DOI: 10.21917/ijct.2015.0178

NOISE ESTIMATION USING STANDARD DEVIATION OF THE FREQUENCY MAGNITUDE SPECTRUM FOR MIXED NON-STATIONARY NOISE

A. Indumathi[1] and E. Chandra[2]

[1]*Department of Master of Computer Applications, Dr. SNS Rajalakshmi College of Arts and Science, India*

E-mail: endhumathi@gmail.com

[2]*Department of Computer Science, Bharathiar University, India*

E-mail: crcspeech@gmail.com

Abstract

Noise estimation and suppression is very important for improving the quality of speech signal. Noises exist in almost all places. In reality, more than one noise degrades the speech signal. It is hard to find and supress various types of noise that affect the speech quality. This paper proposed a method for noise estimation of mixed non-stationary noisy speech signal. This method uses Spectral properties of the noisy speech signal to detect the frequency regions of noise signal. Highest frequency of speech signal is calculated and it is considered as the threshold value for separating noise signal and clean speech signal. Using Spectral subtraction, Standard deviation of noise spectrum is subtracted with noisy spectrum to acquire enhanced speech signal. Performance of the method is evaluated using SNR and Spectrogram. The main focus of this paper is to propose an independent method which estimates the noise of any type and nature.

Keywords:

Noise Estimation, Spectral Subtraction, Standard Deviation, SNR, Spectrogram

1. INTRODUCTION

Speech is the natural mode of communication for human. They express their minds through speech. Speech signals have rich and temporal spectral variation that conveys words, intention, expression, emotion, gender and speaker identification [3]. In real time situation, the speech received by human is not the clear speech. The quality of speech is degraded by various background noises. Noise exits everywhere with different causes and with different features. Some of the noises are Additive noise which gets added to the proposed signal like Background noise due to Bogus sounds, Terminal noise during Signal capture and Environment noise due to machine, Industry, Transport, Reverberation and Echo. Many research in recent years have been implemented various technique to reduce the noise. It is difficult to find the factors for noise and hard to cancel the specific noise. This paper introduces a method to suppress all background noises invariable of noise type.

2. RELATED WORK

In recent studies, several filter designs have been implemented in Speaker recognition to suppress and eliminate the unwanted background noise, as well as to improve speech quality [1]-[4]. Implemented a technique for removal of high frequency noise for speech enhancement using Frequency Response Masking (FRM) based on designing low complexity, narrow transition bandwidth and linear phase Finite Impulse Response (FIR) filters [1]. Noise spectrum is calculated from speech absence frames through Voice Activity Detector (VAD) or Minimum Statistic methods [5].

Many Noise estimation algorithms are proposed for strange non-stationary noise signals [5-7]. Martin [5] proposed an algorithm for estimating the noise by tracking the various noise level using Minimum Statistics (MS) [5]. Cohen proposed an algorithm for noise estimation by tracking the noise only region of the noisy speech spectrum called Minima Controlled Recursive Algorithm (MCRA) [6]. R. Sundarrajan and C.L. Philipos proposed a method for comparing the ratio of the noisy speech signal to the local minimum against a threshold [7]. Doblinger [8] proposed a method for estimating the noise spectrum for the speech signal, the main drawback of this algorithm is noise estimation increases simultaneously with increase in noisy speech power. Cohen and Berdugo proposed a method for noise estimation. In this, noise estimation is updated continuously by averaging the past spectral values of the noisy speech with time and frequency dependent smoothing factors [9].

Hiresh H.G and Ehrlicher C [13] proposed a method based on estimating a histogram of past spectral values, which are compared against a Threshold. Threshold is based on the past noise estimation. This algorithm fails to adapt when the noise estimation suddenly increase.

Marc Karam et al. proposed a method for Noise removal for real time data using Spectral subtraction [14].

Spectral Subtraction [10]-[12] is a traditional approach for reducing additive background noise in Single channel system.

Spectral Subtraction with Partial Differential Equation (PDE) method has been proposed by [15]. In this method, input speech signal is enhanced using PDE and the quality of speech signal is improved through Spectral Subtraction.

