IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 6, Issue 6, Ver. II (Nov. - Dec. 2016), PP 70-76
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Multi-Band Spectral Subtraction for Speech Enhancement Using Sine Multitaper

Supriya.P.Sarvade[1], Dr.Shridhar.K[2], Varun.P.Sarvade[3]
[1](PG Student, Department of Electronics & Communication Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India)
[2](Professor, Department of Electronics & Communication Engineering, Basaveshwar Engineering College, Bagalkot, Karnataka, India)
[3](Assistant Professor, Department of Computer Science & Engineering, Biluru Gurubasava Mahaswamiji Institute of Technology Mudhol, District Bagalkot, Karnataka, India)

Abstract: *This paper intends to reduce musical noise, a bi-product of subtractive type methods of speech enhancement technique which degrades the quality of the speech signal. Hamming Window is most commonly used in Speech Enhancement systems to estimate power spectrum, but it has high variance which in turn degrades the speech quality. To reduce this high variance, Multitaper window is used in the proposed model as a spectral estimator. Reduction in the variance of spectral estimate of noise improves the performance of spectral subtraction algorithms. A Multi-Band Spectral Subtraction approach is exploited by employing different weights to signal bands so as to take into account the fact that colored noise affects the speech spectrum differently at various frequencies. Choice of optimal values of weights adversely affects the spectral variance at different frequency bands. Informal listening tests over NOIZEUS database show that speech reconstructed in this way has little speech distortion and musical noise is nearly inaudible.*

Keywords: *Multi-band Spectral Subtraction, Multitaper Spectral Estimation, Musical Noise, Speech Enhancement, Variance*

I. Introduction

Speech enhancement aims at improving speech quality through various algorithms. Enhancing of speech degraded by noise, or noise reduction, is the most important field of speech enhancement and used for many applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids. The methods for enhancing the speech include removal of background noise, echo suppression and many more. Few of speech enhancement methods are [12]: Spectral subtraction, Minimum Mean Square Estimation, Signal subspace, Linear Minimum Mean Square Estimation, Kalman filtering and Perceptual property of human auditory system.

The spectral subtraction method proposed by Boll [2] was chosen for proposed model for its relative simplicity. It is a noise reduction technique for enhancing speech by degrading noise. It estimates the spectral amplitude of the clean speech by subtracting an estimate of the noise spectral amplitude from that of the observed noisy speech [2][3]. Mainly there are two problems in spectral subtraction [13], one is generation of musical noise and other is presence of cross spectral terms which results in inefficient noise reduction.

In this paper, an effort has been made to reduce Musical noise which is a product of the variance in the spectral estimates. Musical noise occurs when isolated peaks are left in a spectrum after processing with a spectral subtractive type algorithm. In speech pause sections, these isolated components sound like musical tones to our ears. This paper presents segmental SNR method to eliminate musical noise. Concept of SNR depends on the noise estimation technique which is critical in speech enhancement system and here is done from multitaper spectral estimator.

II. Multitaper Spectral Estimation (MTSE)

MTSE is a technique developed by David J. Thomson [1]. Direct spectrum estimation using Hamming window is most commonly used to estimate power spectrum even though windowing will reduce bias but has high variance. The idea behind the multitaper spectrum estimator is to reduce this variance by obtaining multiple independent estimates from the same sample. Taper is multiplied to the signal element wise providing a sub spectral signal. As each taper is orthogonal to all other tapers, sub spectral signals provide statically independent estimate of spectrum thereby minimizing spectral leakage which exists in the finite length data set. Then final spectrum is obtained by averaging over all tapered spectra. Estimation errors in sub spectra will be approximately uncorrelated, which is a key to variance reduction.

In the proposed model Sine Taper is used as spectral estimator. Table 1 gives a comparison of bias and variance for convectional Rectangular and Hamming Window with the Sine Multitaper on signal x(t)=2cos(2π*5000t).

Table 1. Performance measure of conventional windowing techniques and multitaper

	Bias	**Variance**
Rectangular Window	13.65	3.99e04
Hamming Window	2.76	1.59e04
Sine Multitaper	0.04	2.58

Table 1 show that Multitaper has largely reduced the variance. Simulation results showed that Spectrum estimated from Rectangular Window failed to distinguish the frequency components separated by 200Hz. Hamming Window could distinguish the frequencies separated by 50Hz, but failed to distinguish frequencies below 50Hz. Whereas Spectrum estimated from Sine taper was able to distinguish for the smallest separation of 10Hz as shown in Fig. 1. This clearly shows that multitapers possess higher resolution than Rectangular and Hamming Windowing techniques.

