Spectral Subtraction for Enhancement of Speech Quality

Gitu Geevarghese Milind S. Shah

Abstract - **This paper describes a method for enhancing speech corrupted by additive white Gaussian noise. The method is based on the spectral noise subtraction method. The basic method subtracts an estimate of the noise power spectrum from the speech power spectrum, setting negative differences to zero, recombining the new power spectrum with the original phase, and then reconstructing the time waveform with the help of IFFT. While this method reduces the broadband noise, it also introduces an annoying "musical noise". A modified spectral subtraction method is used which consists of subtracting an overestimate of the noise power spectrum, and preventing the resultant spectral components from going below a preset minimum level (spectral floor). Listening tests were performed to determine the quality and intelligibility of speech enhanced by this method.**

IndexTerms— Speech enhancement, spectral subtraction, over-subtraction, spectral floor parameter.

I. INTRODUCTION

This paper is a report on the work done to improve the quality of speech degraded by background noise. The goal of this approach is to improve the listenability of the speech signal by enhancing the quality, without affecting the intelligibility of the speech. The noise is present in the speech signal in such a level that the speech is essentially unintelligible out of context. The average SNR (signal-to-noise ratio) is used to measure the noise levels present in the degraded speech.There is strong correlation between the intelligibility of a sentence and the SNR, but intelligibility also depends on the speaker, on context, and on the phonetic content [1].

Spectral subtraction is historically one of the first speech enhancement algorithm proposed for the removal of additive background noise. The main aim of speech enhancement algorithm is to improve the quality and/or intelligibility of the noisy speech signals by using various techniques and algorithms [2]. Spectral subtraction is a single-input speech enhancement technique developed for use in audio codecs and speech recognition

It involves estimating the noisespectrum, subtracting it from the noisy speech spectrum, and re-synthesizing the speech signal. As the interfering noise may be non-stationary, its spectrum needs to be dynamically estimated. Under-estimation of the noise results in residual noise and its over-estimation results in distortion leading to degraded quality and reduced intelligibility. Noise can be estimated during the silence intervals identified by voice activity detection [2]. But the detection may not be satisfactory under low-SNR conditions and the method may not correctly track the noise spectrum during long speech segments. Several statistical techniques for estimating the noise spectrum, without voice activity detection, have been reported [10]-[20].

The objective of this paper is to implement basic spectral subtraction algorithm and a modified spectral subtraction with over subtraction parameter and spectral floor parameter.The paper is divided into four sections, sectionsII and III explaining the algorithm that were used to implement the spectral subtraction method, section IV describes the implementation and section V results obtained from the algorithm and section VI gives the conclusion.

II. BASIC SPECTRAL SUBTRACTION METHOD

Spectral subtraction is a single-input noise reduction method based on the short time estimation of the magnitude spectrum of the noise. The processing of this method involves estimating the magnitude spectrum of the noise, estimating the magnitude spectrum of the speech signal, and re-synthesizing the speech using the enhanced magnitude spectrum along with the phase spectrum of the noisy speech signal[4][5].

A block diagram of speech enhancement using spectral subtraction is shown in Fig.1.

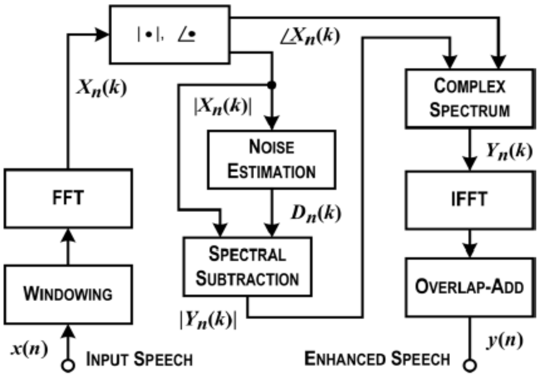

Fig.1 Speech enhancement by Basic Spectral Subtraction method[14].

