. This phonation has a fundamental frequency directly related to the frequency of vibration of the folds. During a voiceless speech sound, the glottis is kept open and the stream of air continues through the larynx without hindrance. The resulting glottal excitation waveform exhibits a flat frequency spectrum [4].

8. ALL POLE LINEAR PREDICTION MODEL

A linear prediction estimate at sample number n for the output signal y by a p^{th} order prediction filter can be given by,

$$\hat{y}(n) = \sum_{k=1}^{p} aky(n-k) \tag{3}$$

The error or residue between the output signal and its estimate at sample n can then be expressed as the difference between the two signals.

$$e(n) = y(n) - \hat{y}(n) \tag{4}$$

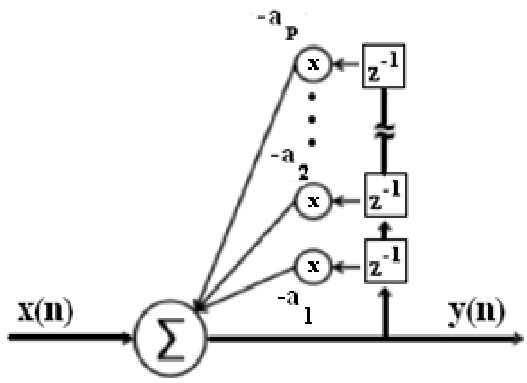

Fig.8. A graphical representation of an all pole linear system, where the output is a linear function of scaled previous outputs and the input

The total squared error for an as of yet unspecified range of signal samples was given by the following equation,

$$\begin{aligned} E &= \sum_n [e(n)]^2 \\ &= \sum_n [y(n) - \hat{y}(n)]^2 \\ &= \sum_n [y(n)]^2 - 2.y(n).\hat{y}(n) + [\hat{y}(n)]^2 \end{aligned} \tag{5}$$

This equation gives a value indicative of the energy in the error signal. Obviously, it was desired to choose the predictor coefficients so that the value of E was minimized over the unspecified interval. The optimal minimizing values can be determined through differential calculus, i.e. by obtaining the derivative of the above equation with respect to each predictor coefficient and setting that value equal to zero.

$$\frac{\partial E}{\partial a_k} = 0 \quad for\, 1 \le k \le p$$

$$\Rightarrow \frac{\partial}{\partial a_k}\left(\sum_n \left([y(n)]^2 - 2.y(n).\hat{y}(n) + [\hat{y}(n)]^2\right)\right) = 0$$

$$-2\sum_n y(n).\frac{\partial}{\partial a_k}\hat{y}(n) + 2\sum_n \hat{y}(n).\frac{\partial}{\partial a_k}\hat{y}(n) = 0$$

$$\sum_n y(n).\frac{\partial}{\partial a_k}\hat{y}(n) = \sum_n \hat{y}(n).\frac{\partial}{\partial a_k}\hat{y}(n)$$

$$-\frac{\partial}{\partial a_k}\hat{y}(n) = -\hat{y}(n-k)$$

$$\Rightarrow \sum_n y(n) - y(n-k) = \sum_n \hat{y}(n) - y(n-k)$$

$$-\sum_n y(n).y(n-k) = \sum_n\left(-\sum_{i=1}^{p} a_i y(n-i) - y(n-k)\right)$$

$$-\sum_n y(n).y(n-k) = \left(-\sum_{i=1}^{p} a_i \sum_n y(n-i) - y(n-k)\right) \tag{6}$$

For the sake of brevity and future utility, a correlation function f was defined. The expansion of this summation describes what will be called the correlation matrix [4].

$$\phi(i,k) = \sum_n y(n-i).y(n-k) \tag{7}$$

Substituting the correlation function into above equation allows it to be written more compactly,

$$-\phi(0,k)=\sum_{i=1}^{p} a_i \phi(i,k) \tag{8}$$

These derived set of equations are called the normal equations of linear prediction [4].

9. THE AUTOCORRELATION METHOD

The autocorrelation method of linear prediction minimizes the error signal over all time, from $-\infty$ to $+\infty$. When dealing with finite digital signals, the signal was windowed such that all samples outside the interval of interest are taken to be zero. If the signal was non-zero from 0 to $N-1$, then the resulting error signal will be non-zero from 0 to $N-1+p$. Thus, summing the total energy over this interval was mathematically equivalent to summing over all time [4].

$$\begin{aligned} E &= \sum_{n=-\infty}^{\infty} [e(n)]^2 \\ &= \sum_{n=0}^{N-1+p} [e(n)]^2 \end{aligned} \tag{9}$$

This form of the correlation function was simply the short-time autocorrelation function of the signal, evaluated with a lag of $(i-k)$ samples. This fact gives this method of solving the normal equations its name. The implication of this convenience was such that the correlation matrix defined by the normal equations exhibits a double-symmetry that can be exploited by a computer algorithm.