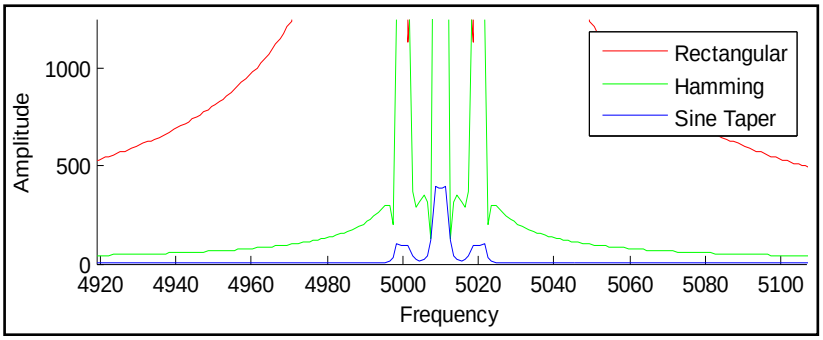

Fig. 1. Magnitude plot of signal x(t)=5cos(2π*5000t)+10sin(2π*5010t)

The variance of the multitaper estimate will be smaller than the variance of each spectral estimate by a factor of 1/k [6], where 'k' is number of tapers. 'k' orthogonal sine tapers are generated using equation given as [1]:

$$\alpha_k(n) = sin\frac{\pi k(n+1)}{(N+1)}\sqrt{\frac{2}{N+1}}, \qquad 0 < n < N-1 \tag{1}$$

where, N is signal length and a_k is k^{th} sine taper

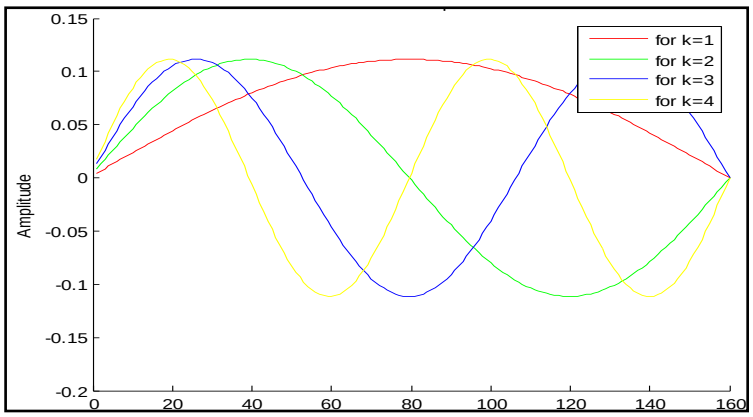

Fig. 2. Sine tapers for k=4 and N=160

'k' value should be chosen carefully, since for higher order windows leakage from side lobes increases resulting in decrease in the resolution. If 'k' value is chosen to be small then there will be not enough averaging to reduce the noise variance.In [8] it was found that multitaper performs the best for k=4 and performed worst for k=8. Thus for proposed model 'k' value has been chosen to be equal to 4.

III. Multiband Spectral Subtraction

Simplest method for getting rid of background noise is spectral subtraction. Multiband spectral subtraction was proposed by Kamath [4]. It is very hard for any speech enhancement algorithms to perform homogeneously over all noise types. For this reason algorithms are built on certain assumptions. Spectral subtraction algorithm of speech enhancement is built under the assumption that the noise is additive and is uncorrelated with the signal which is true for few kinds of noise like background noise and interfering noise. Noise corrupted speech signal can be represented as:

$$\boldsymbol{y(n) = s(n) + d(n)} \tag{2}$$

where s(n) is clean speech and d(n) is noise. Since speech signal is non-stationary, it is split into smaller frames and is assumed to be quasi stationary which allows to apply STFT for signal processing [9][10][11].The power spectrum of noise corrupted speech signal can be approximately estimated as:

$$|Y(k)|^2 \approx |S(k)|^2 + |D(k)|^2 \tag{3}$$

where S(k) and D(k) are magnitude spectra of clean speech and noise respectively. Noise spectrum cannot be obtained directly. An estimate $\widehat{D}(k)$ is calculated from sine tapers spectrum during periods of silence. Estimate of clean speech is obtained as:

$$|\hat{S}(k)|^2 = |Y(k)|^2 - \alpha|\widehat{D}(k)|^2 \tag{4}$$

where α is an over-subtraction factor (weights) which is a function of the segmental SNR [7].
Berouti [7] proposed a modification in spectral subtraction where the resulted spectrum is prevented from going below a minimum level (spectral floor) and is expressed as:

$$|\hat{S}(k)|^2 = \begin{cases} |Y(k)|^2 - \alpha\left|\widehat{D}(k)\right|^2, \text{if } |\hat{S}(k)|^2 > \beta|\widehat{D}(k)|^2 \\ \beta|\widehat{D}(k)|^2, \text{otherwise} \end{cases} \tag{5}$$

