IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

SPEECH ENHANCEMENT BY BAYESIAN ESTIMATION AND DETECTION: A REVIEW

Sarishma.K, Binesh. K
M-Tech Student, Assistant Professor
Department of Electronics & Communication Engineering,
College of Engineering Thalassery, Kerala, India

Abstract: In speech enhancement, one of the most important tasks is the removal or reduction of background noise from a noisy signal. This paper discusses about various speech enhancement techniques and a general framework proposed to estimate short-time spectral amplitudes (STSA) of speech signals in noise by joint speech detection and estimation to remove or reduce background noise, without increasing signal distortion. By combining parametric detection and estimation theories, the main idea is to take into consideration speech presence and absence in each time-frequency bin to improve the performance of Bayesian estimators. The observed signal is frequently segmented, windowed and transformed into the time-frequency domain. Then, the clean signal coeffcients are usually retrieved by applying an enhancement algorithm to the noisy observations in this domain.

***Index Terms*— speech enhancement, parametric method, joint detection and estimation, Bayesian estimator, minimum mean square error (MMSE).**

I. INTRODUCTION

The basic way of communication for humans to convey their message and thoughts is through verbal communication, that is speech. Generally, range of speech frequency is 20Hz to 20kHz. The main drawback in speech processing is the interference of noise in the speech signal, which inturn reduces the quality and thus speech gets adulterated by the presence of noise. This inturn leads to difficulty in listening which arises poor performance in the virtue of speaker identification and speech recognition.To inculcate this problem speech enhrichment algorithm is used to improve the performance of communication system when the speech signals are adulterated with noise signal.
To remove this noise we have to know the characteristics of the noise and the needed voice signal, so that we can separate noise from the original voice signal. The three main characteristics of signals are,

A. Amplitude
This is the strength of the signal. It can be expressed in a number of different ways (as volts, decibels). The higher the amplitude, the stronger (louder) the signal. The decibel is a popular measure of signal strength.

B. Frequency
The measurement of the number of times that a repeated event occurs per unit of time, expressed in Hertz (Hz), or cycles per second. In speech, we also refer to it as the number of vibrations per second. As we speak, the air is forced out of our mouths, being vibrated by our voice box.

C. Phase
The location of a point within a wave cycle of a repetitive waveform. One complete cycle of a wave begins at a certain point, and continues till the same point is reached again. Phase shift occurs when the cycle does not complete, and a new cycle begins before the previous one has fully completed.

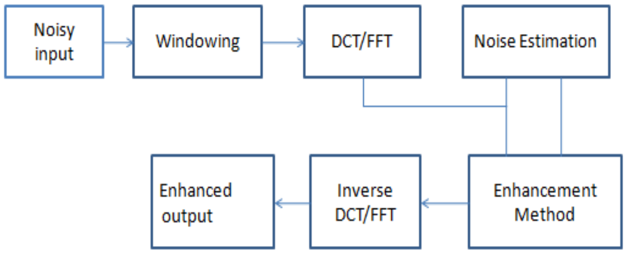

fig.1. block diagram of speech enhancement.

The block diagram of speech enhancement is shown in figure.1. Noise input signal is segmented for 20-30ms samples taken and windowed. Generally hamming window is used for speech is better. Then apply Fourier transform either Discrete or Fast Fourier transform of segmented and windowed. FFT is finest for speech enhancement. Noisy signal obtain and send to noise estimation block. This noise estimation block is used for calculate the overall noise in the original speech. Then if noise estimate is too low, unwanted residual noise will be audible else too high, speech will be unclear. Enhancement block improve speech spectrum is generate and apply the inverse Fourier transform; it gives a clean speech signal[1].

II. SPEECH ENHANCEMENT METHOD

There are several speech enhancement techniques developed for the reduction of background noise and to improve the speech quality. Speech enhancement algorithms are created based on their application. Because only one algorithm is not enough for all types of noise present in the surroundings.

The speech enhancement algorithms can be classified into three types.

1) Filtering Techniques
    i) Spectral subtraction Method
    ii) Wiener Filtering
    iii) Signal Subspace Approach(SSA)
2) Spectral Restoration
    i) Minimum Mean-Square-Error Short-Time Spectral Amplitude Estimator(MMSE-STSA)
3) Speech Model Based

2.1.Spectral Subtraction Algorithm

It is one of the first algorithm proposed for speech enhancement. The spectral subtraction method is a simple and effective method of noise reduction. In this method, an average signal spectrum and an average noise spectrum are estimated in parts of the recording and subtracted from each other, so that average signal-to-noise ratio(SNR) is improved[3].

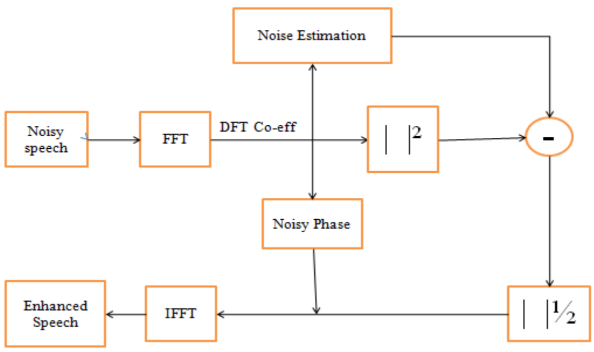

fig.2. block diagram of spectral subtraction.

