IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 7, Issue 2, Ver. I (Mar. - Apr. 2017), PP 41-46
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Speech Enhancement Using Spectral Flatness Measure Based Spectral Subtraction

Supriya.P.Sarvade [1], Dr.Shridhar.K[2]
[1](PG Student, Department of Electronics & Communication Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India)
[2](Professor, Department of Electronics and Communication Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India)

Abstract : *This paper is aimed to reduce background noise introduced in speech signal during capture, storage, transmission and processing using Spectral Subtraction algorithm. To consider the fact that colored noise corrupts the speech signal non-uniformly over different frequency bands, Multi-Band Spectral Subtraction (MBSS) approach is exploited wherein amount of noise subtracted from noisy speech signal is decided by a weighting factor. Choice of optimal values of weights decides the performance of the speech enhancement system. In this paper weights are decided based on SFM (Spectral Flatness Measure) than conventional SNR (Signal to Noise Ratio) based rule. Since SFM is able to provide true distinction between speech signal and noise signal. Spectrogram, Mean Opinion Score show that speech enhanced from proposed SFM based MBSS possess better perceptual quality and improved intelligibility than existing SNR based MBSS.*

Keywords *- Multi-Band Spectral Subtraction, Spectral Flatness Measure, Speech enhancement, SFM, MBSS.*

I. Introduction

Speech is often corrupted by background noise which leads to many negative effects when processing a degraded speech signal. Hearing Aids supported by speech enhancement algorithms help hearing loss people in understanding speech in various noisy environments [7] and lots of research is being carried out in this direction. Speech intelligibility and quality are very important for hearing loss people and can be improved by speech enhancement techniques [7,8]. The spectral subtraction method proposed by Boll [5] is a well-known single channel speech enhancement technique [1,2,3]. Wherein, basically an estimate of noise spectrum is subtracted from noisy speech spectrum to obtain an estimate of clean speech. An estimate of background noise spectrum is used to locate the regions possessing energy level higher than background noise. Higher energy in these regions will be either due to speech or else due to high energy noise components. From instantaneous energy alone, it is not possible to distinguish the two possibilities. Hence convectional SNR based rule fails to differentiate weather the high energy level in the bins is due to speech or due to noise components.

For this reason an effort has been made in this paper to exploit a spectral domain feature, Spectral Flatness Measure to discriminate between speech component and noise component. Tone has more peaks and valleys in its spectrum in comparison to flat spectrum of white noise. Since white noise has flat spectrum, hence one way to determine if the sound is tone or noise is by measuring how flat is its spectrum, which is given by SFM. Experimental results of enhanced speech obtained from proposed model show that signal possess better noise cancellation with improved intelligibility and perceptual quality than traditional SNR based MBSS.

II. Spectral Flatness Measure (SFM)

Spectral flatness [6] or tonality coefficient is the ratio of geometric mean to the arithmetic mean of the power spectrum. Arithmetic mean is average or mean of 'N' sequences whereas geometric mean is Nth root of their products. Therefore SFM is given as:

$$SFM = \left(\frac{Geometric\ mean}{Arithmetic\ mean}\right) = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{1}{N}\sum_{n=0}^{N-1} x(n)} \quad (1)$$

where x(n) is magnitude of bin number 'n'.

If power spectrum is flat (i.e. constant), then its arithmetic and geometric means are equal and hence SFM becomes equal to one. For a sharp spectrum, one or two components will be one's and rest all zero, making geometric mean zero intern value of SFM becomes zero. Hence value of SFM is zero for pure tone and is one for white noise. Usually SFM is measured on logarithmic scale and hence its values lie between -∞ and 0.

III. Proposed SFM Based Multi-Band Spectral Subtraction

Multi-band spectral subtraction, proposed by Kamath [4] is the simplest way of removing background noise. It is very hard for any of the speech enhancement algorithms to perform homogeneously over all types of noise [10] and hence algorithms are built under certain assumptions. Spectral subtraction assumes that noise is additive and uncorrelated with the speech signal and an estimate of noise is subtracted from the noisy speech signal to obtain estimate of clean speech.

Noisy speech signal can be represented as sum of clean speech and noise as:

$$y(n) = x(n) + d(n) \quad (2)$$

where x(n) is clean speech and d(n) is noise.

Since speech signal is non-stationary and changes rapidly, it is divided into smaller frames using windowing techniques where each frame seems to be constant allowing us to apply Short Time Fourier Transform (STFT) for further processing. Hamming window is preferred over rectangular for its smoothness at the edges which reduces distortion.

