

AN EFFECTIVE EVALUATION STUDY OF OBJECTIVE MEASURES USING SPECTRAL SUBTRACTIVE ENHANCED SIGNAL

Sukhvinder Kaur¹, Anil Garg²

¹M.Tech Student MMEC, Mullana, M.M. University, Mullana-Ambala-Haryana, India

²Associate Prof. in ECE department MMEC, Mullana, M.M. University, Mullana-Ambala-Haryana, India

Abstract

Unwanted noises have a negative influence over communication because it disturbs the conversation and make the communication impossible. Speech enhancement algorithms are used for improving the quality and intelligibility or to reduce listener fatigues. Assessment of speech quality can be done by using either subjective listening test or objective quality measure. Evaluation of several objective measures with the speech processed by enhancement algorithms has been performed but these having limitations to assess original speech signal. This paper represents the study of speech quality measures and compute the values used for regression analyses of the objective measures evaluation study using spectral subtraction algorithm based enhanced speech signal.

Keywords: MOS, ITU-T (P.835), SNRseg, log-likelihood ratio and itakura-saito.

1. INTRODUCTION

Speech Communication is an integral part of daily life. Communication systems include various applications such as, mobile phones, speech coding and compression, medical devices such as hearing aids etc. All these are associated with different types of noise or sounds. Unwanted sounds are commonly referred as disturbances. It can be both noise or interferences and which are not acceptable in the communication process. Therefore, unwanted sounds are having a negative influence over the communication because they may disturb the conversation and make it impossible to communicate at all. Signal Processing methods are the effective way for aiding the speech communication by reducing the disturbance levels with respect to the level of the speech. Such speech processing methods are known as speech enhancement methods. Speech enhancement algorithms improve the quality and intelligibility of speech or reduce listener fatigues.

Speech enhancement has several real world applications which include: Telecommunications, Electronic hearing aids and Automatic speaker recognition software. These all are of important quality and intelligibility of speech which vastly improves the user's listening experience. Two criteria are which generally used for measuring the performance of speech: quality and intelligibility. It is quite hard to satisfy both of them at the same time. Quality is the subjective measure which indicates the amount of effort needed to understand the speech material. Whereas Intelligibility, is an objective measure which signifies the amount of speech material correctly understood.

This paper is represented in such a manner that section 2 provides brief review of the different speech enhancement techniques and spectral subtraction algorithm. The speech quality measures are discussed in section 3. In the Section 4

evaluated results are shown and the last section 5 represents the conclusion part. References are given in the last section of the paper.

2. SPEECH ENHANCEMENT TECHNIQUES

Speech enhancement methods can be classified in two ways: (1) Single channel enhancement technique (2) Multi channel enhancement technique. Single channel enhancement techniques are applied to situations in which only one acquisition channel is available. Multi channel enhancement techniques are employed in microphone arrays and take advantage of availability of multiple signal inputs to system, by making possible use of phase alignment to reject the undesired noise components. Single channel enhancement techniques are further divided again into different categories: Short Time Spectral Amplitude Estimation (i.e. Spectral subtraction and Non Causal Weiner Filtering), Periodicity of Voiced Speech based, Adaptive Comb Filtering based, Harmonic Selection based, adaptive noise canceling techniques based, Speech model based (i.e. Kalman Filtering Hidden Markov Modelling(HMM)). Spectral-subtractive algorithms: Is the simplest speech enhancement algorithms based on the principle of estimating the clean signal spectrum by subtracting an estimated noise spectrum from the noisy speech spectrum. Statistical model based algorithm: are based on the estimation of linear and non linear transform coefficients of clean signal by measuring the Fourier transform coefficients of the noisy signal. E.g., Wiener algorithm and minimum mean square error (MMSE) algorithms. Subspace Algorithms: are based on the principle that the clean signals estimation is done by simply nulling the component of the noisy vector residing in the noise subspace.

2.1 Spectral-Subtractive Algorithms

The spectral subtraction algorithm is historically one of the first algorithms proposed for noise reduction based on the simple principle of assuming additive noise. Estimated clean signal spectrum can be obtained by subtracting an average noise spectrum from the noisy speech spectrum.

The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present. Subtraction process needs to be done carefully to avoid speech distortions. If too much subtraction is done, then some speech information might be removed while if the subtraction is too little then much of the interfering noise remains.