Windowed frames of the noisy speech signal $x(n)$, are given to the FFT block to find the magnitude and phase spectra. The magnitude spectra of the initial few frames are used to estimate the noise magnitude spectrum $D_n(k)$. The noise is estimated during non-speech segments using averaging of the first few frames. The enhanced magnitude spectrum $|Y_n(k)|$ is computed using spectral subtraction. IFFT is taken for the complex spectrum formed by enhanced magnitude spectrum and noisy speech phase spectrum. Using overlap-add method the enhanced signal is reconstructed. [6]

The effectiveness of the noise removal process is dependent on obtaining an accurate spectral estimate of the noise from the noisy speech signal. The estimated noise and actual noise which is present in the signal has a significant difference in the short-time speech spectrum which may result in the presence of isolated residual spectral peaks of large variance. These residual spectral contents manifest themselves in the reconstructed signal as varying tonal sounds known as “musical noise” and may result in an unnatural quality[7].The magnitude spectrum or the power spectrum after spectral subtraction may contain some negative values due to errors in the estimated noise spectrum. These values are rectified using half-wave rectification (set to zero) or full-wave rectification (set to its absolute value).Thus the enhancement algorithm consists of thefollowing relationship:

Let $Y_n(k) = |X_n(k)| - D_n(k)$ (1)

Where

$$Y_n(k) = \begin{cases} Y_n(k) & \text{if} \quad Y_n(k) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where$Y_n(k)$ is the modified signal spectrum, $X_n(k)$ is the spectrum of the input noise corrupted speech, and $D_n(k)$ is the estimate of the noise spectrum. The enhanced speechsignal is obtained from both $Y_n(k)$ and the original phase of the noisy speech signal by an inverse Fourier transform:

$$y_n(k) = F^{-1}\left\{\sqrt{Y_n(k)}e^{j\theta(w)}\right\} (3)$$

Where $\theta(w)$ is the phase function of the DFT of the input speech.Since it is assumed that speechsignal and noise are uncorrelated, some of the components of the processed spectrum, $Y_n(k)$,may be negative. Thesenegative values are set to zero as shown in(2).This can lead to further distortions in the resulting time signal. To overcome the shortcomings of spectral subtraction, Berouti *et al.* [17] developed a modified spectral subtraction.

III.MODIFIED SPECTRAL SUBTRACTION

The modified spectral subtraction algorithm consists of reducing the narrow spectral peaks by over estimating the noise spectrum. The enhanced magnitude spectrum $|Y_n(k)|$ computed using modified spectral subtraction is given as following

Let $Y_n(k) = |X_n(k)|^{\gamma} - \alpha D_n(k)^{\gamma}$ (4)

$$Y_n(k) = \begin{cases} Y_n(k)^{\frac{1}{\gamma}} & \text{if} \quad Y_n(k)^{\frac{1}{\gamma}} \geq \beta D_n(k) \\ \beta D_n(k) & \text{otherwise} \end{cases} \quad (5)$$

Here γ is an exponent factor, resulting in power subtraction if$\gamma = 2$ and magnitudesubtraction if$\gamma = 1$. Use of subtraction factor $\alpha > 1$ reduces the broadband peaks in theresidual noise, but it may result in deep valleys, causing warbling or musical noise andadversely affecting the speech quality. The musical noise is masked by a floor noisecontrolled by the spectral floor factor β. These two factors offer a great flexibility in thealgorithm.Assuming that the phase error does not significantly affect the intelligibility andquality of speech, the enhanced magnitude spectrum is combined with the original noisyphase, to get the complex spectrum $y_n(k)$ [8][9].