Neglecting cross spectral terms which is the product of noise and clean speech spectral terms [9], power spectrum of noisy speech signal can be approximately given as:

$$|Y(k)|^2 \approx |X(k)|^2 + |D(k)|^2 \quad (3)$$

where|X(k)| is the magnitude spectrum of clean speech and |D(k)| is magnitude spectrum of noise. An estimate of clean speech can be given as:

$$|\hat{X}(k)|^2 = |Y(k)|^2 - |\hat{D}(k)|^2 \quad (4)$$

Considering the practical fact that a colored noise corrupts the speech signal, multiband spectral subtraction is implemented wherein each frame is divided into 'M' bands of equal lengths and the amount of noise subtracted from each band is decided by a weighting factor α_i. An estimate of clean speech of i^{th} band is given as:

$$|\hat{X}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i|\hat{D}_i(k)|^2 \quad (5)$$

Improved spectral subtraction proposed by Berouti [1] where the resulted spectrum was prevented from going below spectral floor (minimum level) is given as:

$$\left|\hat{X}_i(k)\right|^2 = \begin{cases} |Y_i(\mathrm{k})|^2 - \alpha_i\left|\hat{D}_i(k)\right|^2, if \left|\hat{X}_i(k)\right|^2 > \beta\left|\hat{D}_i(k)\right|^2 \\ \beta\left|\hat{D}_i(k)\right|^2, otherwise \end{cases} \quad (6)$$

where the value of β is chosen to be 0.02

In the proposed model weighting factor α_i is driven by a noise-speech discriminating parameter, SFM than traditional Signal to Noise ratio. SFM in dB can be given as:

$$SFM = 10\, log_{10}\left(\frac{G_m}{A_m}\right) \quad (7)$$

whereG_m and A_m are geometric and arithmetic means of power spectrum respectively.

This paper proposes an empirical relationship between SFM and weighting factor α.For speech signal SFM of -60 dB represents a pure tone and a minimum value of noise power should be subtracted from the input noisy signal, hence a small value of α= 1 was chosen till SFM = -40dB as shown in Fig. 1. Whereas SFM of 0 dB represents complete noise and hence a maximum value of α = 2.5 was chosen. Applying a second order polynomial fit for the above data points, a relation between SFM and weighting factor α of i^{th} band can be given as:

$$\alpha_i = 0.00063\, SFM_i^2 + 0.063\, SFM_i + 2.5 \quad (8)$$

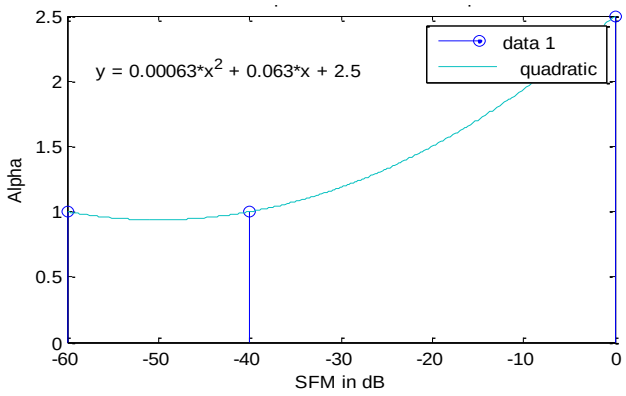

Fig. 1. Relationship between SFM and weighting factor α

IV. Block Diagram of Proposed Model

Block diagram of the proposed model is as shown in Fig. 2.

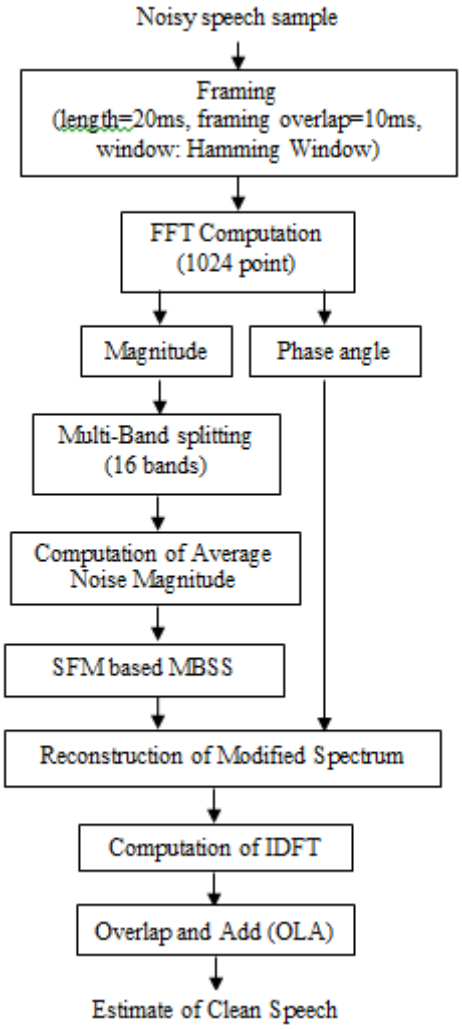

Fig. 2. Block diagram of proposed SFM based MBSS

The proposed SFM based MBSS can be implemented by following steps:

1) Initially speech signal is windowed. Since speech is a long signal, successive windows each of 20ms are taken along the length of the signal with an overlap of 50% so that the deemphasized part of one window becomes middle of the next window.
2) 1024 point Fast Fourier Transform (FFT) is computed on each frame which decomposes the signal into its magnitude and phase. FFT is a technique proposed by [14] that computes coefficients of a Discrete Fourier Series faster than ever it was possible [11,12].
3) Average noise spectrum is computed from speech pause periods. In the proposed work average of first 5 frames i.e. 100ms is considered as estimate of noise power.
4) Multiband concept is implemented by subdividing each frame into 16 bands of equal length.
5) SFM of each band is computed using equation 7.
6) In [13,15] it is revealed that amplitude is more important than the phase information for the quality and intelligibility of speech and hence in proposed model phase of the signal is kept unchanged. An estimate of clean speech is obtained by subtracting an estimate of noise power from each band of noisy speech magnitude as a function of weighting factor α_i using equations 6 and 8.
7) Estimate of clean speech magnitude is combined with the undisturbed phase and then is transformed to time domain by obtaining Inverse Fast Fourier Transform (IFFT).
8) Reverse process of framing is done using Overlap and Add (OLA) method and enhanced speech is obtained.

V. Results and Analysis

Proposed speech enhancement algorithm has been tested on different types of noisy speech samplestaken from NOIZEUS speech database. Performance evaluation of the system is done using both spectrogram analysis and subjective listening tests.

Mean Opinion Score (MOS) for different types of noise:

MOS is a measure of representing overall quality of the system. On a predefined scale of 1 to 5 subjects were asked to rate over the performance of the system, where 1 representing the lowest quality and 5 representing highest quality.

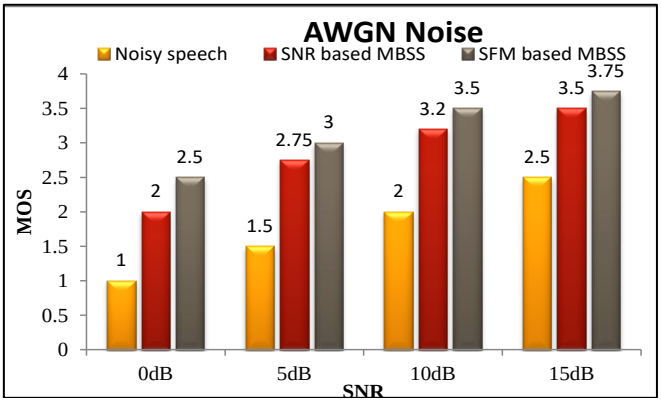

Fig . 3. MOS for Additive White Gaussian Noise (AWGN)