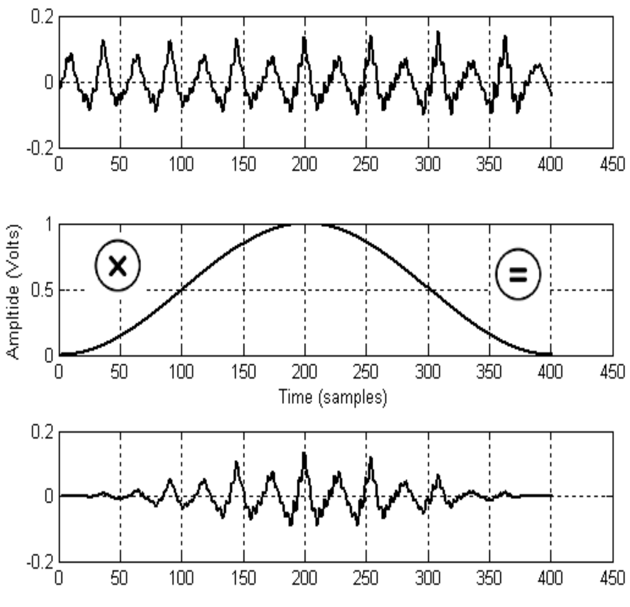

Fig.9. Autocorrelation of a time frame

Given that $a_{i,j}$ was the member of the correlation matrix on the i^{th} row and j^{th} column, the correlation matrix demonstrates,

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & . & . & . & a_{1,m} \\ a_{2,1} & a_{1,1} & a_{1,2} & . & . & . & a_{m,2} \\ a_{3,1} & a_{2,1} & a_{1,1} & . & . & . & a_{m,3} \\ . & . & . & a_{1,1} & . & . & . \\ . & . & . & . & a_{1,1} & . & . \\ . & . & . & . & . & a_{1,1} & . \\ a_{m,1} & a_{m,2} & a_{m,3} & . & . & . & a_{1,1} \end{bmatrix}$$

These redundancies mean that the normal equations can be solved using the Levinson-Durbin method, a recursive procedure that greatly reduces computational load [4].

10. LEVINSON-DURBIN METHOD

By exploiting the Toeplitz nature of the matrix of coefficients, several efficient recursive procedures have been devised for solving this system of equations. The well known of these methods are the Levinson and Robertson algorithms. Durbin's recursive algorithm followed earlier work of Levinson [2]. The following is the Durbin's recursive solution for autocorrelation equations,

$$\begin{gathered} E^{(0)} = R^{(0)} \\ k_i = \left(R(i) - \sum_{j=1}^{i=1} \alpha_j^{(i-1)} R(i-j) \right) \Big/ E^{(i-j)} \\ \alpha_i^{(i)} = k_i \\ \alpha_i^{(i)} = \alpha_i^{(i-1)} k_i \alpha_{i-j}^{(i-1)} \\ E(i) = \left(1-k_i^2\right) E^{(i-1)} \end{gathered} \tag{10}$$

These equations are recursively solved for i = 1, 2….p and the final solution was given as,

$$\alpha_j = \alpha_j^p \quad 1 \le j \le p \tag{11}$$

11. THE SPEECH ENHANCEMENT SYSTEM

The single channel Input corrupted speech is sampled at 8 KHz at 16 bit resolution per sample. A short-time Fourier transform (STFT) is performed on the signal with frame size of 20 milli seconds at 75% overlap rate to avoid spectral leakage. Before doing STFT, the section of the speech signal is multiplied with a hamming window. Spectral Subtraction is done on the current frame. The output of the SS algorithm for the current frame is passed onto LPC block for further enhancement of speech.

12. SPEECH QUALITY MEASUREMENTS

It is very essential to benchmark the software of interest in order to evaluate its performance based on sound quality. Speech Quality Measurements are of two types:

- Objective measurements
- Subjective measurements

Below is Fig.10 that vividly captures the categories of speech quality measurements.