Multi-band spectral subtraction is implemented to take into account the fact that colored noise affects the signal at different frequencies. Complete signal is divided into bands of equal length which do not overlap with each other. Estimate of clean speech of i^{th} band is given as:

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i|\widehat{D}_i(k)|^2 \tag{6}$$

The band specific over subtraction factor α_i is a function of the segmental SNRi of the ith frequency band which is calculated as:

$$\mathrm{SNR}_i(\mathrm{dB}) = 10\log_{10}\frac{\sum_{k=0}^{N-1}|Y_i(k)|^2}{\sum_{k=0}^{N-1}|\widehat{D}_i(k)|^2} \tag{7}$$

Over subtraction factor α_i can be calculated as:

$$\alpha_i = \begin{cases} 4 + \frac{3}{4} & ; \quad \mathrm{SNR} \le -5\mathrm{dB} \\ 4 - \frac{3}{20}\mathrm{SNR}_i & ; \quad -5\mathrm{dB} \le \mathrm{SNR} \le 20\mathrm{dB} \\ 0 & ; \quad \mathrm{SNR} \ge 20\mathrm{dB} \end{cases} \tag{8}$$

IV. Implementation Details

1. **Frame size and overlap:** Frame size can be anywhere between 5 to 50 milliseconds. In [7] it was found that frames shorter than 20-ms result in roughness, and increasing the frame size although decreases the musical noise, might also result in slurring. Thus, 20-ms was chosen as the optimum frame size. The amount of overlap between consecutive frames is also associated with the frame-size and is required to prevent discontinuities at frame boundaries. For the proposed model overlap is chosen to be 50%.
2. **Window type:** Here windowing is done using 4 Sine Tapers, then is averaged over all 4 tapers.
3. **FFT Size:** It is generally possible that while reconstructing the signal after modifications made by enhancement application, distortion may occur due to aliasing if the time-domain signal is not augmented with sufficient number of zeros prior to performing the DFT. Hence sufficiently large FFT of 1024 point was computed over all frames.

4. **Speech/Noise detection and Noise Estimation:** Since noise is estimated during non-speech periods in the proposed method, it requires a robust speech/noise detector. The initial estimate of noise statistics is made from averaging over a few frames of silence. In this implementation five frames (100-ms) of silence period were added. Since the noise is stationary, the estimate of the noise can be updated during speech pauses.

V. Block Diagram of Proposed Model

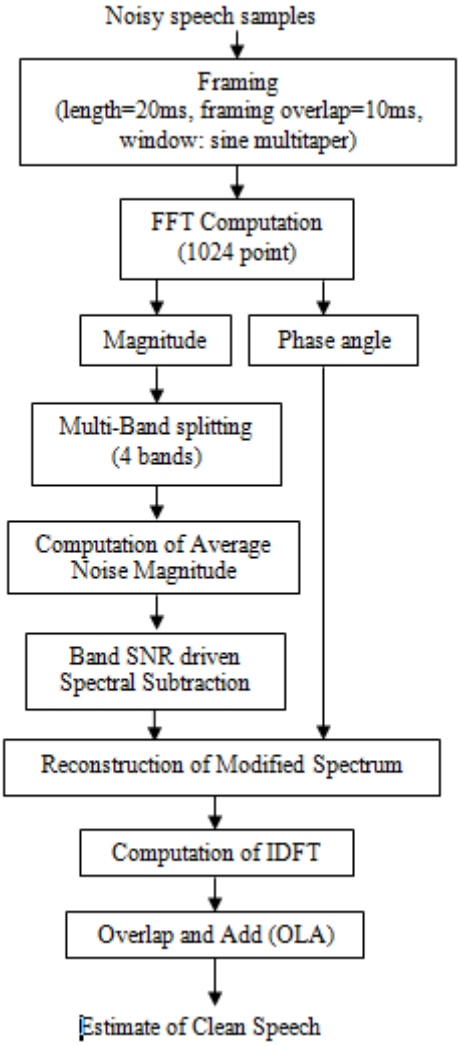

Fig. 3. Block diagram of proposed model

The proposed model can be implemented by following steps:

