International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.8, August 2013 DOI:10.15693/ijaist/2013.v2i8.22-25

Modified Spectral Subtraction Scheme for Enhancing Speech Signal

Purushotham.U,
Research Associate
Department of Electronics and Communication
Dayananda Sagar College of Engineering
Bangalore, India -560078

Suresh.K,
Principal
SDM Institute of Technology
Ujire, Dakshina Kannada, India-574240

***Abstract*—** Development of preprocessing algorithms for speech enhancement is always of great interest. This research area is extensively utilized in Mobile phones. The Performance of these devices is fine in practically noise-free conditions, but their performance deteriorates rapidly in noisy conditions. The speech signal to be transmitted may be polluted by echo and background noise. Hence in numerous cases, there exists a need for digital voice communications, human-machine interfaces, and automatic speech recognition systems to perform reliably in noisy environments. The objectives of speech enhancement vary according to specific applications, such as to boost the overall speech quality, to increase intelligibility, and to improve the performance of voice communication devices. Most of the speech enhancement algorithms rely on frequency domain weighting approach, commonly consisting of a noise spectral density estimator and a spectral amplitude estimator. But, the nature of speech signals and restriction on algorithmic delay and complexity require that the noisy signal must be processed in short frames. In this paper we propose a method of estimation based on, filtering frequency components of noisy speech signal using multiband filters. The time-varying components of speech are modeled by autoregressive process, incorporated in the multiband filter. Experimental results show that the proposed method provides the better performance as compared to the other approach.

Key words – Amplitude estimator, multiband, Autoregressive

INTRODUCTION

In speech communication, the speech signal is always accompanied by a quantity of noise. The interfering noise generally degrades the quality and intelligibility of speech. In most cases the background noise of the environment where the source of speech lies, is the main element of noise that adds to the speech signal. Though the apparent effect of this noise addition makes the listening job difficult for a direct listener, an associated problem is processing degraded speech in preparation for coding by a bandwidth compression system. Hence speech enhancement not only involves processing speech signals for human listening but also for further processing prior to listening. The spectral subtraction method as proposed by Boll [1] is based on the direct estimation of the short-term spectral magnitude. The basic principle of the spectral subtraction method is to subtract the magnitude spectrum of noise from that of the noisy speech. The noise is assumed to be uncorrelated and additive to the speech signal. The usual power spectral subtraction method substantially reduces the noise levels in the noisy speech [2]-[8]. However, it also introduces an annoying distortion in the speech signal called musical noise. Due to the inaccuracies in the short-time noise spectrum estimate, large spectral variations exist in the enhanced spectrum causing these distortions. The negative spectral components are floored to zero or to some minimal value, causing further distortions in the time signal.

Unlike white Gaussian noise, which has a flat spectrum, the spectrum of real-world noise is not flat. Thus, the noise signal does not affect the speech signal uniformly over the whole spectrum. Some frequencies are affected more adversely than others [9]. In multi-talker babble, for instance, the low frequencies, where most of the speech energy resides, are affected more than the high frequencies. Hence it becomes essential to estimate a suitable factor that will subtract just the necessary amount of the noise spectrum from each frequency bin, to prevent destructive subtraction of the speech while removing most of the residual noise. In this paper, we propose a multi-band approach to the spectral subtraction method that accomplishes just that, i.e., it reduces the above-mentioned distortions to a large extent while maintaining a high level of speech quality. Section II of the discusses about some of the usual approaches for speech enhancement, Section III Presents the Proposed technique, describes the implementation of the proposed method, section IV gives the Experimental results, Conclusions and comments are given in section V.