3. SPEECH QUALITY MEASURES

Speech processing algorithms usage is rapidly increasing, which raises the need of speech quality evaluation. The main evaluation method of speech enhancement system affects the intelligibility of the speech signal and improves the overall quality of the signal. The assessment of speech quality can be done by either using subjective listening tests or objective quality measures i.e. Subjective measures and Objective measures.

3.1 Subjective Measures

Subjective evaluation involves comparisons of original and processed speech signals by a group of listeners who are asked to rate the quality of speech along a pre-determined scale. This evaluation method requires the judgment of human listeners. These can be further divided into two categories: (1) Relative Preference Methods - multiple signals are compared. (2) Absolute Category Rating Methods - a single stimulus is tested.

Table 1.1 MOS rating scale

Rating	Speech quality	Level of distortion
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 and objectionable

MOS (Mean Opinion score) is the most widely used subjective evaluation method [8]. The MOS test is based on five category rating scale of the speech quality given in Table 1.1. ITU-T standard (P.835) is designed for integrating the effect of signal and background distortions.

Subjective evaluation is the reliable method for assessment of speech quality but these are time consuming and expensive. This evaluation method is stringent in nature as it needed a set of trained listeners for determining the quality of speech. Therefore need of objective evaluation arises.

3.2 Objective Measures

Objective evaluation involves a mathematical comparison of original and processed speech. In this evaluation method the judgment is predicted with some analysis of the system. In this method quality is quantified by measuring the numerical distance between the original and processed signals. There are different types of objective quality measures which are further categorized as: (1) Time and frequency signal to noise ratio measures, (2) Spectral distance measure based on LPC, (3) Perceptually Motivated measures and (4) Composite measures.

The objective measures accesses the quality of the processed speech without accessing the original speech signal. Objective measures of speech quality are implemented by first segmenting the speech signal into 10-30 ms frames, and then computing a distortion measure between the original and processed signals. Distortion measure computation can either be done in the time domain (e.g. signal to noise ratio measures) or in the frequency domain (LPC spectral distance measures). In the frequency domain measure distortions and differences detected in the magnitude spectra are correlated with speech quality.

In Time and Frequency Signal to Noise Ratio Measures, the segmental signal to noise ratio can be evaluated either in terms of time or frequency domain. The original and processed signals both are aligned either in time domain or in frequency domain for the calculation. The main advantage of using frequency based segmental SNR over the time domain (SNRseg) is the flexibility to place different weights for different frequency bands of the spectrum.

In the Spectral Distance Measure Based on LPC several LPC based objective measures are proposed on the basis of all pole models of the clean and enhanced speech signals. Three different LPC based objective measures include: The LLR (log-likelihood ratio) measure, the IS (Itakura-Saito) measure and the Cepstrum distance measure. Cepstral distance measure is derived from the LPC coefficients.

Perceptually Motivated Measures are used for the modeling of auditory processes like normal listening frequency selectivity and the perceived loudness. Weighted spectral slope (WSS), Bark distortion measures (BSD) and Perceptual evaluation of speech quality (PESQ) measures are the perceptually motivated measures.

Composite Measures are formed by combining multiple objective measures. Regression analysis is used to compute the optimum combination of objective measure for maximum correlation. The selection of objective measures for composite measures depends basically on the experimental evidence and intuition. MARS technique can be used for composite measure evaluation.

Figure of Merits are obtained by statistical analysis. This analysis is necessary for the validity of evaluation method of objective measures. Statistical analysis is used to assess the correlation between subjective score and the values of the

objective measures. Two types of figure of merit are calculated in statistical analysis: (1) Pearson's correlation coefficient. (2) Standard error of the estimate.

The first figure of merit Pearson's correlation coefficient is used to obtain the correlation between subjective listening scores and objective measures. Second figure of merit is an estimate of the standard deviation of the error obtained by objective measures to predict subjective listening scores.

