Speech Enhancement using Multiband Spectral Subtraction with Cross Spectral Component Reduction

Megha S. Karabashetti
Department of Electronics and Communication Engineering,
Basaveshwar Engineering College Bagalkot

Dr. Shridhar S. K.
Department of Electronics and Communication Engineering,
Basaveshwar Engineering College Bagalkot

***Abstract* — To preserve the message content, it is necessary to improve the quality and intelligibility of the speech signal. Quality of the speech signal can be improved by enhancing the noisy speech signal. This paper presents two algorithms to reduce the additive background noise considerably.First method is modified multiband spectral subtraction to reduce additive noise, which is non-stationary with respect to the speech signal. In this method spectral subtraction is performed based on the SNR values in different frames of the noisy speech.Secondmethod is implemented to reduce the cross spectral components, where the noise signal is correlated in some extentwith noisy signal.These methods are implemented to overcome the limitations of basic spectral subtraction method. Both the methods are combined to enhance the noisy speech signal.**

Keywords—spectral subtraction, cross spectral components, SNR value, multi-band spectral subtraction.

I. INTRODUCTION

Speech has been evolved as the primary way of communication. In any communication system speech signal is usually accompanied by the background noise, which not only affects the task of listening but also degrades the performance of digital signal processor. Hence it is necessary to reduce the background noise for the effective communication.

Speech enhancement is one among the processes to improve the quality, intelligibility, perceptibility of the speech signal, by the reduction of background noise from noisy speech signal. Some of the applications of speech enhancement are, in Mobile communication, Tele communication, Hearing aids, Recording systems, Tele-conferencing. This paper provides algorithms for the reduction of additive noise, such as Babble noise, White noise, Flicker noise, etc. Both the algorithms are improvised versions of Basic spectral subtraction method.

In paper [1] Boll presentsdifferent techniques to reduce additive noise one such method is basic spectral subtraction method, which involves the subtraction of power spectral magnitude of the estimated noise from that of the noisy speech signal to obtain the speech signal. There are two assumptions made during the process. First one is, speech signal assumed to be stationary on short time basis. Second assumption is that the noise signal is uncorrelated to the speech signal. Noise signal is estimated from the silent periods, where there is the absence of the speech signal. In practical these assumptions are not true in all cases; noise is not uniformly distributed throughout the noisy signal. Thus this method of speech enhancement introduces the musical noise.

To overpower the basic spectral subtraction method and to avoid the musical noise that can be caused due to this method, proposed work introduces two speech enhancement algorithms. The First method is modified multi-band spectral subtraction technique. This algorithm is implemented to process the noisy speech signal degraded by the additive noise, where speech signal is non stationary [13]. Spectral subtraction is performed on the basis of the SNR of the current frame. Complete details of the method are given in the section-1.

Second method involves the computation of the correlation between the speech signal and noise signal, to process the noisy speech signal corrupted by noise correlated to the speech signal. The computation details are explained in the section-2.

1. Modified multi-band spectral

Let *s(n)*be the noisy speech signal to be enhanced, corrupted by the additive noise *d(n).* And *c(n)* be the clean speech signal. Hence S(n) is given by,

$$s(n) = c(n) + d(n) \quad (1)$$

let *s(n)* be converted into transfer domain and is written as

$$S(f) = C(f) + D(f) \quad (2)$$

Power spectrum of the corrupted speech signal is given as

$$|S(f)|^2 = |C(f)|^2 + |D(f)|^2 + C(f).D(f)^* + C(f)^*.D(f) \quad (3)$$

According to the assumption made in basic spectral subtraction method noise signal is uncorrelated to the corrupted signal, $C(f).D(f)^* + C(f)^*.D(f)$ terms in the equation (3) are neglected. Thus clean speech $C(f)$ can be obtained by the equation as follows,

$$|C(f)|^2 = |S(f)|^2 - |D(f)|^2 \quad (4)$$

But in this method it is assumed that noise is uniformly distributed throughout the corrupted speech signal, which is not possible in practical aspects. Hence if we follow the same method it subtracts the same amount of, estimated noise from the noisy speech signal. To avoid this another method of speech enhancement is required, where noise to be subtracted depends on the SNR in the corresponding portion of the signal *S(n)*. Modified multi-band spectral subtraction is performed to compute the over subtraction factor α which depends on the SNR value. Thus clean speech can be computed by introducing over subtraction into equation (4) factor can be given by,

$$|C(f)|^2 = |S(f)|^2 - \alpha |D(f)|^2 \quad (5)$$

In paper [13] the author has given the relationship between α and SNR. The relation is given as,

$$\alpha = \begin{cases} 5 & SNR < -5 \\ 4 - 3/20\,(SNR) & -5 < SNR < 20 \\ 1 & SNR > 20 \end{cases} \quad (6)$$