APPROACHES FOR SPEECH ENHANCEMENT

Approaches to retrieve enhanced speeches are plentiful. Among them the spectral subtraction methods are the most widely used due to the simplicity of implementation and also low computational load makes them the primary choice for real time Applications. In general, using the family of Subtraction type algorithms, the enhanced speech spectrum is obtained by subtracting an average noise spectrum from the noisy speech spectrum. The phase of the noisy speech is kept unchanged, since it is assumed that the phase distortion is not perceived by human ear. However, the spectrum of real world

noise is not stable. Thus, the noise signal does not affect the speech signal uniformly over the whole spectrum. Some frequencies are affected more adversely than others. Hence it becomes imperative to estimate a scale factor that will subtract just the necessary amount of the noise spectrum from each frequency bin, to prevent destructive subtraction of the speech while removing most of the residual noise.

Sub band filtering approach proposed in [8] was basically used to increase convergence speed in comparison to a full band solution. This is due to reduced spectral magnitude range, i.e. sub band filtering has a de-correlating effect because colored input signals are decomposed into sub bands with "whiter" sub–spectra. Ephraim and Malah [10] have proposed a system that utilizes the MMSE criteria using models for the distribution of the spectral components of speech and noise signals. The MMSE-short time spectral amplitude (STSA) estimator for speech enhancement aims to minimize the mean square error between the short time spectral magnitude of the clean and enhanced speech signal. MMSE-LSA estimator for speech enhancement was also proposed by Ephraim and Malah in [11]. Both the proposals do not consider any of the nonlinear characteristics observable in human perception [12].

I. PROPOSED METHODOLOGY

Since colored noise affects the speech spectrum differently at various frequencies [13]-[15], we propose a Modified approach spectral subtraction approach. The speech spectrum is divided into L non-overlapping bands, and spectral subtraction is performed independently in each band. So, the estimate of the clean speech spectrum in the i^{th} band is obtained by:

$$|S_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \delta_i |D_i(k)|^2 \quad \text{---- (1)}$$

k is chosen between the beginning and ending frequency bins of the i^{th} frequency band, α_i is the over-subtraction factor of the i^{th} band and δ_i is a tuning factor that can be individually set for each frequency band to customize the noise removal properties. The band specific over subtraction factor α_i is a function of the segmental SNR_i of the i^{th} frequency band which is calculated as:

$$SNR_i(db) = 10\ log \frac{\sum |Y_i(k)|^2}{\sum |D_i(k)|^2} \quad \text{---- (2)}$$

While the use of the over-subtraction factor α_i provides a degree of control over the noise subtraction level in each band, the use of multiple frequency bands and the use of the δ_i weights provide an additional degree of control within each band. The value of α_i empirically determined and set to

$$\delta_i = \begin{cases} 1 & f_i \le 1\ kHz \\ 2.5 & 1\ kHz < f_i \le \frac{F_s}{2} - 2\ kHz \\ 1.5 & f_i > \frac{F_s}{2} - 2\ kHz \end{cases} \quad \text{----(3)}$$

where f is the upper frequency of the ith band, and $_s F$ is the sampling frequency. Speech was hamming windowed using an 18-ms window and a 9-ms overlap between frames. The Fast Fourier Transform of the windowed speech was smoothed and a weighted spectral average is taken over preceding and succeeding frames of data as:

$$\bar{Y}_j(k) = \sum_{I=-M}^{M} W_I Y_{j-I}(k) \quad \text{----(4)}$$

Where j is the frame index. The number of frames, M was limited to 2. The motivation for using smaller δ_i values for the low frequency bands is to minimize speech distortion, since most of the speech energy is present in the lower frequencies. Relaxed subtraction was also used for the high frequency bands. The enhanced spectrum within each band is combined, and the enhanced signal is obtained by taking the inverse Fourier transform of the enhanced spectrum using the phase of the original noisy spectrum. Finally, the standard overlap-and-add method is used to obtain the enhanced signal.

IV.EXPERIMENTAL RESULTS

Two sentences from the IEEE database [16] spoken by a male and female speaker were used to evaluate the proposed spectral subtraction approach. Speech signal of 8 kHz sampling frequency at 5 dB and 10 dB SNR with car noise is considered for both male speaker and female speaker. This method ensures a reasonable overall measure of performance [17-[20]. To determine the speech quality, number of bands, where varied and speech performance is examined.

