IJCAT International Journal of Computing and Technology, Volume 1, Issue 4, May 2014
ISSN : 2348 - 6090
www.IJCAT.org

Speech Enhancement using Modified Spectral Subtraction Algorithm

[1] Vijay Kumar M A, [2] Kishore T K , [3] Mahesha M R, [4] Karthik S, [5] Gagan H R

[1] Asst. Prof, Department of ECE, VTU, PESIT&M
Shivamogga, 577204, India

[2,3,4,5] Student Department of ECE, VTU, PESIT&M
Shivamogga, 577204, India

Abstract - The objective of Speech Enhancement is to improve the quality of audio signal that are degraded speech signal used in various application i.e. VOIP, hearing aids, teleconferencing systems. Currently the Speech Enhancement deals with several problems to improve requirements of speech recognition such as background noise that is presented in speech signal. The proposed work introduces alternative ways for providing best quality of speech signal aimed at speech enhances using Modified Spectral Subtraction algorithms (MSS). The scope of the proposed solution is to improve the SNR value of audio signal as well as to reduce completely background noise without affecting signal power. This paper defines the modified spectral subtraction method to estimate noise in non-stationary speech environment. Finally, we have discussed some of the performance parameters such as: time required as number of samples increases and Improved SNR.

***Keywords* - Speech Enhancement, Spectral Subtraction, SNR.**

1. Introduction

Speech signals are often contaminated by noise during acquisition and transmission. Removing/reducing noise in speech signals (denoising) is highly desirable, as it can enhance perceived speech quality, increase the intelligibility, effectiveness and increase the speech quality by using spectral subtraction algorithm for speech enhancement technique.

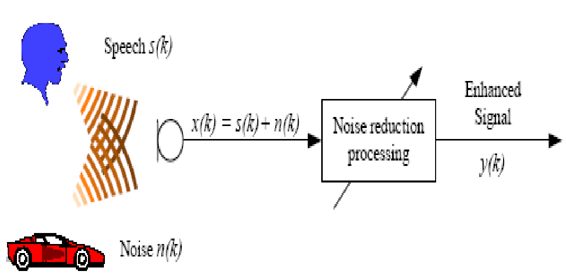

Fig. 1 Speech processing system

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signal. Aspects of speech processing include the acquisition, manipulation, storage, transfer and output of speech signals. The main applications of speech processing are the recognition, synthesis and compression and manipulation of human speech. Fig 1.1 shows how the noise is added in speech signal and how it is processed to get an enhanced speech signal.
The spectral subtraction [1] is based on the principle that the enhanced speech can be obtained by subtracting the estimated spectral components of the noise from the spectrum of the input noisy signal. Assuming that noise is additive to the speech signal.

2. Theoretical Background

The noise-suppressed spectral estimator is obtained by subtracting an estimate of the noise spectrum from the noisy speech spectrum. Noise spectrum is obtained from the noisy signal during non-speech activity. After developing the spectral estimator, the spectral error is computed and four methods for reducing it are presented.

The following assumptions were used in developing the analysis. The background noise is acoustically or digitally added to the speech. The background noise environment remains locally stationary to the degree that its spectral magnitude expected value just prior to speech activity equals its expected value during speech activity. If the environment changes to a new stationary state there exist enough time (about 300 ms) to estimate a new background noise spectral magnitude expected value before speech activity commences. For the slowly varying non-stationary noise environment, the algorithm requires a speech activity detector to signal the program that speech has ceased and a new noise bias can be estimated. Finally, it is assumed that significant noise reduction is possible by removing the effect of noise from the magnitude spectrum.