2. Cross-correlation technique

In equation (3) $C(f).D(f)^* + C(f)^*.D(f)$ are considered as the cross correlation terms, which are neglected in spectral subtraction technique. But in real time applications there is certain amount of correlation between speech signal and noise. Hence it is necessary to find these correlation terms, r_{cd} and r_{dc}

Respectively, but we don't have access to the clean speech hence we can find the correlation between corrupted speech signal and noise signal. ie., r_{yd}. where r_{sd} is gives

$$R_{sd} = r_{cd} + r_{dd}$$

r_{sd} gives required correlation between clean speech signal and noise signal. Paper [6] gives equation for correlation parameter δ introduced into equation (5) as follows,

$$|C(f)|^2 = \begin{cases} |S(f)|^2 - \alpha\, |D(f)|^2 - \delta |S(f)| * |D(f)| & \text{if } |S(f)|^2 > \alpha |D(f)|^2 \\ \beta\, |D(f)|^2 & \text{else} \end{cases} \quad (7)$$

Where α is the over spectral subtraction factor estimated by the equation (6), β is the spectral floor factor whose value is 0.002 given in the paper [3]. δ is correlation factor which gives the estimate of the correlation between the noisy speech signal and estimated noise signal. Equation to compute δ is

$$\delta = \left| \frac{\chi_{sd} - \mu_s \mu_d}{\sigma_s \sigma_d} \right| \quad (8)$$

where,

$$\left.\begin{aligned} \chi_{sd} &= \frac{1}{N/2} \sum_k |S(k)| * |D(k)| \\ \mu_s &= \frac{1}{N/2} \sum_k |Y(k)| \\ \mu_d &= \frac{1}{N/2} \sum_k |D(k)| \end{aligned}\right\} \quad (9)$$

Where μ_s, μ_d are the mean of noisy speech signal and noise signal respectively where $0 < k < N/2$, N being the size of FFT. And σ_s^2, σ_d^2 are the variances of the corrupted speech signal and estimated noise signal.

II. PROPOSED WORK

Initially noisy speech signal is divided into frames of20ms (160 samples per frame). Hamming Window is used for this purpose (with 160 window size). Windowing method may introduce spectral leakages at the edges of the window, which will cause loss of information, hence to avoid the same50% overlapping is done before processing of the signal. Windowed noisy speech signal can be written as

$$S_w(n) = s(n) * w(n)$$

From equation (1)

$$S_w(n) = [c(n) + d(n)] * w(n)$$
$$= c_w(n) + d_w(n)$$

FFT of the noisy speech signal is computed followed by the computation of the power spectrum magnitude as in the equations (2) and (3). In modified multiband spectral subtraction each frame magnitude spectrum of the noisy speech signal is divided into bands with 40 samples each. Spectral subtraction is performed separately for these bands based on there SNR values using the equations (5) and (6) by computing the value of over subtraction factor.

Now finally by using equations (8) and (9) δ, correlation factor is calculated, and by using equation (7). Magnitude spectrum of the clean speech is obtained.

Magnitude spectrum estimated clean speech signal and unchanged phase spectrum of the original speech signal are combined to form complex spectrum. Inverse Frequency Fourier transform is performed to convert complex spectrum into time domain signal. As 50% overlapping is used in the framing process, 50% overlap adding is done to get the enhanced speech signal.

The Fig-1 block diagram shows different steps involved in the implementation of proposed method.