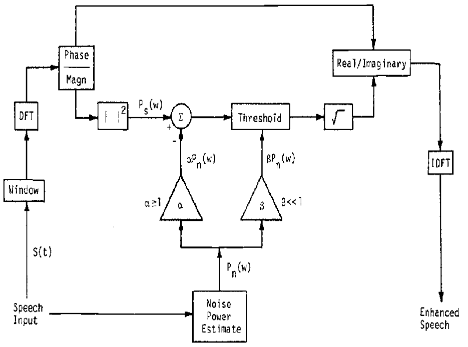

Fig.2. Modified spectral subtraction with spectral floor parameter [17]

The modified spectral subtraction technique is as shown in the fig.2 of section 3. The resulting complex spectra after the magnitude or power spectrum subtraction are used to re-synthesize the speech signal. As spectralsubtraction involves modification of the STFT, there may bediscontinuities between the signal segments corresponding to the modified complex spectra ofthe consecutive frames. The value of α is changed according to the frame wise SNR across the same sentence as,

$$\alpha = \begin{cases} 5 & SNR \leq -5dB \\ 4 - (3/20) * SNR & -5 \leq SNR \leq 20 \\ 1 & SNR \geq 20 \end{cases} \quad (6)$$

Overlap-add method is used after doing the IFFT of the frames in the synthesis part in order to reduce the distortion related to their discontinuities, which helps in masking.In the generalized spectral subtraction, it is assumed that the noise affects theentire spectrum uniformly, which is generally not valid in the case of real time noises. The generalized spectral subtraction is based on theassumption that speech and additive noises are uncorrelated and hence cross terms are madezero. However, this assumption is not valid when the speech is processed on frame-to-framebasis as mentioned in[17].

IV. IMPLEMENTATION

In order to compare the results of both basic spectral subtraction and modified spectral subtraction the implementation was carried out in MATLAB software. The speech was divided into frames using a hamming window. Each frame had 400 samples of 25ms window. The 512-point FFT length was used.Testing involved processing of speech with additive white, pink, babble, car, and train noises at SNR 0 dB. The results obtained from MATLAB are of a sentence “This is a speech project”, which is a recording of a female speaker which is being added with additive white Gaussian noise of SNR 0dB.

V. Results

Informal listening test showed that the processing significantly enhanced the speech for all noises and there was no audible roughness. While spectral floor factor β in between 0.02 and 0.06 for SNR below -5dB and β value between 0.005 and 0.02 for SNR above 0 was found to be appropriate in all cases, most appropriate value of subtraction factor α varied over 1.5-3 which is varied over different frames according to the SNR value of each frame.The graph shows the clean speech, noisy speech and clean speech and their respective spectrograms for half wave rectification and full wave rectification.

A. Noise estimation

The magnitude of noise spectrum is taken as a average value during the non-speech activity $Dn(k)$ and the phase of $Dn(k)$ is replaced by the phase $X_n e^{j\theta(w)}$ of $|Xn(k)|$. Assuming that the phase error does not significantly affect the intelligibility and quality of speech, the enhanced magnitude spectrum is combined with the original noisy phase, to get the complex spectrum $y_n(k)$.

B. Synthesis of the enhanced signal

To examine the effect of the processing parameters, thetechnique was implemented using MATLAB for offlineprocessing. Implementation of 50% overlap is selected for implementation. It was observedthat for different noises and SNR of 0dB, appropriate selectionof subtraction factor α and floor factor β resulted inalmost similar results for magnitude subtraction (exponentfactor γ = 1) and power subtraction (γ = 2). The results ofmagnitude subtraction showed higher tolerances tovariation in the values of α and β, and hence only magnitudesubtraction was used for processing.

The graphs below show the results of basic spectral subtraction using half-wave rectification and modified spectral subtraction using full wave rectification with spectral floor and over subtraction parameter. The results show that the recovered speech for half wave rectification have better recovery of speech.

VI. Conclusion

The main differences between the basic spectral subtraction method and modified spectral subtractionimplementation is that here there is subtraction of an overestimate of the noise spectrum which prevent the resultant spectral components from going below a spectral floor. This implementation of the spectral noise subtraction method affords a great reduction in the background noise with very little effect on the intelligibility of the speech.