In TABLE 1 and SNR values are tabulated for 2 sentences at 5 and 10 dB SNR for MMSE and Proposed methodology respectively, In Figure 1 and 2, Time domain plot and STFT plot are shown are shown for the modified spectral subtraction methodology. For comparative purposes, we also plot the performance of the MMSE method as implemented by Ephraim and Malah in [10] as shown in fig 3 and 4. The proposed multi-band spectral subtraction approach consistently outperformed the MMSE approach for both SNRs. The improvement in speech quality can be seen. Informal listening tests indicated that the multiband approach yielded very good speech quality with very little trace of musical noise and with minimal, if any, speech distortion. The lack of musical noise can also be seen in Figure 2, which shows the spectrograms of enhanced speech obtained with multi-band spectral subtraction and enhanced speech obtained with power spectrum subtraction.

Table 1: Comparsion of SNR

Input SNR	MMSE		PROPOSED Method	
	Male	Female	Male	Female
5db	10.6665	11.3821	11.1179	11.5154
10db	14.7504	14.3721	14.9710	14.7169

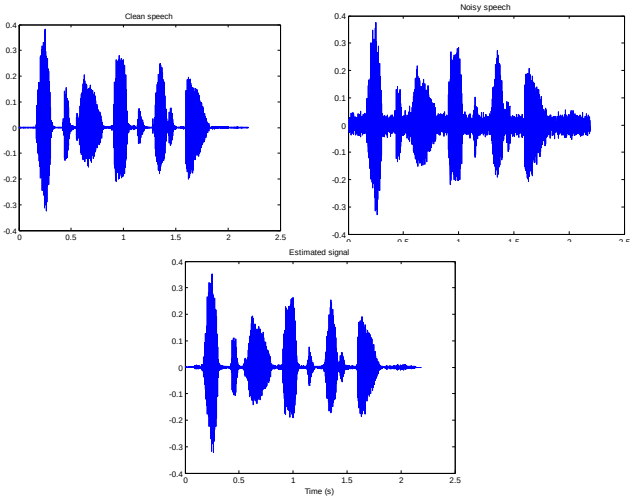

Fig 1a: Speech utterance with background car noise(right), clean speech (left).Effect of modified spectral subraction (center) for Male speaker.

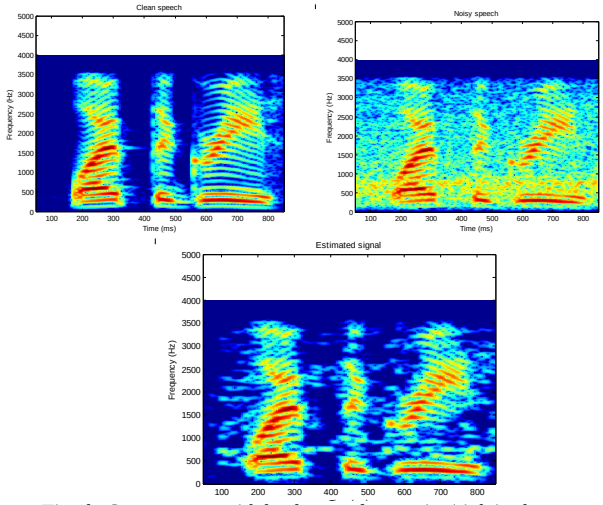

Fig 1b: Spectrogram with background car noise(right), clean speech (left).Effect of modified spectral subraction (center) for Male speaker.