3.2.1 Time and Frequency Signal to Noise Ratio Measures

The segmental signal to noise ratio is evaluated either in terms of time or frequency domain. The time domain measure is the simplest objective measure which is used to evaluate speech enhancement and speech coding algorithms. In this measure original and processed signal are aligned in time and phase errors are present. The segmental signal to noise ratio is given by eqn. (3.1)

$$SNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} (x(n) - \hat{x}(n))^2} \quad (3.1)$$

Where $x(n)$ is the original signal (clean), $\hat{x}(n)$ is the enhanced signal, N is the frame length and M is the number of frames in the signal. The segmental SNR in terms of frequency domain produce the frequency weighted segmental SNR $fwSNR_{seg}$ is given in eqn. (3.2)

$$fwSNR_{seg} = \frac{10}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^k B_j \log_{10} \left[\frac{F^2(m,j)}{(F(m,j) - \hat{F}(m,j))^2} \right]}{\sum_{j=1}^k B_j} \quad (3.2)$$

Where B_j is the weight placed on the j th frequency band, k is the number of bands, M is the total number of frames in the signal. $F(m,j)$ is the filter bank amplitude (excitation spectrum) of the clean signal and $\hat{F}(m,j)$ is the filter bank amplitude of enhanced signal in the same bank.

3.2.2 LPC based Objective Measures

LPC based objective measures including the log likelihood ratio (LLR), Itakura Saito distance measure (IS), cepstrum distance measure (CEP). The two most commonly used measures are LLR and IS. Cepstral distance measure is derived from the LPC coefficients. The log likelihood ratio (LLR) measure is defined as in eqn.

$$d_{LLR}(a_x, \bar{a}_{\hat{x}}) = \log \frac{\bar{a}_{\hat{x}}^T R_x \bar{a}_{\hat{x}}}{a_x^T R_x a_x} \quad (3.3)$$

Where a_x^T the coefficients of the clean signal are, $\bar{a}_{\hat{x}}^T$ are the coefficients of enhanced signal and R_x is the autocorrelation matrix of clean signal. The Itakura Saito (IS) measure is defined as in eqn. (3.4)

$$d_{IS}(a_x, \bar{a}_{\hat{x}}) = \frac{G_x \bar{a}_{\hat{x}}^T R_x \bar{a}_{\hat{x}}}{\bar{G}_{\hat{x}} a_x^T R_x a_x} + \log \left(\frac{\bar{G}_{\hat{x}}}{G_x} \right) - 1 \quad (3.4)$$

Where G_x and $\bar{G}_{\hat{x}}$ are the all pole gains of the clean and enhanced signals. The cepstrum distance measure (CEP) is defined as in eqn. (3.5)

$$d_{CEP}(c_x, \bar{c}_{\hat{x}}) = \frac{10}{\log_e 10} \sqrt{2 \sum_{k=1}^p [c_x(k) - \bar{c}_{\hat{x}}(k)]^2} \quad (3.5)$$

Where $c_x(k)$ and $\bar{c}_{\hat{x}}(k)$ are the cepstrum coefficients of the clean and enhanced signals.

4. EVALUATION RESULTS

Results evaluated for the Statistical Analysis of spectral subtractive enhanced signal is discussed in this section. In this paper, the clean and processed signal is used for statistical analysis. Speech signal is first segmented into frames and then standard deviation and mean values are computed for the noisy and enhanced signals shown in table 1.2. The original time waveform for clean noise, noisy and enhanced signals are shown in fig 1. These are used as the primary tool for the result analysis.