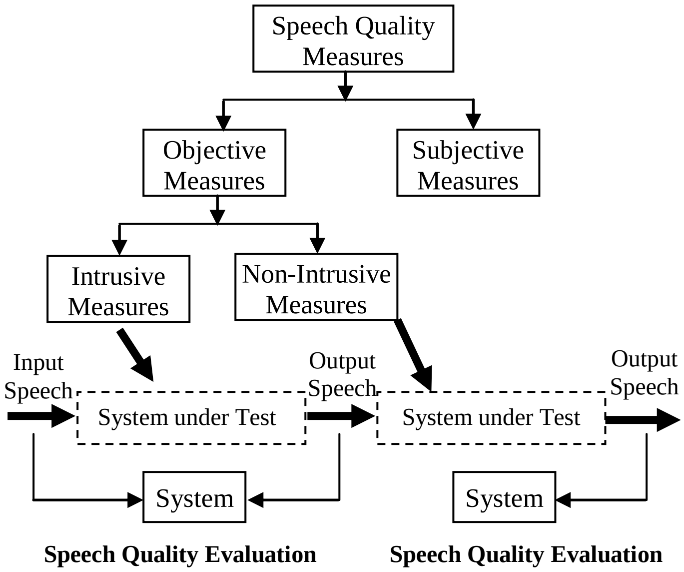

Fig.10. The classification of speech quality measurement

This paper work focuses only on objective speech quality measurements because the subjective measurements are time consuming and expensive. In industry, it is very critical to meet software deadlines often. Hence it would be handy to objectively test the software's performance. A combination of Itakura-Saito scheme mentioned in [3] is used.

12.1 PESQ

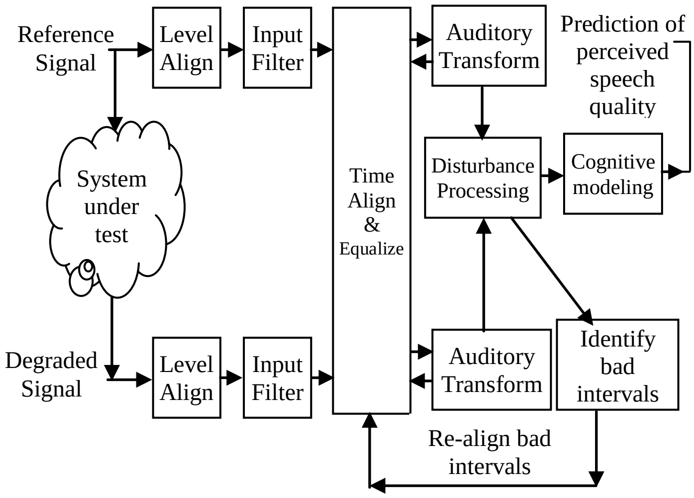

Fig.11. Structure of perceptual evaluation of speech quality (PESQ) model [1]

PESQ stands for 'Perceptual Evaluation of Speech Quality' and is an enhanced perceptual quality measurement for voice quality in telecommunications.

PESQ [11] was specifically developed to be applicable to end-to-end voice quality testing under real network conditions, like VoIP, POTS, ISDN, and GSM [6]. The structure of the PESQ measure is shown in Fig.11. The clean and degraded signals are first level-equalized to a standard listening level. Then they are filtered by a filter with response of a standard telephone handset. The signals are then synchronized in time to compensate for any time delays, and then processed through an auditory transform to obtain the loudness spectra. The auditory transform in PESQ uses a psychoacoustic model which translates the reference and degraded signals into a representation of perceived loudness in time and frequency.

12.2 ITAKURA-SAITO (IS)

The Itakura–Saito distance is a measure of the perceptual difference between a reference power spectrum $S(\omega)$ and a test spectrum $X(\omega)$. It was proposed by Fumitada Itakura and Shuzo Saito in the 1970s while they were with Nippon Telegraph and Telephone [1].

$$d_{iS}\left(X(\omega), S(\omega)\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{S(\omega)}{X(\omega)} - \log\left(\frac{S(\omega)}{X(\omega)}\right) - 1\right] d\omega \quad (12)$$

Owing to its asymmetric nature, the IS measure provides more emphasis on spectral peaks than spectral valleys. The IS distortion measure between the estimated and true short-time power spectra at the k^{th} frequency bin is given by,

$$d_{iS}t_{IS}\left(X_k^2, \hat{X}_k^2\right) = \frac{X_k^2}{\hat{X}_k^2} - \log\left(\frac{X_k^2}{\hat{X}_k^2}\right) - 1 \quad (13)$$