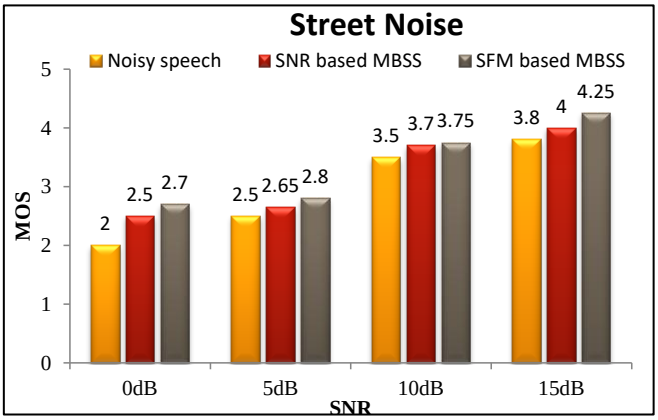

Fig . 4. MOS for street noise

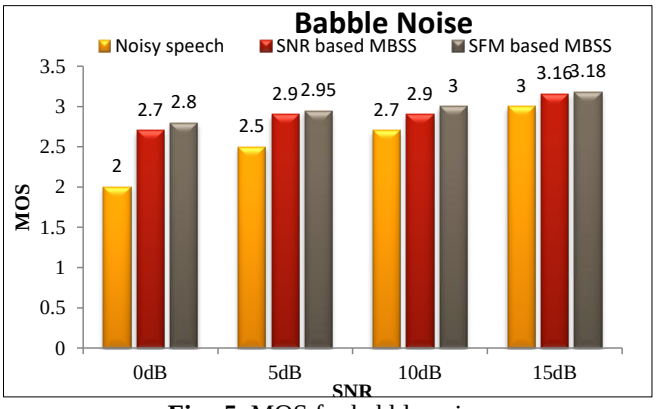

Fig . 5. MOS for babble noise

Fig. 3,4,5 shows MOS computed from listening tests for different kinds of noise sources with different SNR levels. It is observed that MOS decreases with increase in SNR of the noisy speech samples. It is also evident that MOS for proposed model is more than the traditional SNR based model.

Spectrogram Analysis for different types of noise:

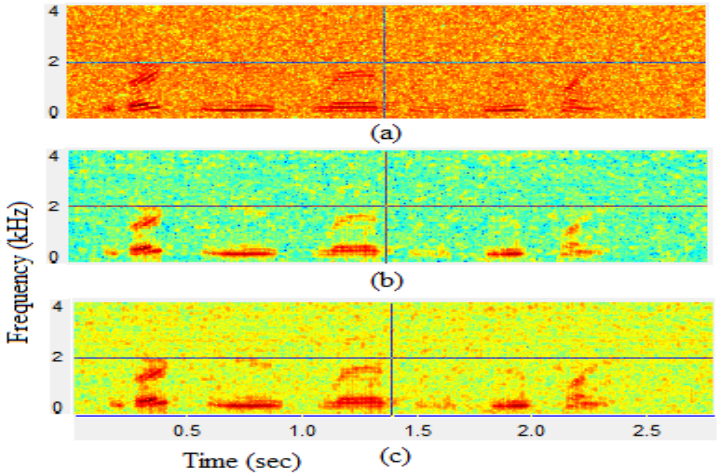

Fig . 5. Spectrograms for AWGN (a) 0 dB SNR noisy speech; (b),(c)Enhanced speech obtained using convectional SNR based rule and proposed SFM based rule respectively.

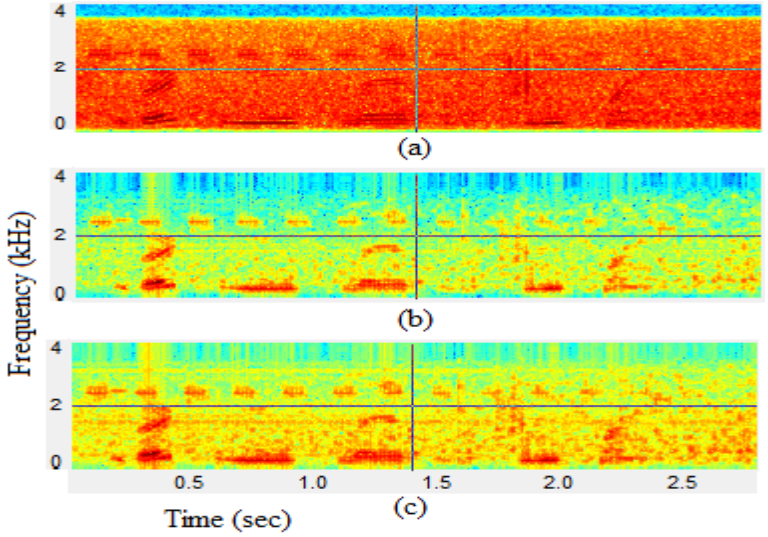

Fig . 6. Spectrograms for Street Noise (a) 0 dB SNR noisy speech; (b),(c)Enhanced speech obtained using convectional SNR based rule and proposed SFM based rule respectively.

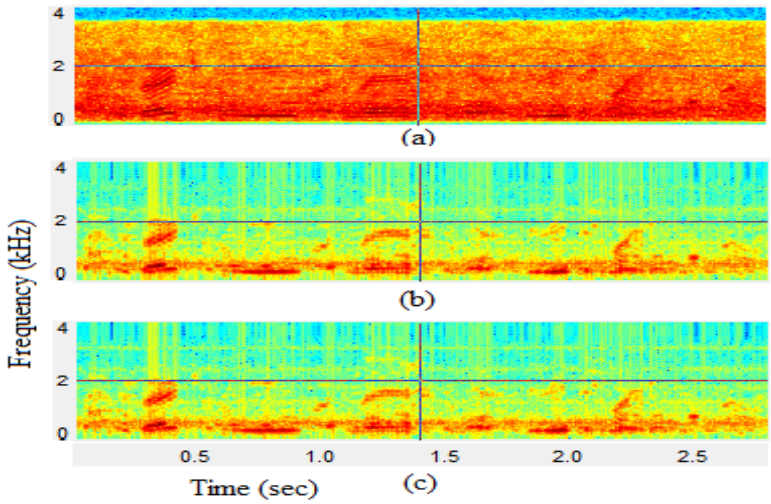

Fig . 7. Spectrograms for Babble Noise (a) 0 dB SNR noisy speech; (b),(c)Enhanced speech obtained using convectional SNR based rule and proposed SFM based rule respectively.