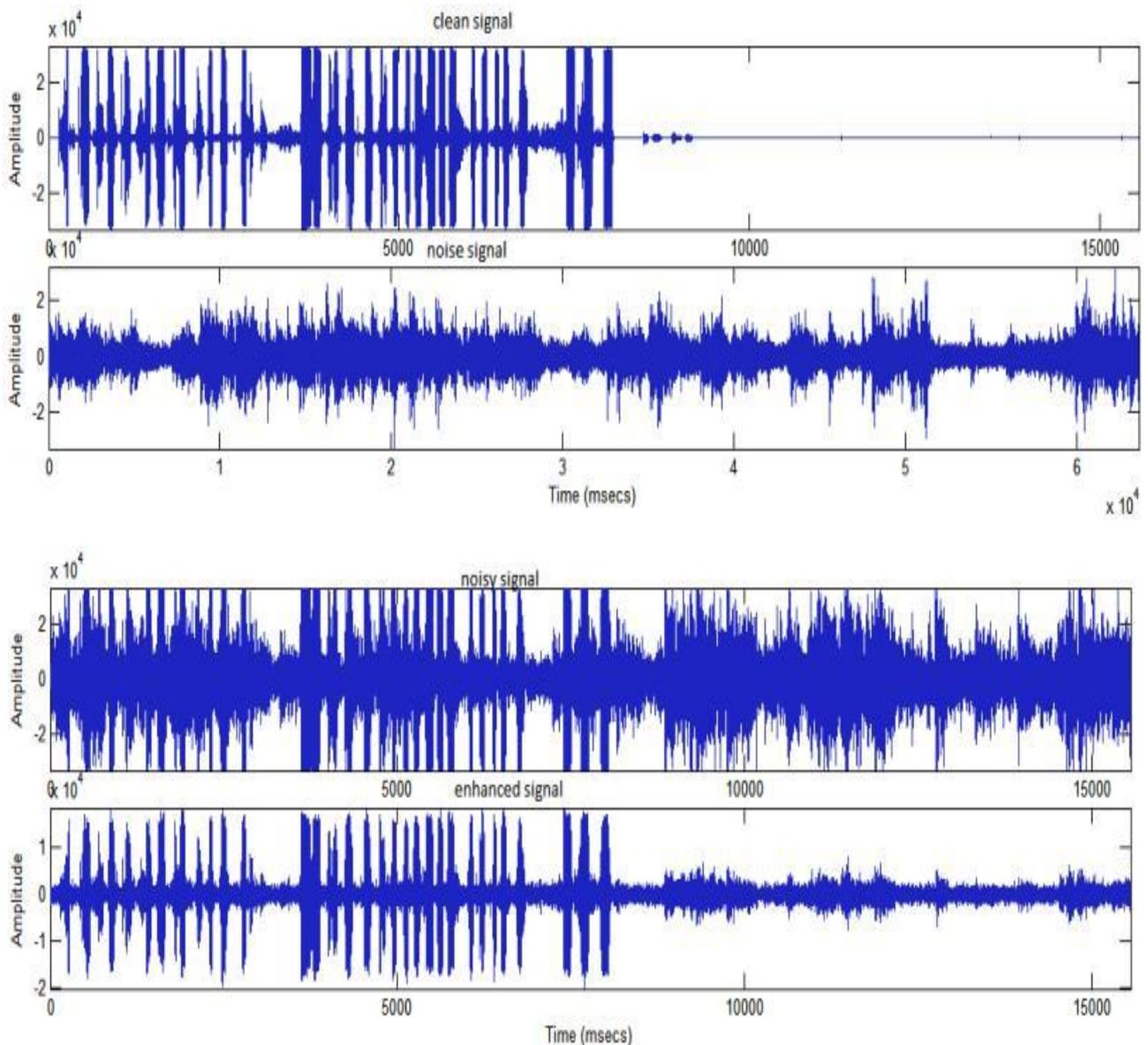
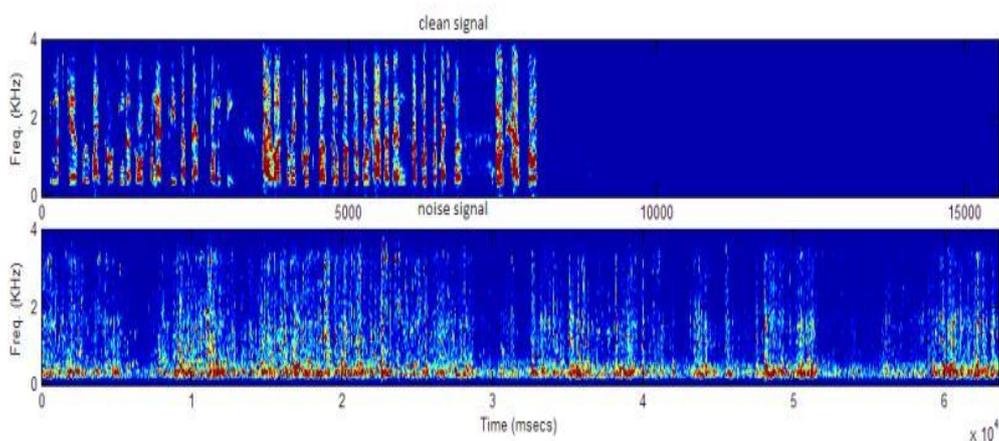


Fig 1: Time waveform representation of clean, noise, noisy and spectrally subtractive enhanced signal

The spectrogram of clean, noise, noisy and enhanced signal shown in fig 2. Spectrogram is utilized for detecting the differences in the magnitude spectra of waveform

information. The quality of the processed speech is determined by without needing access to the original speech signal.