M. Thirumarai Chellapandi and P. Kabilan [16] proposed a method to improve the quality of speech for mobile application by reducing acoustic noise using Spectral Subtraction and Linear Prediction Coding (LPC).

Poornapriya G et al. [17] proposed a method to isolate the noise from noisy speech signal for mobile phone users. Wavelet based Spectral Subtraction technique is implemented to enhance the speech signal by reducing the interferences.

Israel Cohen, Baruch Berdugo [18] presented a speech estimator OM-LSA and a noise estimation method for Non-stationary environment.

Lin. L, E. Ambikairajan, W.H. Holmes [19] proposed a technique using auditory filter bank for denoising the speech signal which is suitable for Non-stationary and colored noise environment.

To overcome the drawback of Short time Spectral attenuation (STSA) Kotta [20] proposed a post processing method to detect noise domination region which are attenuated using a SNR Based rule.

3. SYSTEM OUTLINE

Spectral Subtraction and Linear Prediction suits good for eradication of Background noise and residual Noise [20]. Block diagram of our proposed system is shown below.

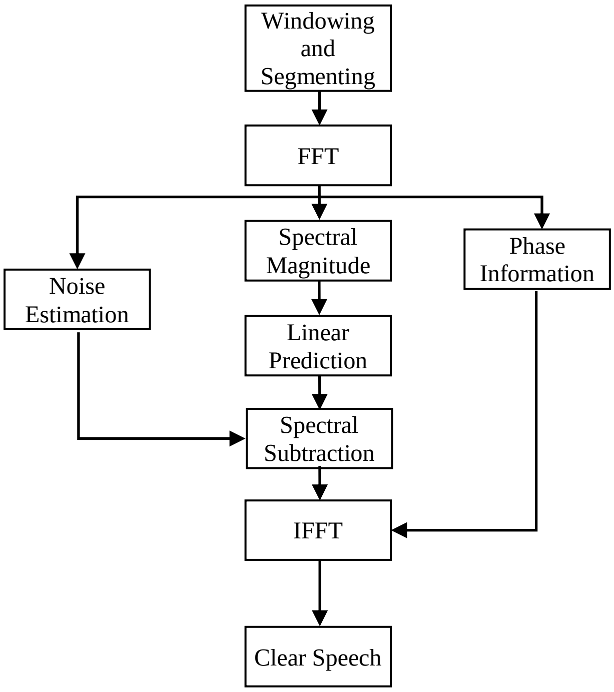

Fig.1. Block diagram of proposed method

Speech is non-stationary signal which is affected by various environmental noises. Properties of speech signal changes rapidly over time. Speech signal is represented as,

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

where, $x(n)$ is clean speech signal, $d(n)$ is noise signal and $s(n)$ is Noisy speech signal.

In general, noise properties change quit rapidly over time in non-stationary environment. Noisy speech signal remains invariant for a short period of time [13]. The speech signal is divided into the frames by using Hamming window. In windowing, the beginning and end of the signal is attenuated in the calculation of the spectrum. Hence overlapping is performed to recover a portion of each previous frame that is lost due to windowing. Overlapping ensure better temporal continuity in the transform domain. An overlap of half the window size is typical [6]. Thus input speech signal is segmented into frames with overlap of 1/2 of the frame size. The time domain of framed speech signal is denoted as,

$$s_i(n) = s_1(n) + s_2(n) + s_3(n) + \ldots + s_i(n) \tag{2}$$

where, n ranges over 1-256 samples and i denotes the frame number corresponding to the time-domain frame. Phase value is calculated and stored for later. FFT is performed to obtain the magnitude frequency response of each frame. In FFT, the signal within a frame is considered as periodic, and continuous when wrapping around.

Linear Prediction is one of the most important tools in speech processing. LP is a useful method for estimating parameters of recorded speech signal. The output samples are predicted as a linear combination of filter co-efficient and previous samples.