From spectrogram analysis of input noisy speech, enhanced speech obtained from traditional SNR based MBSS and enhanced speech obtained from proposed SFM based MBSS, it is evident that the performance of proposed model is superior that the existing SNR based model. Performance of proposed model is best for Additive White Gaussian Noise since model was designed under the assumption that additive noise corrupts the speech signal and performance of the proposed model decreases for babble noise since the frequency and characteristics of babble noise are very similar to the speech signal of interest.

VI. Conclusion

This paper intended to preserve the perceptual quality of speech by exploiting one of the spectral characteristic of noise called SFM. From results and analysis it can be concluded that the performance of proposed SFM based MBSS is superior than the traditional SNR based MBSS. Proposed model proved to have better noise cancellation preserving perceptual quality of the speech signal with minimum distortion and musical noise is nearly inaudible.

References

[1] M. Berouti, R. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise,"*Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*,pp. 208–211, April 1979.

[2] C.-T. Lin, "Single-channel speech enhancement in variable noise-level environment," *IEEE Trans. Syst. Man Cybernet.* A 33 (1) (2003) 137–143.

[3] Radu Mihnea Udrea, Nicolae D. Vizireanu, Silviu Ciochina, "An improved spectral subtraction method for speech enhancement using a perceptual weighting filter," *Elsevier Digital Signal Processing 18,* pp. 581-587, Aug 2007.

[4] S. Kamath, and P. C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in Proceedings of *Int. Conf. on Acoustics, Speech, and Signal Processing*, Orlando, USA, May 2002, vol. 4, pp. 4160 4164.

[5] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech*.

[6] GRAY, A.H., and MARKEL, J.D. "A spectral-flatness measure for studying the autocorrelation method of linear prediction of speech analysis," *IEEE Trans. Acoust. Speech Signal Process.*, 1974, 22, pp. 207–217.

[7] Dr. (Smt). S.D. Apte and Shridhar, "Speech Enhancement in Hearing Aids Using Conjugate Symmetry of DFT and SNR-Perception Models," *International Journal of Computer Applications*, vol. 1,no. 21, pp. 44-51, 2010.

[8] Dr. (Mrs). S.D. Apte, Shridhar, "Speech Enhancement in Hearing Aids Using Conjugate Symmetry Proprety of Short Time Fourier Transform," *International Journal of Recent Trends in Engineering*, vol. 2, no. 5, pp. 346-351, November 2009.

[9] Soumya Jolad, Shridhar,"Speech Enhancement Using Spectral Subtraction Technique with Minimized Cross Spectral Components," *International Journal of Research in Engineering and Technology,* vol. 5, no.3, pp. 197-200, March 2016.

[10] Supriya.P.Sarvade, Dr.Shridhar. K and Varun.P.Sarvade, "Multi-Band Spectral Subtraction for Speech Enhancement Using Sine Multitaper," *IOSR Journal of VLSI and Signal Processing,*vol. 6, issue 6, ver. II, pp. 70-76, Nov.-Dec. 2016.

[11] Supriya.P.Sarvade, Dr.Shridhar. K and Varun.P.Sarvade, "Radix-2 DIT-FFT Algorithm for Real Valued Sequence," *International Journal of Emerging Trends in Science and Technology,* vol. 3, issue 2, pp. 3534-3536, Feb. 2016.

[12] Supriya.P.Sarvade, Dr.Shridhar. K and Varun.P.Sarvade, "Time Efficient Structure for DFT Filter Bank," *International Journal of Emerging Trends in Science and Technology,* vol. 3, issue 11, pp. 4791-4794, Nov. 2016.

[13] J. S. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech,"*Proceedings of the IEEE,* vol. 67, pp. 1586–1604, (1979).

[14] W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Coinput,* vol. 19, pp.297—301, 1965.

[15] P. C. Loizou, "Speech Enhancement: Theory and Practice,"*Ist ed. Taylor and Francis,* (2007).