In [3], Loizou found that a PESQ- IS combination ended in a correlation coefficient that is greater than 0.9 between predicted and actual quality scores. This combination is given by,

$$\begin{aligned}
&\mathrm{BF1} = \max(0,\ \mathrm{PESQ} - 1.696);\\
&\mathrm{BF2} = \max(0,\ \mathrm{IS} - 11.708);\\
&\mathrm{BF3} = \max(0,\ \mathrm{IS} - 3.559);\\
&\mathrm{BF4} = \max(0,\ \mathrm{PESQ} - 2.431);\\
&\mathrm{BF5} = \max(0,\ \mathrm{PESQ} - 2.564);\\
&\mathrm{Y_all} = 1.757 + 1.740 \times \mathrm{BF1} + 0.047\\
&\quad \times \mathrm{BF2} - 0.049 \times \mathrm{BF3} - 2.593\\
&\quad \times \mathrm{BF4} + 11.549 \times \mathrm{BF5};
\end{aligned} \quad (14)$$

In [3], the authors evaluated a number of speech enhancement algorithms both objectively and subjectively. These algorithms are tabulated in Table.1. In our paper, all these algorithms will not be discussed widely but their names are mentioned here because the MOS scores of these algorithms published in [3] will be compared against our Speech Enhancement algorithm for the sake of justifying any drawn conclusion. We will evaluate Speech Enhancement for speech corrupted by multi-talker babble noise, restaurant noise and airport noise.

Table.1. List of speech enhancement algorithms mentioned in [3] included for comparative purposes

Sl. No.	Abbreviation	Full form of algorithm
1	MMSE SPU	Minimum Mean Square Estimation Speech Presence Uncertainty
2	logMMSE	Log Minimum Mean Square Estimation
3	logMMSE SPU	Log Minimum Mean Square Estimation Speech Presence Uncertainty

4	pMMSE	Speech Enhancement based on perceptually motivated Bayesian Estimators of the Magnitude Spectrum
5	AudSup	Speech Enhancement based on Audible Noise Suppression
6	Wiener-as	Speech Enhancement based on A Priori Signal To Noise Estimation
7	WT	Speech Enhancement based on Wavelet Thresholding the Multitaper Spectrum
8	MB	Multi-Band Spectral Subtraction
9	RDC-ne	RDC Algorithm That Included Noise Estimation
10	RDC	Spectral Subtraction using Reduced Delay Convolution and Adaptive Averaging.
11	KLT	Karhunen-Loeve Transform
12	pKLT	Perceptual Karhunen-Loeve Transform

13. PLOTS AND RESULTS

As shown in Fig.12 and Fig.13, the corrupted speech was successfully cleaned and processed for speech enhancement. The mean opinion scores of the algorithms mentioned in Table.1 was compared against MOS for SS-LPC algorithm.

It was inferred that the SS-LPC algorithm slightly outperformed the existing algorithms when the speech is corrupted by babble noise, restaurant noise and airport background noise at SNR 10 dB. The MOS scores was collected for a set of 16 sentences mentioned in [3] and average of all 16 scores were computed and plotted for SS-LPC algorithm in Fig.14, Fig.15 and Fig.16.

The MB (Multi band Spectral subtraction) was known to perform best at 10 dB SNR out of rest of the algorithms known. While an objective score improvement was observed for SS-LPC algorithm as against MB algorithm, [4] and [7] mentions that only a change in MOS score by 0.25 will cause a change perceptually. A change by negative 0.25 means a slight degradation of speech quality and +0.25 improvement means speech quality improvements. Our SS-LPC algorithm does neither shows scores that degrade speech quality subjectively nor does it seem to improve speech quality subjectively as the scores objectively have increased slightly than MB MOS scores.