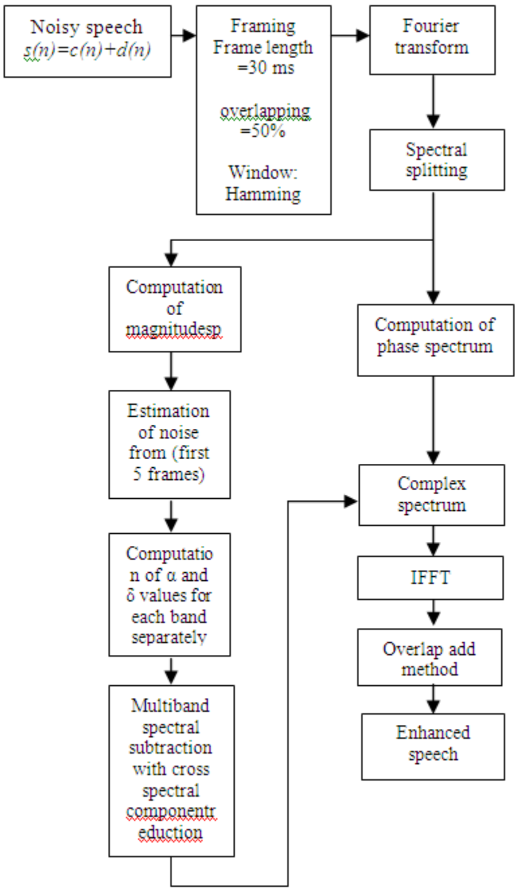

Fig.1 Block diagram for the proposed method.

III. RESULTS

In this paper subjective listening test and spectrogram analysis is used for the assessment of speech quality. By using these analysis methods performance of the proposed method is compared with the existing speech enhancement techniques.

In subjective listening test processed speech is compared with the unprocessed speech signal, with the help of listeners. listeners are allowed to rate the speech quality based on a predefined scale.

Spectrogram is the time-frequency representation of any speech signal, where frequency of the signal vary as the time varies. color of the spectrogram represents the energy of the speech at that frequency. Dark color depicts that the speech signal is of high energy.

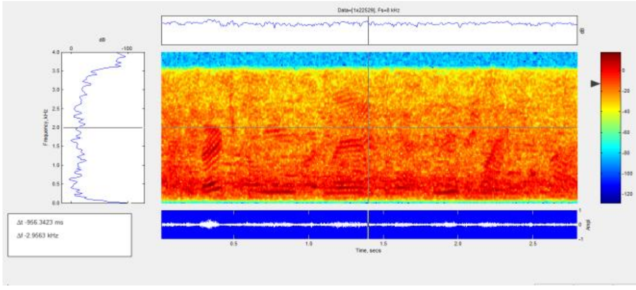

Fig.2signal 1 -0dB SNR noisy speech with Babble noise

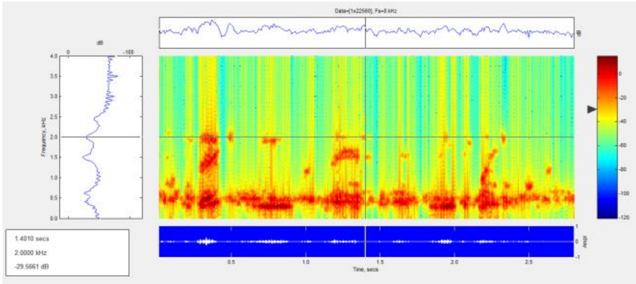

Fig.3signal enhanced by multiband spectral subtraction

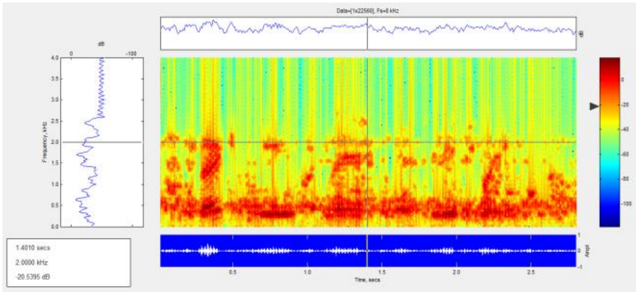

Fig.4signal enhanced by the proposed method

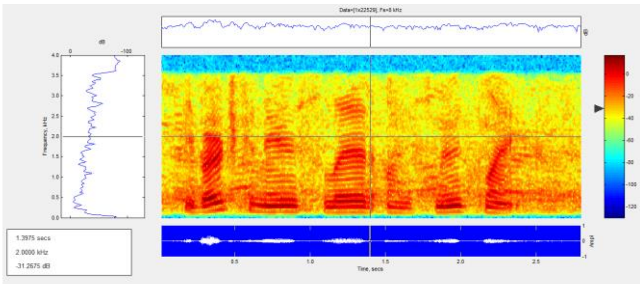

Fig.5signal 2 - 15 dB SNR noisy speech signal with Babble noise

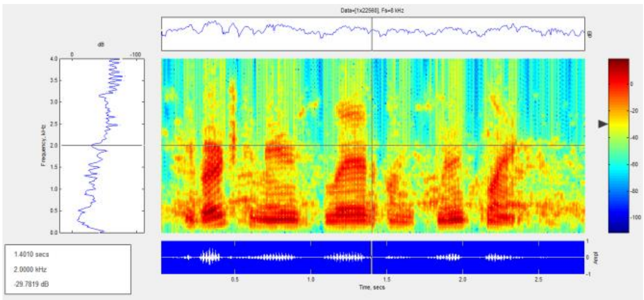

Fig.6signal 2 enhanced by the proposed method

Fig.2 and Fig.5 shows spectrogram analysis of 0dB and 15dB noisy speech signal corrupted by Babble noise respectively.