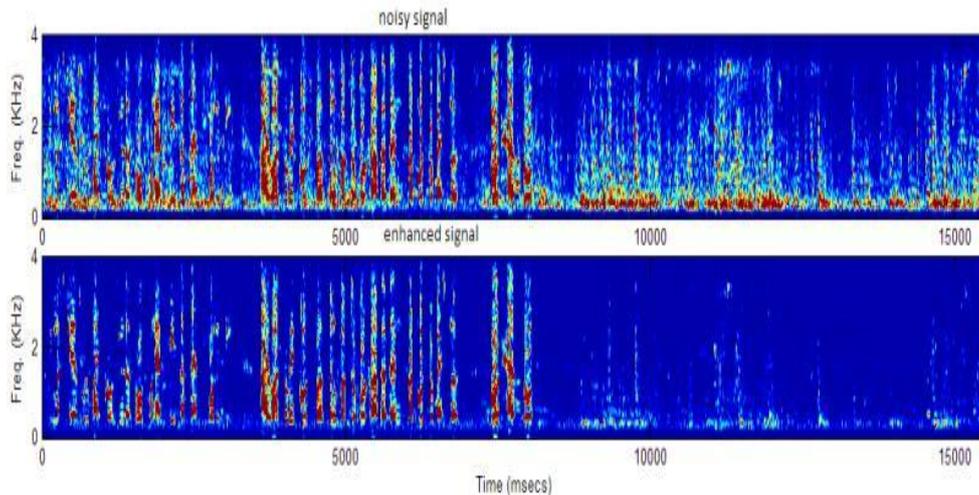


Fig 2: Spectrogram representation of clean, noise, noisy and spectrally subtractive enhanced signal

Computation of the LPC measure is done by estimating the correlation coefficient by using the mean value and standard deviation values. Table 1.2 shows the respective values of mean and standard deviation for different objective measures such as SNR, IS, LR, LLR, WSM. These values are the LPC coefficients that can be used for computing the objective measures.

Table 1.2 Table for mean and standard deviation values of objective measure

Objective measure	Mean value	Standard deviation
SNR	1.388178	1.727957
IS	30.170003	21.972701

LR	0.205720	0.189852
LLR	0.176183	0.139966
WSM	3.686912	2.256240

SNR plot of enhanced signal with respect to the noisy signal is shown in fig 3. The plot shows the smaller values of the signal energy during the interval of silence in the speech signal results the larger negative value of (SNR).