For any noisy speech signal, back ground noise is to be suppressed. Spectral subtraction is commonly used algorithm for speech enhancement. The basic principle of spectral subtraction is subtracting noise spectrum from the noisy speech spectrum, which results in clean speech spectrum.

IFFT is used to reconstruct the spectral speech signal and transforms Frequency domain signal into Time domain enhanced speech signal.

4. PROPOSED NOISE ESTIMATION ALGORITHM

Noise suppression is an important factor of modern communication systems. Pre-processing of noisy speech signal improves the performance of speech communication system for signals that was corrupted by stationary environment noise, through improving the speech quality or intelligibility. It is difficult to suppress non stationary noise with pre-processing. The proposed method presents a Noise estimation technique with Spectral subtraction which well suited for rapidly varying noise in non-stationary environment. Noise estimation and Spectral Subtraction cannot be successfully implemented in time domain. Hence, time domain signal is converted into frequency domain signal using FFT.

Noisy speech signal is segmented, windowed and half overlapped. Spectral magnitude of noisy speech signal is obtained through FFT. Spectral magnitude of the noisy speech signal is filtered using Linear Prediction. Highest frequency of speech signal is calculated from noisy speech signal and it is treated as "Threshold value". This value helps to calculate the frequency range of Noise signal from Noisy speech signal.

In proposed method, Noise is estimated by using mean and Standard deviation. As the characteristics of signal changes rapidly in non-stationary signal, the slowly changing mean interferences with the calculation of the standard deviation leads to inaccurate value [14]. This problem is overcome by dividing the signal into small sections and calculating the statistics for each section individually and Standard deviation for each of the section is averaged to produce a single accurate value.

The mean value of signal is calculated from,

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x_i \tag{3}$$

where, μ is mean value of signal, x_i is a signal, i is an index which varies from 0 to N-1. Sum of the values in signal x is calculated as,

$$x_0 + x_1 + x_2 + x_3 + \ldots x_{N-1} \tag{4}$$

This resultant value is divided by N to derive the mean of the signal.

$$|x_i - \mu| \tag{5}$$

The above expression describe how far the i^{th} sample deviation from the mean. Standard Deviation of the signal is found by summing the derivations of all the individual samples and then dividing by the number of samples *N*. The Standard Deviation is calculated from,

$$\sigma^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}(x_i-\mu)^2 \quad (6)$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N-1}(x_i-\mu)^2} \quad (7)$$

where, σ is Standard Deviation, *N* is number of samples.

Mean value describes what is being measured and Standard Deviation represents noise and other interference in signal.

5. SPECTRAL SUBTRACTION

Spectral subtraction is used to subtract estimated Standard Deviation of the noise spectrum from the spectrum of noisy speech. The signal received is the sum of clean speech and noise.

$$s(n) = x(n) + d(n) \quad (8)$$

where, $s(n)$ is noisy speech signal, $x(n)$ is clean speech signal and $d(n)$ is noise signal. Once it is framed and transformed into frequency domain it is represented as,

$$s_i(k) = x_i(k) + d_i(k) \quad (9)$$

where, k ranges over 1-256 samples and i ranges over the number of frames. Magnitude spectrum of noisy speech $s(n)$ is represented as,

$$M_i(k) = |S_i(k)| \quad (10)$$

Likewise magnitude of each frame is calculated.

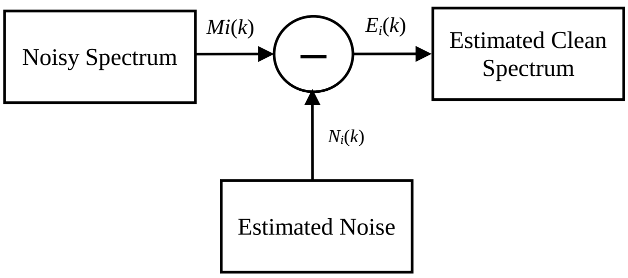

Fig.2. Spectral subtraction

This magnitude and noise estimate which is already estimated through proposed method is used for Spectral subtraction.

$$E_i(k) = \begin{cases} M_i(k) - N_i(k) & \text{if } M_i(k) \geq N_i(k) \\ 0 & \text{if } M_i(k) < N_i(k) \end{cases} \quad (11)$$

where,

$E_i(k)$ represents the estimated clean spectrum

$M_i(k)$ represents the Noisy spectrum

$N_i(k)$ represents noise estimate of our proposed method

Spectral subtraction is restricted to use positive magnitude spectrum. Thus clean spectrum of each frame is estimated.