1) Generation of four orthogonal sine tapers using equation 1 of 20-ms duration.
2) Speech was windowed with all four sine taper of 20-ms window length and overlap of 10-ms between the frames.
3) Fast Fourier Transform (FFT) on windowed speech is obtained then spectrum is averaged over four tapers.
4) Average noise spectrum is estimated from the periods where signal is absent and only noise is present, i.e. first 100-ms duration of noisy speech signal.
5) Each frame is divided into 4 sub bands. SNR of each sub band is calculated using equation 7.
6) Estimate of clean speech is obtained by subtracting estimate of average noise spectrum from each sub band as a function of over subtraction factor α_i as given in equation (8).
7) Estimate of magnitude spectrum is combined with the phase of the noisy speech signal and then transformed to time domain using Inverse Fast Fourier Transform (IFFT). Phase of the noisy signal is not altered since from the perceptual point of view human ear is not sensitive to change in the phase.
8) De-framing is done using overlap and add method.

VI. Results and Analysis

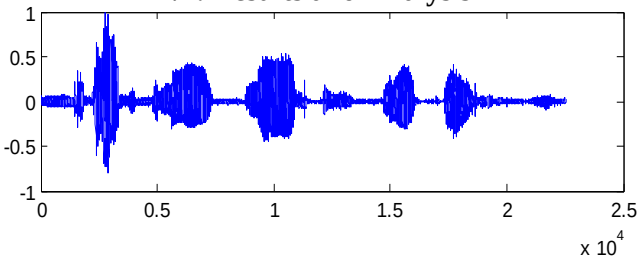

Fig. 4. Clean Speech

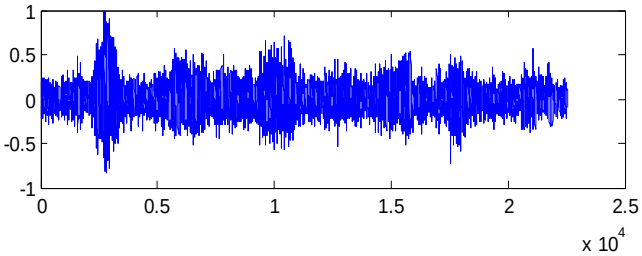

Fig. 5. Input Noisy Speech at SNR 0dB

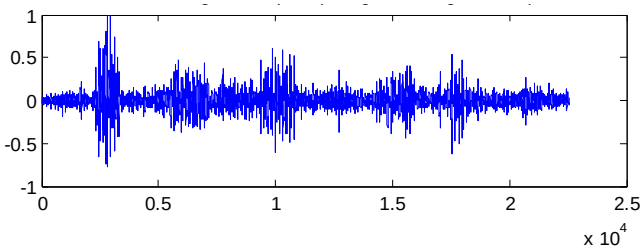

Fig. 6. Enhanced Speech using Hamming Window

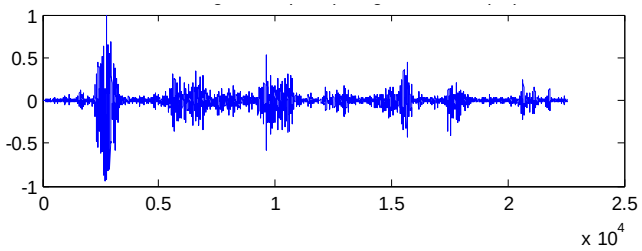

Fig. 7. Enhanced Speech from Proposed model

The proposed speech enhancement algorithm has been tested on the spoken English sentence which has been chosen from NOIZEUS database. Fig. 4 is the clean speech sample on the sentence "The birch canoe slid on the smooth planks" with a sampling rate of 8 kHz and spoken by a male speaker. Fig. 6 is the enhanced speech signal for convectional multiband hamming window technique with a noisy speech input with SNR 0dB as shown in Fig. 5. Whereas Fig. 7 shows enhanced speech from proposed multiband multitaper based spectral subtraction for the same noisy input speech input of 0 dB SNR, wherein noise has been reduced to a larger extent in comparison with the hamming window technique.

Performance evaluation of proposed model was also done using spectrogram analysis as shown in figures. Spectogram shown in Fig. 9 is speech corrupted by noise at 0 dB SNR, wherein Fig. 8 is spectrogram of clean speech with SNR 15dB. It can be observed that proposed model presented in Fig. 11 gave better results in noise cancellation in comparison to hamming window method shown in Fig. 10.