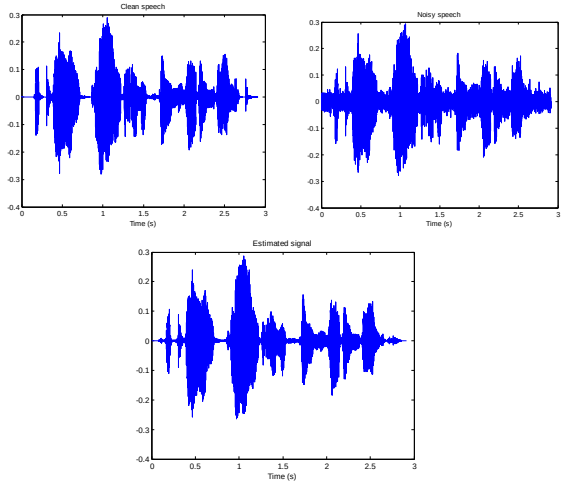

Fig 2a: Speech utterance with background car noise(right), clean speech (left).Effect of modified spectral subraction (center) for female speaker.

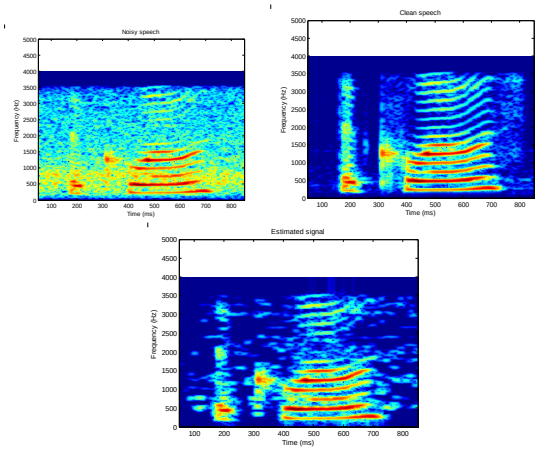

Fig 2b: Spectrogram with background car noise(right), clean speech (left).Effect of modified spectral subraction (center) for female speaker.

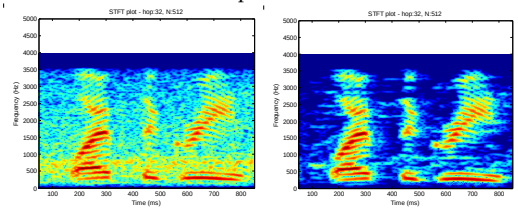

Fig 3a: Spectrogram with background car noise(right), clean Effect of MMSE for male speaker

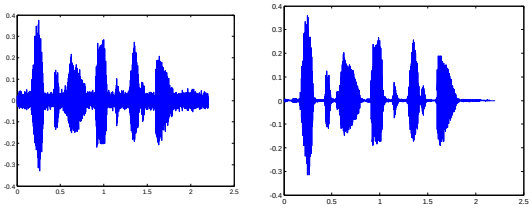

Fig 3b: Speech utterance with background car noise(right), clean Effect of MMSE for male speaker

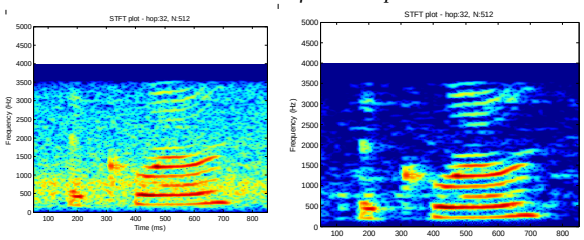

Fig 4a: Spectrogram with background car noise(right), clean Effect of MMSE for female speaker

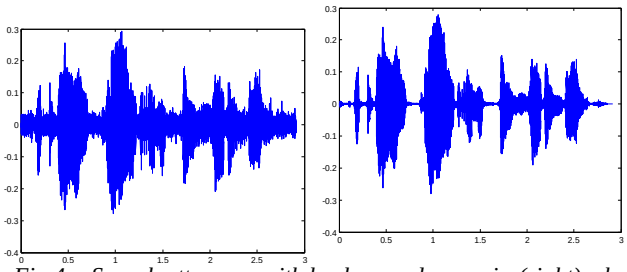

Fig 4a: Speech utterance with background car noise(right), clean Effect of MMSE for female speaker

V. CONCLUSIONS

The Proposed spectral subtraction method provides a definite improvement over the MMSE method. We consider that the improvement is due to the fact that the multi-band approach takes into account the non-uniform effect of colored noise on the spectrum of speech. The added computational complexity of the algorithm is minimal. We found that four linearly-spaced frequency bands were adequate in obtaining good speech quality.