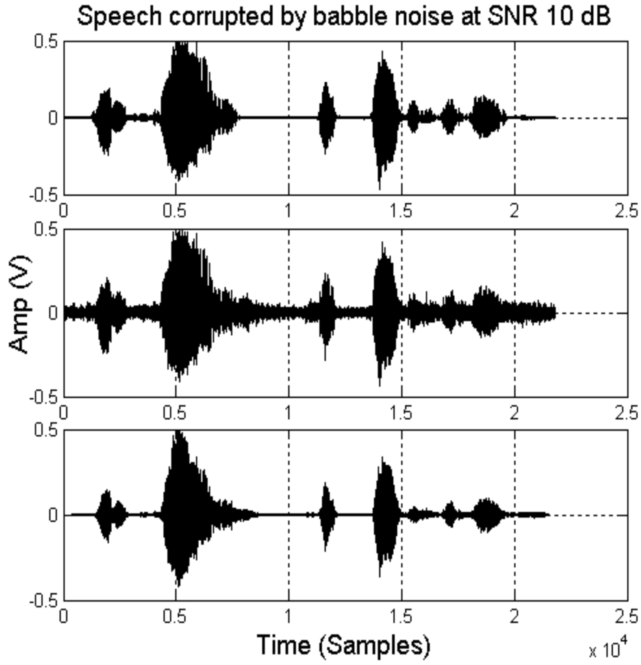

Fig.12. Time domain plots: (Top) clean speech; (mid) Speech corrupted by babble noise at 10 dB SNR; (bottom) Output of SS-LPC algorithm with processed speech

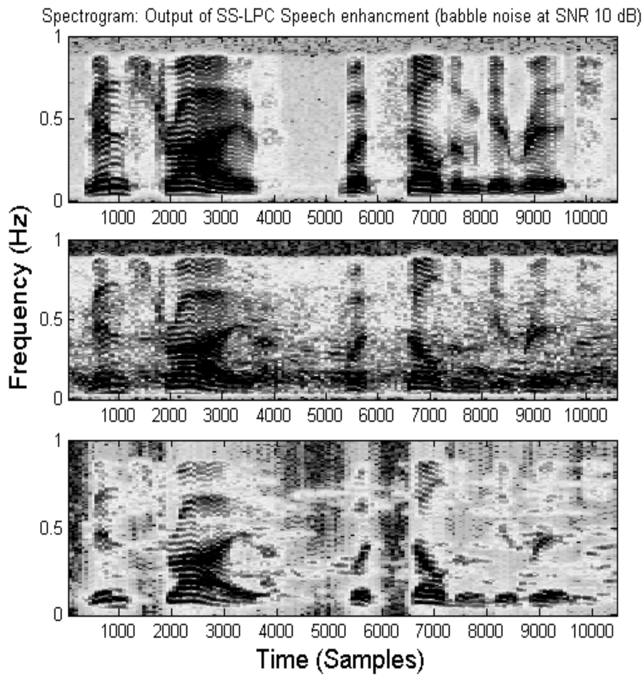

Fig.13. Spectrograms: (Top) clean speech; (mid) Speech corrupted by babble noise at 10 dB SNR; (bottom) Output of SS-LPC algorithm with processed speech.