Speech is analyzed by windowing data from half-overlapped input data buffers. The magnitude spectra of the windowed data are calculated and the spectral noise bias calculated during non-speech activity is subtracted off. Resulting negative amplitudes are then zeroed out. Secondary residual noise suppression is then applied. A time waveform is recalculated from the modified magnitude. The waveform is then overlap added to the previous data to generate the output speech. Suppression of acoustic background noise in single-channel noisy speech has been typically carried out by spectral subtraction (SS). The basic SS technique involves subtracting an estimate of the noise spectrum from the noisy speech spectrum. Crucial to the performance of the technique is the voice activity detector (VAD) required to determine regions of speech pause for the noise estimate update.

In stationary background noise, the SS method is simple and quite effective. Specific shortcomings include the presence of residual noise and an artefact known as musical noise. Satisfactory solutions to the problem of musical noise have been obtained by advanced approaches such as parametric spectral subtraction (PSS) [2] which adapt the subtraction parameters based on a computed apriori SNR. A different approach to noisy speech enhancement is filtering by relative spectral processing based on the assumption that the corrupting noise varies only slowly with respect to speech. The temporal trajectory of each short-time spectrum component is filtered to separate speech from noise. This approach has the advantage of not requiring a VAD.

3. Proposed Approach

Spectral Subtraction is a method to enhance the quality of speech that has been degraded by additive noise. For this method it is assumed that the speech and noise signals are uncorrelated and for the noise signal to be stationary. Therefore, in Spectral Subtraction the noise in the degraded speech is estimated from the pauses in the speech signal, since speech in general is made up of many pauses such as between words or the next person talking.

Spectral subtraction is based on the principle that one can obtain an estimate of the clean signal spectrum by subtracting an estimate of the noise spectrum from the noisy speech spectrum. Spectral subtraction performs many operations to remove noise such as windowing, fft, ifft, overlap add method, noise detection, noise estimation and calculating SNR value. Fig 2 shows the block diagram of modified spectral subtraction.

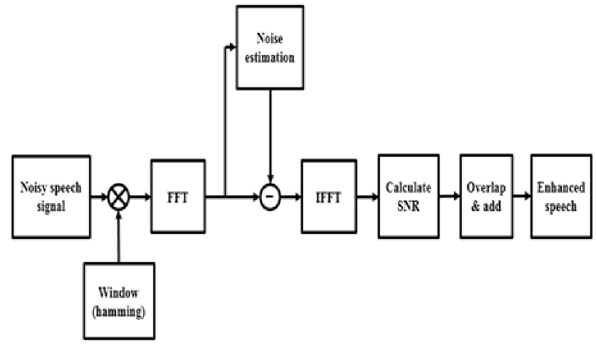

Fig. 2 Block Diagram of modified spectral subtraction

The spectral subtraction is based on the principle that the enhanced speech can be obtained by subtracting the estimated spectral components of the noise from the spectrum of the input noisy signal. Assuming that noise *d(n)* is additive to the speech signal *x(n)*, the noisy speech *y(n)* can be written as,

$$y(n) = x(n) + d(n), \quad \text{for } 0 \le n \le N-1 \qquad (1)$$

Where x(n) represents the pure speech signal, which is assumed to be a stationary signal whenever processing is done on a short time basis, d(n) is the uncorrelated additive noise and y(n) represents the degraded speech signal. The noise spectrum can be estimated, and updated, during the periods when the signal is absent or when only noise is present. Assumption is noise is additive¸ its spectrum does not change with time means noise is stationary or it's slowly time varying signal¸ whose spectrum does not change significantly between the updating periods. The noise corrupted input speech signal which is composed of the clean speech signal x(n) and the additive noise signal d(n) is shown in eq.(1) above. The above eq. can be given in Fourier domain as shown:

$$Y[w] = X[w] + D[w] \qquad (2)$$

Y[w] can be expressed in terms of Magnitude and phase as

$$Y[w] = Y\,|(w)|\ e^{j \text{Ø} y} \qquad (3)$$

Here |Y(w)| is the magnitude spectrum and Ø is the phase spectra of the corrupted noisy speech signal. Noise spectrum in terms of magnitude and phase spectra as