ACKNOWLEDGMENT

The Authors are Grateful to R. C. Hendriks, R. Heusdens and J. Jensen for Providing Matlab Implementation of MMSE Based Noise PSD Tracking with Low Complexity

REFERENCES

[1] S.F.Boll," Suppression of acoustic noise in speech using spectral subtraction,"IEEE Trans. Acoust., Speech, Signal Process., vol 27,pp. 113-120,Apr.1979.

[2] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden markov models and the projection, for robust speech recognition in cars," *Speech Communication*, Vol. 11, Nos. 2-3, pp. 215-228, 1992.

[3] I. Soon, S. Koh and C. Yeo, "Selective magnitude subtraction for speech enhancement," *Proceedings. The Fourth International Conference/Exhibition on High Performance Computing in the Asia-Pacific Region*, vol.2, pp. 692-695, 2000.

[4] K. Wu and P. Chen, "Efficient speech enhancement using spectral subtraction for car hands-free application," *International Conference on Consumer Electronics*, vol. 2, pp. 220-221, 2001.

[5] C. He and G. Zweig, "Adaptive two-band spectral subtraction with multi-window spectral estimation," *ICASSP*, vol.2, pp. 793-796, 1999.

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

[7] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech and Audio Processing*, pp 126-137, vol. 7, March 1999.

[8] P. Krishnamoorthy; S. R. Mahadeva Prasanna "Enhancement of Noisy Speech by Spectral Subtraction and Residual Modification "IEEE Transactions on Signal Processing , pp.1-5,Sep 2006.

[9] Crozier, P.M.; Cheetham, B.M.G.; Holt, C.; Munday, E.;"Speech enhancement employing spectral subtraction and linear predictive analysis "IEEE Transactions on Electronics Letters,Vol 29 pp.1094-1095, June 2002.

[10] Y.Epharin, and D.Malah,"Speech enhancement a minimum mean square error log –spectral amplitude estimator," IEEE Trans .Acoustic., speech signal process vol ASSP-32 pp.1109-21,dec 1984.

[11 Y.Epharin, and D.Mallah, " Speech enhancement using a minimum-mean square error log-spectral amplitude estimator," IEEE Trans. Acoustic., speech, signal process., vol ASSP-333 pp.443-5, apr.1985.

[12] B.J.Shannon,"Speech recognition and enhancement using autocorrelation domain processing," Ph.D.dissertation, school of engineering, Griffith University, Brisbane, Australia, Aug.2006

[13] Malte sandrock and, Stefan Schmitt Realization of an adaptive algorithm with subband filtering Approach for acoustic Echo Cancellation in telecommunication Applications. Proc ICASSP 2004

[14] Chowdhury, F.A. Alam, J. Alam, F. O'Shaughnessy, D. "Perceptual multiband spectral subtraction for noise reduction in hearing aids" IEEE Transactions on Speech Processing, pp. 395-399, Dec 2008.

[15] IEEE Subcommittee (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Trans. Audio and Electroacoustics*, AU-17(3), 225-246.

[16] Hu, Y. and Loizou, P. (2007). "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Communication*, 49, 588-601.

[17] Hu, Y. and Loizou, P. (2008). "Evaluation of objective Quality measures for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, 229-238.

[18] Ma, J., Hu, Y. and Loizou, P. (2009). "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions", *Journal of the Acoustical Society of America*, 125(5), 3387-3405.

[19] ITU P.862 (2000). Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. *ITU-T Recommendation P. 862*