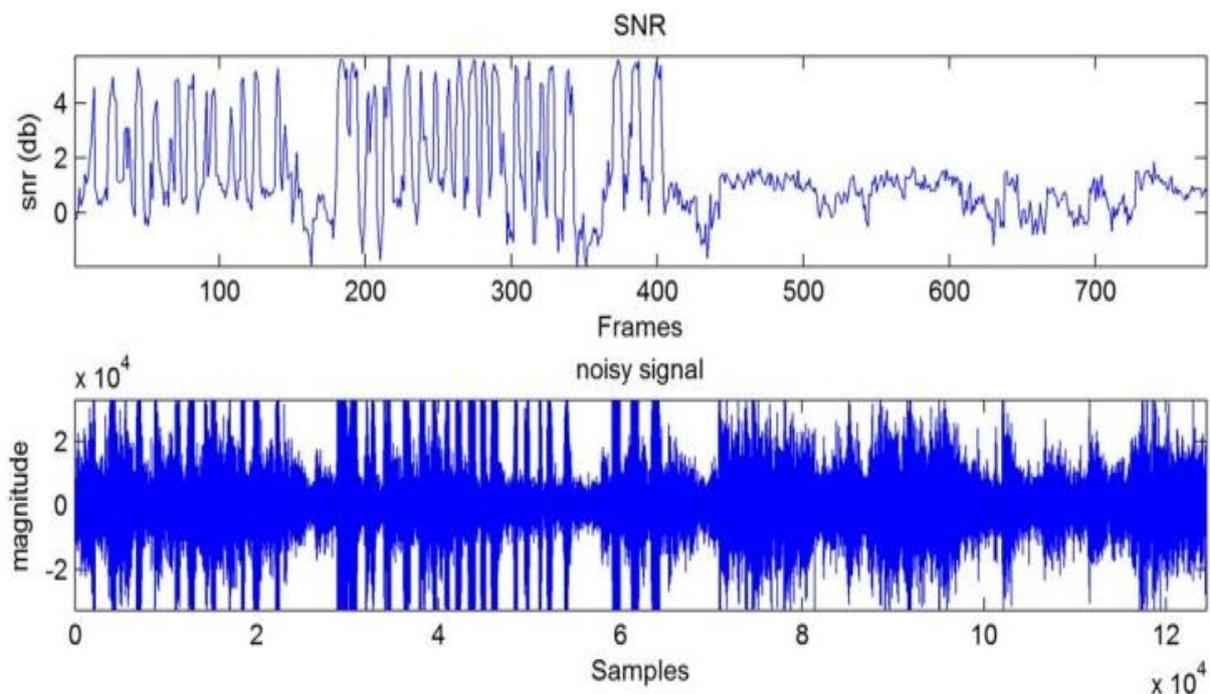


Fig 3: SNR plot with respect to the noisy signal spectrum

Itakura saito ratio (IS) plot of enhanced signal with respect to the noisy signal is shown in fig 4. The plot shows the smaller values of the signal energy during the interval when the speech in the processed signal is of highest amplitude

i.e., speech is clearly understandable and distractive and larger values of IS shown for the noise present in processed signal.

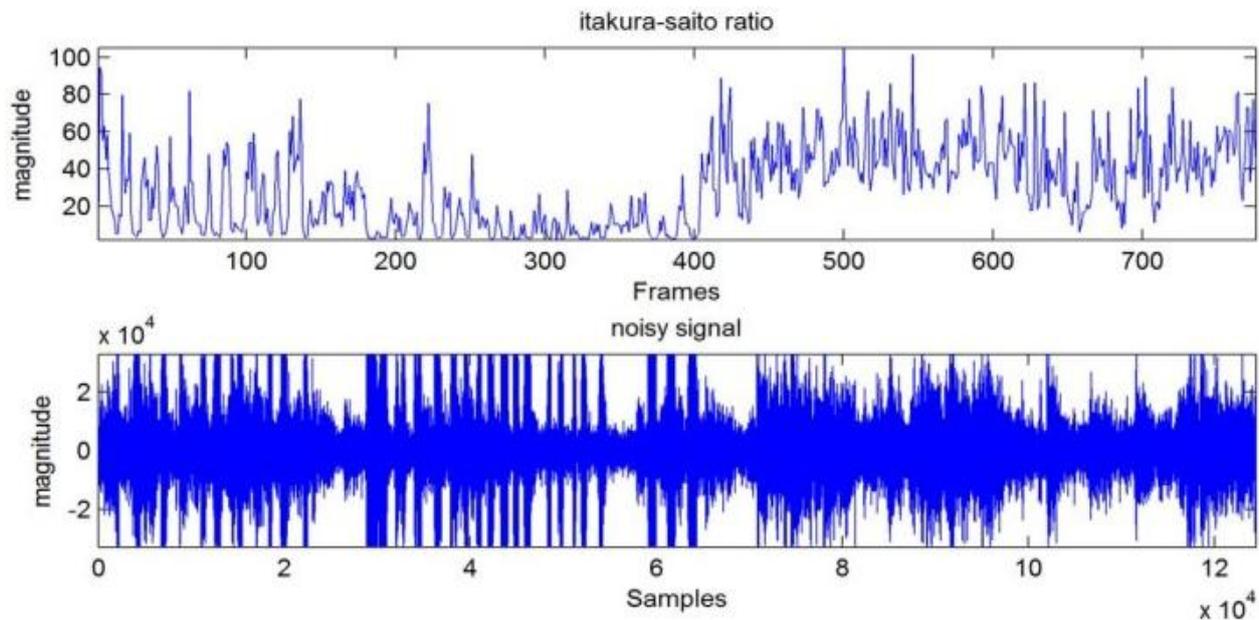


Fig 4: IS (itakura- saito ratio) plot with respect to the noisy signal spectrum

LLR (log likelihood ratio) plot of enhanced signal with respect to the noisy signal is shown in fig 5. The plot shows the lower values of the signal energy during the interval of

speech signal and results the larger peak value of (LLR) for the noise in the processed signal spectrum.