And the Fig.2 and Fig.6 shows the spectrogram analysis of the enhanced speech signals. It is observed that the speech quality has been increased by using proposed method. Mean opinion of subjective listening test of modified multiband spectral subtraction for signal with 0dB and 15dB SNR is 2.7 and 2.6 (moderate) respectively. Mean opinion of the proposed method for signal with 0dB and 15dB SNR is 3.7 and 3.6 respectively (greater than that of previous method).

CONCLUSION

Problems and limitations of the basis spectral subtraction method is considered in this paper. In this paper we have performed multiband spectral subtraction by computing the value of over subtraction factor. Further cross spectral components were computed by cross-correlation technique. By the result analysis, it is concluded that the quality of the speech signal has increased by the proposed method than that in the spectral subtraction method.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Shridhar K. sir for there helpful and fruitful guidance during the course of the work. The authors also thank authorities of Basaveshwar Engineering college for providing all required facilities during the course of this work.

REFERENCES

[1] S.F. Boll, "Suppression of Acoustic Noise In Speech Using Spectral Subtraction", IEEE Trans. Acoust., Speech, Signal Processing, vol. 27, no. 2, pp. 113–120, 1979 .

[2] M Berouti, et.al, "Enhancement of Speech Corrupted by Acoustic Noise", Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp. 208-211, April 1997.

[3] T.F.Quatieri and R. Baxter, "Noise Reduction Based o n Spectral Change", Proc. IEEE Workshop on Applications of Signal Processing To Audio and Acoustics, pp. 8.2.1-8.2.4, New Paltz NY ,Oct. 1997.

[4] Sofia Ben Jebara, et.al, "Reduction of Musical Noise Generated by Spectral Subtraction by Combining Wavelet Packet Transform and Wiener Filtering", IEEE 10th European Signal Processing Conference, pp. 1-4, 2000.

[5] Yi Hu , Mukul Bhatnagar and Philip C. Loizou, "A Cross-Correlation Technique for Enhancing Speech Corrupted With Correlated Noise ", in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001 (ICASSP"01) Salt Lake City, UT, vol. 1, pp. 673-676, May 2001.

[6] Masatsugu Okazaki, Toshifumi Kunimoto, Takao Kobayashi, "Multi-Stage Spectral Subtraction for Enhancement of Audio Signals", in Proc. IEEE ICASSP 2004, pp. 805-808, 2004.

[7] Saeed Ayat, Mohamad T. Manzuri, Roohollah Dianat. "An improved spectral subtraction speech enhancement system by using an adaptive spectral estimator", IEEE CCECE/CCGEI, pp. 261-264, (2005).

[8] Chiung-Wen Li, et.al, "Signal Subspace Approach for Speech Enhancement in Nonstationary Noises", IEEE International Symposium on Communications and Information Technologies, pp. 1580 – 1585, 2007.

[9] Kamil Wojcicki, et.al, "Exploiting Conjugate Symmetry of the Short-Time Fourier Spectrum for Speech Enhancement", IEEE Signal Processing Letters, vol. 15, 2008.

[10] Radu Mihnea Udrea, et.al, "An Improved Multi-band Speech Enhancement Method for Coloured Noise Estimation and Reduction",International Journal on Advances in Telecommunications, vol. 3 no 3 & 4, year 2010.

[11] Chao Li, and Wen-Ju Liu " A Novel Multi-Band Spectral Subtraction Method Based On Phase Modification and Magnitude Compensation", in Proc. 2011 International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Prague, pp. 4760 – 4763, May 2011.

[12] Nitya Tiwari, et.al, "Speech Enhancement and Multi-band Frequency Compression for Suppression of Noise and Intra speech Spectral Masking in Hearing Aids" Annual IEEE India Conference (INDICON), pp. 1-6, 2013.

[13] Sunil D. Kamath and Philipos C. Loizou et.al," A Multi-band Spectral Subtraction Method for Enhancing speechcorrupted by Coloured noise", Department of Electrical EngineeringUniversity of Texas at Dallas.