References

[1] S. V. Vaseghi, Advanced digital speech processing and noise reduction,2nd ed.

[2] Z. Goh, K. Tan and B. T. Tan, “Post Processing Method For Suppressing Musical Noise Generated By Spectral Subtraction,” *IEEE Trans. Speech Audio Process*. Vol. 6 no. 3, pp. 287-292, 1998.

[3] A. Drygajlo and M. Maliki, “Speaker Verification In Noisy Environments With Combined Spectral Subtraction And Missing Feature Theory,” *IEEE Trans. Speech Audio Process*, pp. 121-124, 1998.

[4] M. Padilla and T. Quatieri, “A Comparison Of Soft And Hard Spectral Subtraction For Speaker Verification,” in *EUROSPEECH’99*, Hungary,2000.

[5] Y. Lu and C. Loizou, “A Geometric Approach to Spectral Subtraction,” *Speech Comm.* vol. 50, pp.453-466, 2008

[6] R. MARTIN “Spectral Subtraction Based On Minimum Statistics,” *in Proc. EUSIPCO ’94*, 1994, pp. 1182-85.

[7] P. Sovka, P.Pollak and J.Kybic, “Extended Spectral Subtraction,” *EUSIPCO ’96*, 1996.

[8] V. stahl, A.Fischer and R.Bippus, “Quantile Based Noise Estimation for Spectral Subtraction and Weiner Filtering,” in Proc. *ICASSP’00*, 2000, 1875-1878.

[9] S. F. Boll, “Suppression Of Acoustic Noise In Speech Using Spectral Subtraction,” *IEEE Trans. On acoustics, speech, and signal processing*, vol.ASSP-27, No.2, April 1979.

[10] R. Martin, “spectral subtraction based on minimum statistics,” in *proc EUSIPCO*, pp.1182-1185.

[11] A. Vizinho, P. Green, M. Cooke and L. Josifovski, “Missing Data Theory, Spectral Subtraction And Signal-to-noise Ratio Estimation For Robust ASR: An Integrated Study,” *in EUROSPEECH ’99*, budapest, Hungary, 1999.

[12] R. Martin, “Spectral subtraction based on minimum statistics,” in *Proc. Eur. Signal Process. Conf.*, 1994, pp. 1182-1185.

[13] V. Stahl, A. Fisher, and R. Bipus, “Quantile based noise estimation for spectral subtraction and Wiener filtering,” in *Proc. IEEE ICASSP*, 2000, Istanbul, Turkey,
pp. 1875–1878.

[14] S. K. Waddi, P. C. Pandey, and N. Tiwari, “Speech enhancement using spectral subtraction and cascaded-median based noise estimation for hearing impaired listeners,” in *Proc. Nat. Conf. Commun. (NCC 2013)*, Delhi, India, 2013, pp. 1–5.

[15] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in *Proc. IEEE ICASSP* 1979, Washington, DC, pp. 208–211.

[16] P. C. Loizou, *Speech Enhancement: Theory and Practice*. New York: CRC, 2007.

[17] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in *Proc. IEEE ICASSP* 1979, Washington, DC, pp. 208–211.

[18] S. Kamath and P. Loizou, “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise,” in *Proc. IEEE ICASSP*, 2002, Orlando, Florida, vol. 4, pp. IV–4164.

[19] Y. Lu and P. C. Loizou, “ A geometric approach to spectral subtraction,” *Speech Commun.*, vol. 50, no. 6, pp. 453–466, 2008.

[20] K. Paliwal, K. Wojcicki, and B. Schwerin, “Single-channel speech enhancement using spectral subtraction in the short-time modulation domain,” *Speech Commun.*, vol. 52, no. 5, pp. 450–475, 2010.