$$D[w] = |\,D[w]\,|\ e^{j \text{Ø} y} \qquad (4)$$

Estimate the clean speech signal simply by subtracting noise spectrum from noisy speech spectrum¸ in equation form

$$X^{\wedge}(w) = [|Y(w)|-|D(w)|]\, e^{j\varnothing y} \quad (5)$$

The magnitude of noise spectrum |D(w)| is unknown but can be replaced by its average value computed during non-speech activity i.e. during speech pauses. The spectral subtraction algorithm is computationally simple as it only involves a forward and inverse Fourier Transform. The basic block diagram of spectral subtraction method is shown above in fig 2

Noisy speech signal is segmented and then windowed using Hamming Window. Then Discrete Fourier transform (DFT) of segmented and windowed Noisy speech signal is taken [4]. DFT of noisy signal is then given to noise estimation block and speech enhancement block. Noise estimation block estimate the noise during the pauses and find the noise spectrum. In most speech-enhancement algorithms, it is make assumed that an estimate of the noise spectrum is available. The noise estimate can have a major impact on the quality and intelligibility of the enhanced signal.

If the noise estimate is too low, unwanted residual noise will be audible, if the noise estimate is too high, speech will be distorted. From the above discussion it is clear that subtraction process needs to be done carefully to avoid any speech distortion. If too much is subtracted, then some speech information might be removed as well, while if too little is subtracted then much of the interfering noise remains. It is clear from equation (5) that spectral subtraction method can lead to negative values, resulting from differences among the estimated noise and actual noise frame. Simple solution is set the negative values to zero, to ensure a non-negative magnitude spectrum.

$$|Xe(\omega)| = |Y(\omega)| - |De(\omega)| \quad (6)$$

The adaption of subtraction parameters has been discussed earlier. Here the parameters are chosen in such a way that the residual noise stays below the masking threshold of the auditory system. This would ensure that the residual noise is masked, and remains inaudible. However, if the noise level increases, the masking threshold is too low to completely mask the residual noise without increasing the speech distortion [5], leading to a synthetic sound.

4. Simulation Procedure

The subtraction parameters are kept to their minimal values. However, if the masking threshold is low, residual noise will be annoying to the human listener and it is necessary to reduce it. This is done by increasing the subtraction parameters.

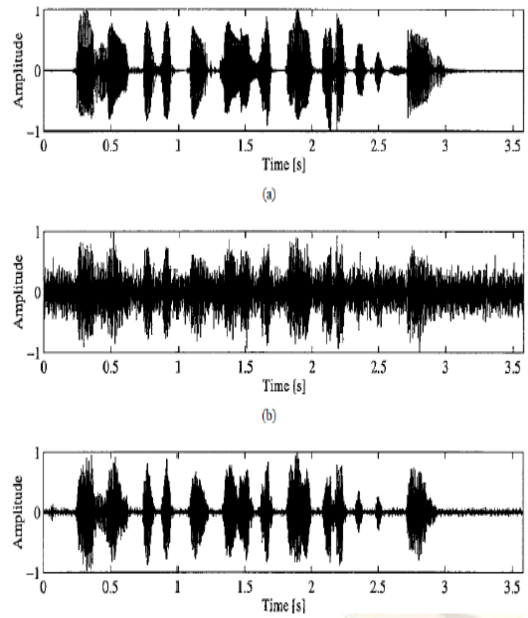

Fig 3. (a) Speech signal (b) Noisy signal (c) Enhanced speech signal

In Fig. 3 There are three types of signals, first is speech signal which is the normal human speech signal which contains background noise, echo, Gaussian noise and many other type of noise. The second one is the noise signal extracted from the speech signal, in this noise is first estimated and them extracted separately. And finally the third signal is of enhanced speech signal where the noise in the speech signal is removed and clean speech signal is obtained for this signal.