Estimated clean magnitude spectrum $E_i(k)$ is combined with phase spectrum $\theta_i(k)$

$$C_i(k) = E_i(k)e^{j\theta_i(k)}. \quad (12)$$

According to Boll [10], the phase value of noisy speech spectrum is sufficient to use for an estimation of clean speech signal after Spectral subtraction.

$$\phi_s(w) \approx \phi_y(w) \quad (13)$$

Inverse FFT of $C_i(k)$ results time domain frames to reconstruct enhanced speech signal.

6. PERFORMANCE EVALUATION

Noise signal have different properties and the way it affects the speech signal also varies. Hence the performance of the proposed method is tested with various Noisy speech samples. White noise, Pink noise and Factory noise are considered as background noises which were taken from the NOISEX-92 database. The Fig.3 shows the temporal results of Noisy speech, Clean speech and Estimated Speech signal. This Figure says Clean Speech Signal and Estimated Speech signal resembles same.

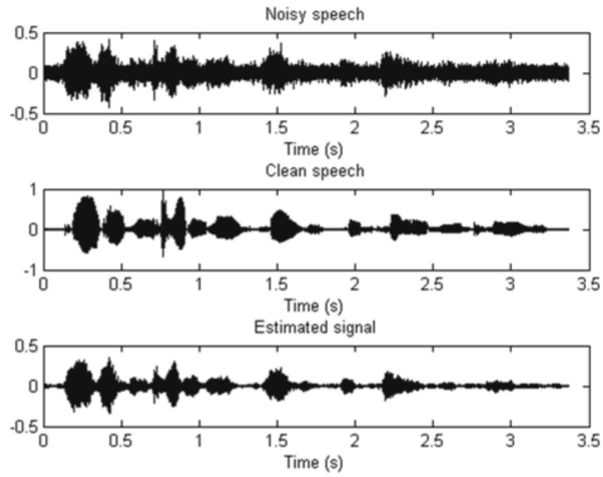

Fig.3. Noisy speech, clean speech and estimated speech signal

The performance of proposed method is evaluated through (i) SNR, (ii) Spectrogram, (iii) Perceptual Evaluation of Speech Quality, (iv) Time and Frequency Signal to Noise Ratio.

6.1 SIGNAL TO NOISE RATIO (SNR)

As speech is processed in frames, SNR is computed for each segment. SNR value for i^{th} segment is defined as defined as,

$$SNRi = 10\log\frac{\sum_{n=0}^{M-1}s_i^2[n]}{\sum_{n=0}^{M-1}n_i^2[n]} \quad (14)$$

where, S_i and n_i are speech and noise sample in the i^{th} segment. Averages of these segments are calculated as SNR value of speech signal.

SNR value for Noisy Speech signal (Pink, White and Babble) are compared with its corresponding clean speech signal. The Table.1 shows that the SNR value for clean speech signal is high which results increase in hearing ability and intelligibility of speech signal.

Table.1. Performance Comparison

Type of Noise	Noise (dB)	SNR Before Noise Estimation	SNR After Proposed Noise Estimation
Noisy Speech With White Noise	5	3.24	8.85
Noisy Speech With Pink Noise	5	4.01	10.72
Noisy Speech With Factory Noise	5	3.12	9.12
Noisy Speech With Multiple Noise	5	2.17	9.89

6.2 SPECTROGRAM

Spectrogram is a visual quality comparison of speech signal [16].The spectrogram of noisy speech signal and enhanced speech signal is shown in the Fig.4.