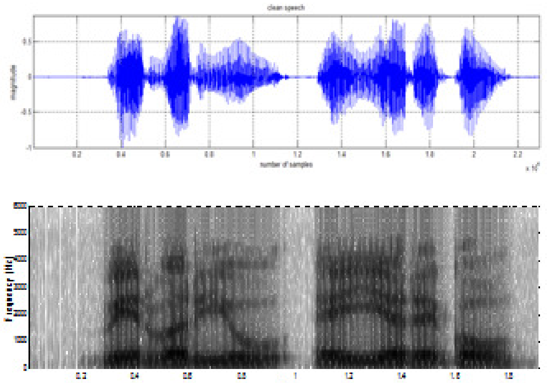

Fig.3 Waveform of clean speech

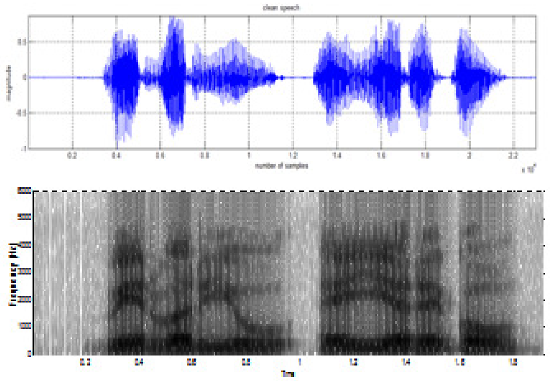

Fig 6 Waveform of clean speech

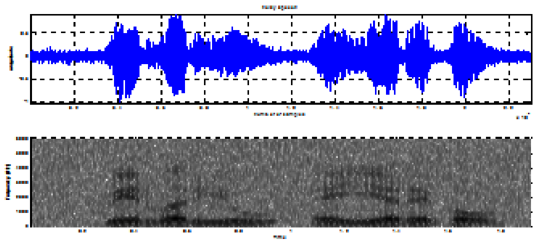

Fig 4 Waveform of clean speech

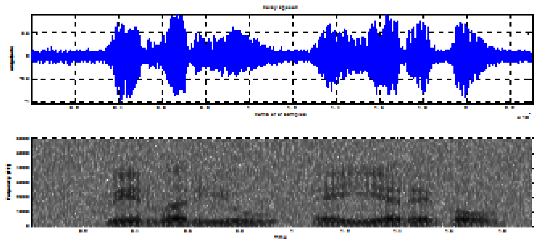

Fig 7 Waveform of noisy speech

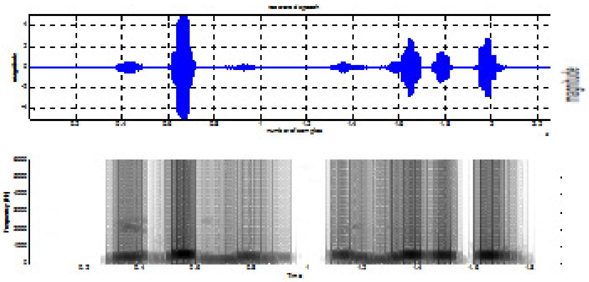

Fig 5 Waveform of recovered speech using half wave rectification

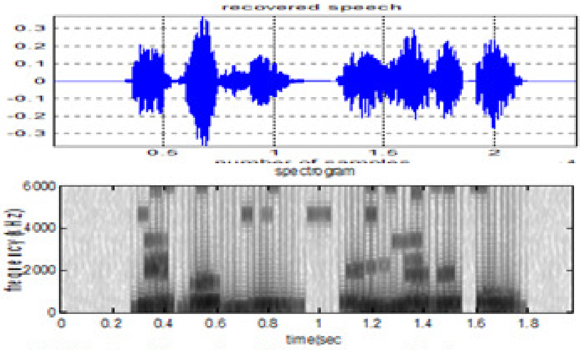

Fig 8 Waveform of recovered speech using full wave rectification