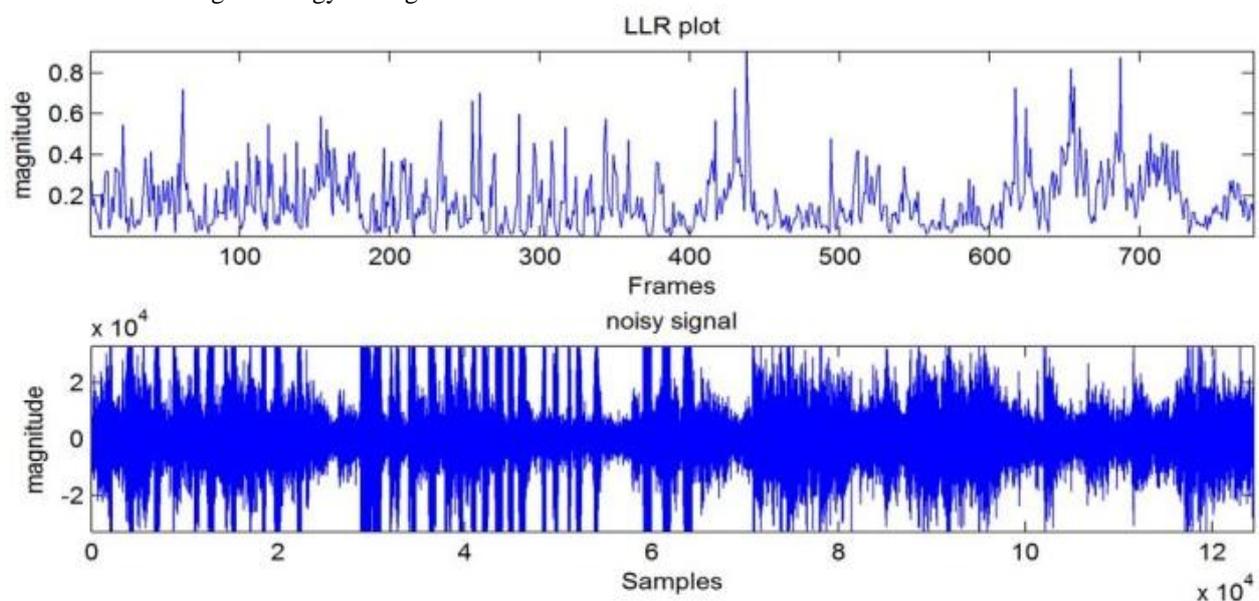


Fig 5: LLR (log likelihood ratio) plot with respect to the noisy signal spectrum.

5. CONCLUSION

Different speech enhancement algorithms have been proposed to improve the performance of modern communication devices in noisy environments. The key conclusion drawn from this paper is that speech quality

assessment is done through subjective and objective evaluation measures. In this paper study of objective measure and computed values of the standard deviation and mean for noisy and enhanced signal are given.

It is very difficult to find reliable and fair comparison among different algorithms due to lack of common speech database used in algorithms evaluation, different types of noise and differences in the testing methodology. The most accurate method for evaluating speech quality is subjective listening tests but these are time consuming and expensive. This evaluation method is stringent in nature as it needed a set of trained listeners for determining the quality of speech. Therefore need of objective evaluation arises. As the frame length increases SNR value decreases monotonically.

REFERENCES

- [1]. Loizou, P., 2007. *Speech Enhancement: Theory and Practice*. CRC Press, Boca Raton, FL.
- [2]. Y. Hu and P. C. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," submitted to *Speech Communication*.
- [3]. Yi Hu and Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement", *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 16, No. 1, January 2008.
- [4]. P. Chu and D. Messerschmitt, "A frequency weighted itakura-saito spectral distance measure", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 30, pp. 545—560, Aug. 1982.
- [5]. Anuradha R. Fukane, "Enhancement of Noisy Speech Signals for Hearing Aids", 2011 International Conference on Communication Systems and Network Technologies, 978-0-7695-4437-3/11© 2011 IEEE
- [6]. Sunil D. Kamath, "A Multi-Band Spectral Subtraction Method for Enhancing Speech Corrupted By Colored Noise", 2002
- [7]. Jianfen Ma, Yi Hu and Philipos C. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions", *J. Acoust. Soc. Am.*, Vol. 125, No. 5, May 2009.
- [8]. Kuldip Paliwal, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain", Elsevier, *Speech Communication* 52 450–475, 2010.