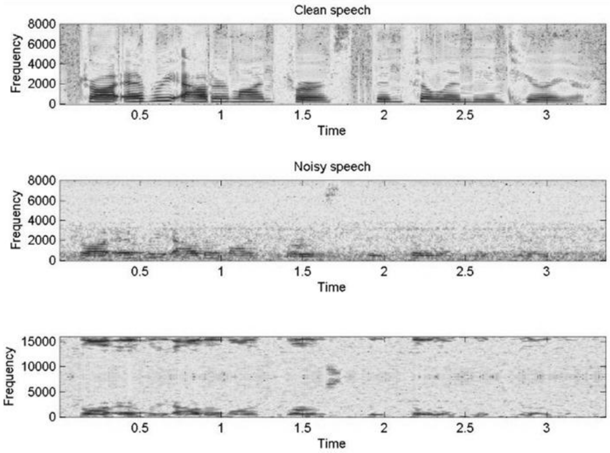

Fig.4. Spectrograms of Noisy Speech, Clean Speech and Estimated Speech

6.3 PERCEPTUAL EVALUATION OF SPEECH QUALITY (PESQ)

The PESQ is the most complex objective measure to compute the Quality of speech. This evaluation method focus on end-to end behaviour including the effects of filtering, listening frequency and the perceived loudness.

The PESQ provides a speech quality score based on the comparison between the original speech signal and the signal degraded by the network under test. Psychoacoustic and Cognitive models are used to perform this test.

6.4 TIME AND FREQUENCY SIGNAL TO NOISE RATIO

The segmental signal to noise ratio is evaluated either in terms of time or frequency domain. Segmental signal to Noise ratio is measured using original speech signal and processed signal. It is calculated through,

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\log\frac{\sum_{n=N_m}^{N_m+N-1}x^2(n)}{\sum_{n=N_m}^{N_m+N-1}\left(x(n)-\hat{x}(n)\right)^2} \quad (15)$$

where, $x(n)$ is the original signal, $\hat{x}(n)$ 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 yields the frequency weighted segmental SNR. f_wSNR_{seg} is given by,

$$fwSNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{k}B_j\log_{10}\left[\frac{F^2(m,j)}{\left(F(m,j)-\hat{F}(m,j)\right)^2}\right]}{E_{j=1}^{k}B_j} \quad (16)$$

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, $F(m, j)$ is the filter bank amplitude of the original signal and $\hat{F}(m,j)$ is the filter bank amplitude of resultant speech signal.

7. CONCLUSION

This paper proposed a new method for noise estimation which suits for noise suppression Spectral Subtraction algorithm. Noise is estimated by fixing the threshold value which separates noise range from speech signal range. Unlike other methods, Noise estimation did not depend on specific noise and it suits for any specific noise. From the experimental results and performance of SNR and Spectrogram, noise estimation method effectively supports for noise suppression.

REFERENCES

[1] K.P. Hymavathy and P. Janardhanan, "Noise Filtering in Speech Using Frequency Response Masking Technique", *International Journal of Emerging Trends in Engineering and Development*, Vol. 2, No. 3, pp. 208-213, 2013.

[2] Supavit Muangjaroen and Thaweesak Yingthawornsuk, "A Study of Noise Reduction in Speech Signal Using FIR Filtering", *Proceedings of International Conference on Advances in Electrical and Electronics Engineering*, pp. 51-54, 2012.

[3] T.L. Kumar and K.S. Rajan, "Noise Suppression in Speech Signals Using Adaptive Algorithms", *International Journal of Engineering Research and Applications*, Vol. 2, No. 1, pp. 718-721, 2012.

[4] Rajeev Aggarwal, Jai Karan Singh, Vijay Kumar Gupta, Sanjay Rathore, Mukesh Tiwari and Anubhuti Khare, "Noise Reduction of Speech Signal Using Wavelet Transform with Modified Universal Threshold", *International Journal of Computer Applications*, Vol. 20, No. 5, pp. 14-19, 2011.