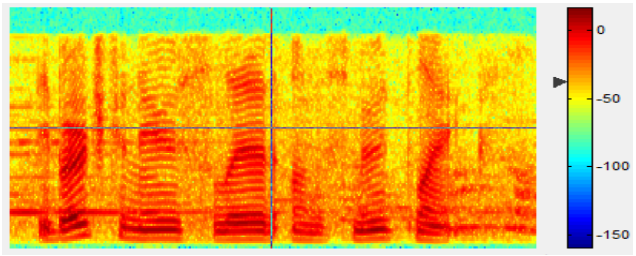

Fig. 8. Spectogram of Clean Speech

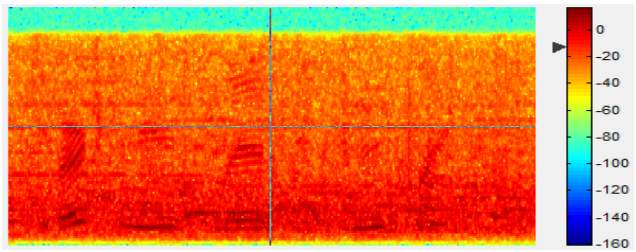

Fig. 9. Spectogram of Input Noisy Speech at SNR 0 dB

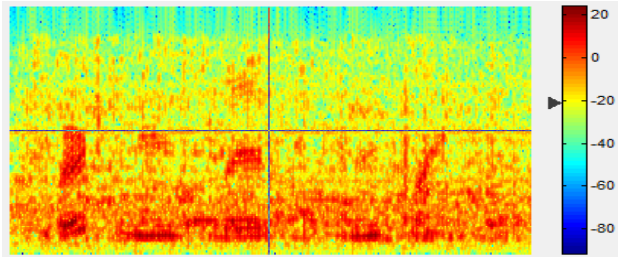

Fig. 10. Spectogram of Enhanced Speech using Hamming Window

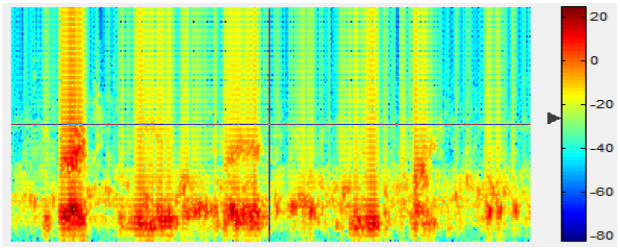

Fig. 11. Spectogram of Enhanced Speech from Proposed model

VII. Conclusion

The main attraction of spectral subtraction is its relative simplicity, in that it only requires an estimate of the noise power spectrum. The multitaper based spectral subtraction method provides a definite improvement over the conventional hamming window. Listening tests showed that the musical noise was eliminated with the multitaper based spectral estimation because of the fact that multitapers reduced variance. The main problem in spectral subtraction is the presence of processing distortions caused by the random variations of the noise, and fails for colored noise and this problem has been solved using proposed non-linear Multiband Spectral Subtraction algorithm.

References

[1] Thomson, D.J., "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, 70, 1055-1096, 1982.
[2] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech.*
[3] H. Gustafsson, S. Nordholm, I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Trans. Speech Audio Process.* 9(2001) 799–807.
[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] D. B. Percival and A. T. Walden, "Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques," Cambridge, MA: Cambridge Univ. Press, 1993.
[6] Yi Hu, and Philipos C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Trans. Speech Audio Process*, vol. 12, pp.59-67 January 2004.

[7] M. Berouti, R. Schwartz and J. Makhoul, "Enhancementof speech corrupted by acoustic noise," *Proc.IEEE Int. Conf. Acoust., Speech, Signal Process*., pp.208-211, Apr. 1979.
[8] Selin Aviyente, William J. Williams, "Multitaper marginal time–frequency distributions," *Elsevier Signal Processing*, pp.279–295, June 28, 2005.
[9] J.Allen and L.Rabiner, "A Unified approach to short-timeFourier analysis and synthesis," Proc. IEEE,vol. 65, no.11, pp. 1558 - 1564, Nov.1977.
[10] 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.
[11] Bagher BabaAli, Hossein Sameti, andMehran Safayani, "Likelihood-Maximizing-BasedMultiband Spectral Subtractionfor Robust Speech Recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 1-15, 19 January 2009.
[12] 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.
[13] 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.