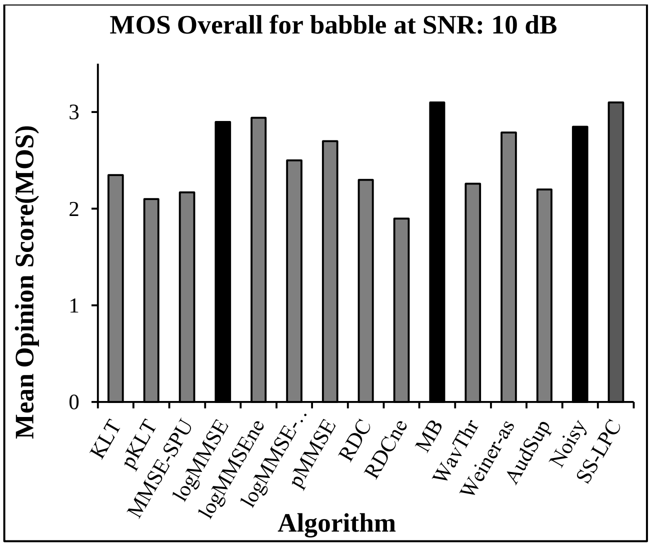

Fig.14. MOS overall for babble at SNR 10 dB

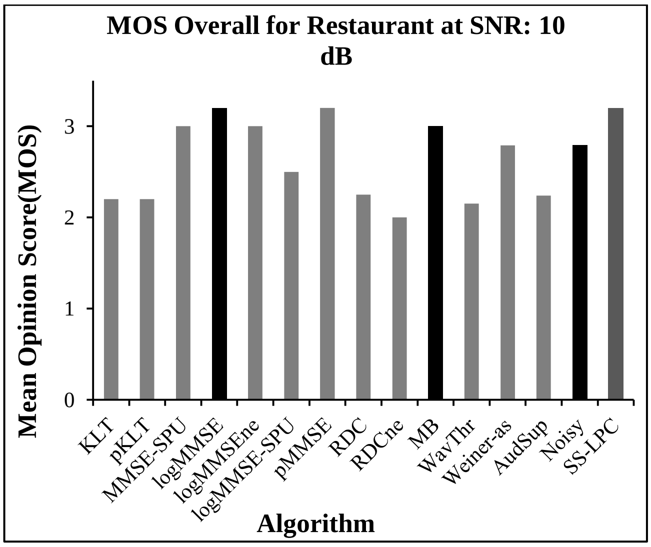

Fig.15. MOS overall for restaurant at SNR 10 dB

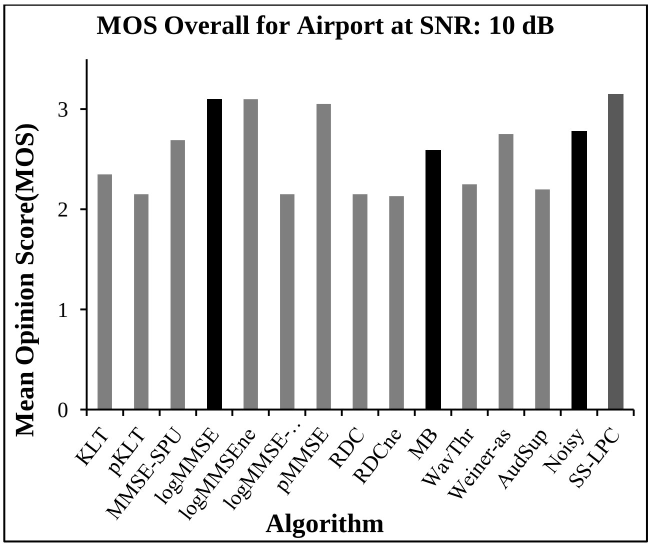

Fig.16. MOS overall for airport at SNR 10 dB

14. CONCLUSION

In this work, we have proposed an algorithmic strategy for single channel speech enhancement in noisy conditions for mobile speech processing applications. A combination of spectral subtraction and linear prediction coding is used. A PESQ-IS evaluation strategy is used to benchmark the software written in MATLAB and the synergistic effect of the SS-LPC combination in improving speech quality is discussed. The results could be summarized as follows,

- A minimal improvement in objective MOS scores was observed for speech corrupted by babble noise, restaurant noise and airport noise at SNR 10 dB. Babble noise is stationary noise but restaurant and airport noise is unstationary.
- Marginal subjective speech quality could be possible based on correlation equation provided for objective-subjective scores in [4] and [7].
- Further improvements need to be made in future for attaining greater performance.

REFERENCES

[1] P.C. Loizou,"*Speech Enhancement: Theory and Practice*", CRC Press, Boca Raton, FL, 2007.

[2] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 27, No. 2, pp. 113-120, 1979.

[3] Yi Hu and Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement", *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 16, No. 1, pp. 229-238, 2008.

[4] Alan O'Cinneide, David Dorran and Mikel Gainza, "Linear prediction-The technique, ITS solution and Application to Speech", Dublin Institute of Technology, Internal Technical Report, 2008.

[5] Jin Wen, Xin Liu, M.S. Scordilis and Lu Han, "Speech Enhancement Using Harmonic Emphasis and Adaptive Comb Filtering", *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 18, No. 2, pp. 356-368, 2010.

[6] Yi Hu and Philipos C. Loizou, "*Subjective comparison and evaluation of speech enhancement algorithms*", presented at Speech Communication, Vol. 49, No. 7, pp. 588-601, 2007.

[7] K. Lebart, J.M. Boucher and P.N. Denbigh, "A New Method Based on Spectral Subtraction for Speech Dereverberation", *Acta Acustica united with Acustica*, Vol. 87, No. 3, pp. 359-366, 2001.

[8] Ekaterina Verteletskaya and Boris Simak, "Noise Reduction Based on Modified Spectral Subtraction Method", *IAENG International Journal of Computer Science*, Vol. 38, No. 1, 2011.

[9] Jeremy Bradbury, "*Linear Predictive Coding*", 2000.

[10] ITU-T Recommendation P.862, "Perceptual Evaluation of Speech Quality (PESQ), an objective method for end to end speech quality assessment of narrowband telephone networks and speech coders", 2001.

[11] http://www.pesq.org