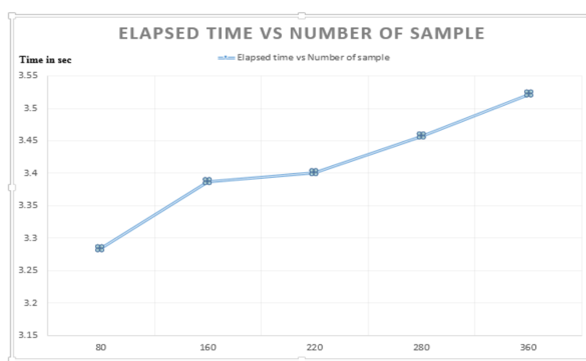

Fig. 4 Elapsed Time vs Number of Sample

As the samples are increased, the time taken to compute or process will also increase, it is shown in Fig. 4 where elapsed time is a time taken to complete the given operation.

After completion of complete program the improved SNR value and Output SNR value is obtained. It shows how much the input SNR is improved and what is the final value of output SNR. This is shown in fig 5 for various samples and finds which is better

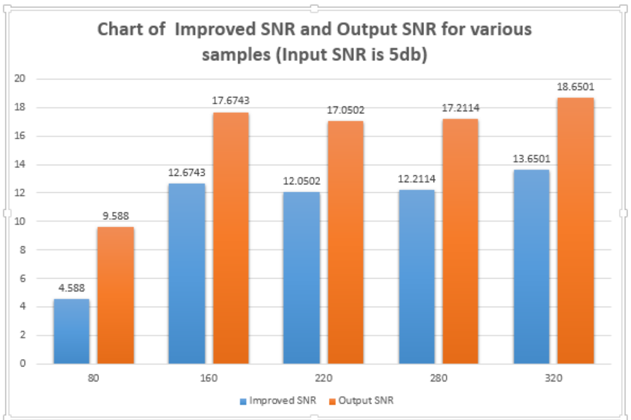

Fig 4.3 Improved SNR and Output SNR for various samples

In this chapter each and every step to compute the simulation process is described briefly and the output of the program is seen. The result obtained is analysed with various samples and SNR value. In this section speech enhancing method based on improved spectral subtraction algorithm was processed. For effective noise reduction with minimal distortion, proposed algorithm takes in account perceptual aspects of human ear. It can be seen from the experimental results that proposed method effectively reduces background noise. Proposed method results in greater improvement of SNR and considerably improvement of perceptual speech quality.

5. Conclusion

This proposed work presents a method to improve speech enhancing based on modified spectral subtraction algorithm was introduced. For effective noise reduction with minimal distortion proposed algorithm takes in account perceptual aspects of human ear. It can be seen from the experimental results that proposed method effectively reduces background noise. The scope of the proposed solution is to improve the SNR value of audio signal as well as to reduce completely background noise without affecting signal power. Proposed method results in greater improvement of SNR and considerably improvement of perceptual speech quality in compassion to conventional spectral subtraction method.

References

[1] Ekaterina Verteletskaya, Boris Simak, ‘Noise Reduction Based on Modified Spectral Subtraction Method in IAENG International Journal of Computer Science, 38:1, IJCS_38_1_10, published in 2011

[2] Zhixin Chen ILX Lightwave Corporation, Bozeman, Montana, USA. Simulation of Spectral Subtraction Based Noise Reduction Method, (IJACSA) in 2011

[3] Anuradha R. Fukane, Shashikant L. Sahare, Noise estimation Algorithms for Speech Enhancement in highly non-stationary Environments , published in 2011

[4] Bo Li1, Yu Tsao, Khe Chai Sim, An Investigation of Spectral Restoration Algorithms for Deep Neural Networks based Noise Robust Speech Recognition in 2009

[5] Yukihiro NOMURAy, Jianming LUy, Hiroo SEKIYAy, and Takashi YAHAGI , Spectral Subtraction Based On Speech/Noise-Dominant Classification at International Workshop on Acoustic Echo and Noise Control, Kyoto, Japan, Sept. 2003.