[5] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics", *IEEE*

Transactions on Speech Audio Processing, Vol. 9, No. 5, pp. 504-512, 2001.

[6] Israel Cohen, "Speech Enhancement Using a Noncausal a Priori SNR Estimator", *IEEE Signal Processing Letters*, Vol. 11, No. 9, pp. 725-728, 2004.

[7] Sundarrajan Rangachari and C.L. Philipos, "A Noise-Estimation Algorithm for Highly Non-Stationary Environments", *Speech Communication*, Vol. 48, No. 2, pp. 220-231, 2006.

[8] Gerhard Doblinger, "Computationally Efficient Speech Enhancement by Spectral Minima Tracking in Subbands", *Proceedings of European Conference on Speech Communication and Technology*, pp. 1513-1516, 1995.

[9] I. Cohan and B. Berdugo, "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement", *IEEE Signal Processing Letters*, Vol. 9, No. 1, pp. 12-15, 2002.

[10] S. 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.

[11] P. Lockwood and J. Boudy, "Experiments with a Nonlinear Spectral Subtraction (NSS), Hidden Markov Models and Projection for Robust Recognition in Cars", *Speech Communication*, Vol. 11, No. 2-3, pp. 215-228, 1992.

[12] Hidetoshi Nakashima, Yoshifumi Chisaki, Tsuyoshi Usagawa and Masanao Ebata, "Spectral Subtraction based on Statistical Criteria of the Spectral Distribution", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E85-A, No. 10, pp. 2283-2292, 2002.

[13] H.G. Hirsch and C. Ehrlicher, "Noise Estimation Techniques for Robust Speech Recognition", *Proceedings of International Conference of Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 153-156, 1995.

[14] Marc Karam, Hasan F. Khazaal, Heshmat Aglan and Cliston Cole, "Noise Removal in Speech Processing Using Spectral Subtraction", *Journal of Signal and Information Processing*, Vol. 5, No. 2, pp. 32-41, 2014.

[15] D. Deepa and A. Shanmugam, "Spectral Subtraction Method of Speech Enhancement Using Adaptive Estimation of Noise with PDE Method as a Preprocessing Technique", *ICTACT Journal of Communication Technology*, Vol. 1, No. 1, pp. 1-6, 2010

[16] M. Thirumarai Chellapandi and P. Kabilan, "Evaluation of Speech Enhancement in Noisy Conditions Using a Spectral Subtraction and Linear Prediction Combination", *ICTACT Journal of Communication Technology*, Vol. 3, No. 4, pp. 626-633, 2012.

[17] G. Poornapriya, R. Shanmugasundaram and N. Santhiyakumari, "Clatter Diminishing for Mobile Telephony", *Proceedings of International Conference of Innovations in Information, Embedded and Communication Systems*, pp. 1-4, 2015.

[18] Israel Cohen and Baruch Berdugo, "Speech Enhancement for Non-stationary noise environment", *Signal processing*, Vol. 81, No. 11, pp. 2403-2418, 2001.

[19] L. Lin, E. Ambikairajah and W.H Holmes, "Speech Enhancement for Non-stationary Noise Environment", *Proceedings of Asia–Pacific Conference on Circuits and Systems*, Vol. 1, pp. 177-180, 2002.

[20] Kotta monahar and Preeti Rao, "Speech Enhancement in Non-stationary Noise Environment Using Noise Properties", *Speech Communication*, Vol. 48, No. 1, pp. 96-109, 2006.

[21] Sundarrajan Rangachari, Philipos C. Loizou and Yi Hu, "A Noise Estimation Algorithm with Rapid Adaptation for Highly Non-stationary Environments", *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, Vol. 1, pp. 305-308, 2004.

[22] Soo-Jeong Lee and Soon-Hyob Kim, "Noise Estimation based on Standard Deviation and Sigmoid Function Using a Posterior Signal to Noise Ratio in Non-stationary Noisy Environments", *International Journal of Control, Automation and Systems*, Vol. 6, No. 6, pp. 818-827, 2008.