Block diagram of Spectral Subtraction method is shown in figure 2. First the noisy speech is inputed to reduce the noise content in the speech, so that the listener can actually understand the speech. The noisy speech is a nonstationary signal. To make them stationary, apply framing and windowing techniques to the inputed noise speech signal. After that apply FFT on each short signal. That is time domain signal is converted into frequency domain signal. DFT coefficients is used to eliminate the noise from the noisy speech. After estimating the noise, the estimated noise is subtracted from the noisy speech. Then we get the enhanced speech in time domain after applying the IFFT.

2.2.Wiener Filtering

Wiener Filtering is an alternative method to spectral subtraction for enhancing the speech signal. The wiener filter is a linear filter to recover the clean speech signal from the noisy signal by minimizing the mean square error(MSE) between the estimated signal and the original one[2].

$$H(w)=Ss(w)/(Ss(w)+Sn(w)) \quad (1)$$

Ss(w) is the estimated power spectra of the noise free signal. Sn(w) is the background signal.

Recovered speech signal,

$$\hat{S}(w)=X(w)H(w) \quad (2)$$

H(w) is the transfer function of Wiener filter. X(w) is the noisy observation magnitude spectrum.

2.3.Signal Subspace Method

This method introduce the use of a signal dependant transform to decompose a noisy speech into two separate subspaces. They are signal with noise subspace and the noise –only subspace. Here, Karhuenen Loeve Transform (KLT) is used[3]. Figure 3 is the block diagram representation of signal subspace method.

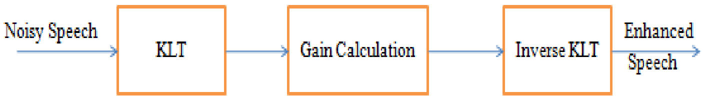

fig.3. block diagram of Signal Subspace Method

2.4.Minimum Mean- Square-Error Short-Time- Spectral Amplitude Estimator(MMSE-STSA)

The Speech Enhancement system MMSE-STSA which focuses on the major importance of the short time spectral amplitude (STSA) of the speech signal. Here the STSA of speech signal is estimated. It is then combined with the short time phase of the noisy speech or the degraded speech for constructing the enhanced signal[4].

It is an optimal STSA estimator which is derived directly from the noisy observations. To derive MMSE STSA estimator, the apriori probability distribution of speech and noise Fourier expansion coefficients should be known. Practically they are unknown, one can think of measuring each probability distribution or assume a reasonable statistical model. This model utilizes asymptotic statistical properties of the Fourier expansion coefficients. We assume that the fourier expansion coefficients of each process can be modeled as statistically independent Gaussian random variables. Mean is of each coefficient is assumed to be zero, since the processes involved here are assumed to have zero mean. The variance of coefficients is time varying due to nonstationarity of speech.

2.5.Model Based Speech Enhancement

In this technique partial spectral reconstruction methods for improving noisy speech signals are included. The reconstruction process is performed on the basis of speech models for the short-term spectral envelope and for the so-called excited signal: the signal that would be recorded directly behind the vocal cords. The idea of model based speech enhancement is first to detect the time–frequency areas that seem to be appropriate for reconstruction. For the successful reconstruction it is necessary that at least a few time-frequency areas have a sufficiently high SNR. These signals are then used to reconstruct those parts with lower SNR.

III. PROPOSED SYSTEM

In speech enhancement, one of the most important tasks is the removal or reduction of background noise from a noisy signal $y[n]$,

$$y[n] = s[n] + x[n] \quad (3)$$

where s and x are respectively the clean signal and independent noise in the time domain and n is the sampling time index.

The speech and noise material which are collected from the NOIZEUS database. It includes the clean speech produced by male and female speakers and also provides the noisy signals. Figure 4 shows the representation of a train noise speech signal having 10dB SNR.

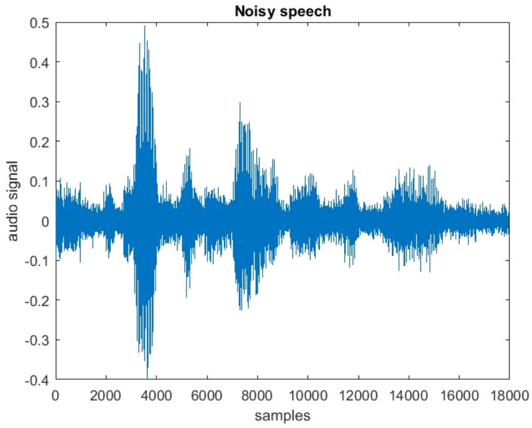

fig.4. representation of a noise speech signal.

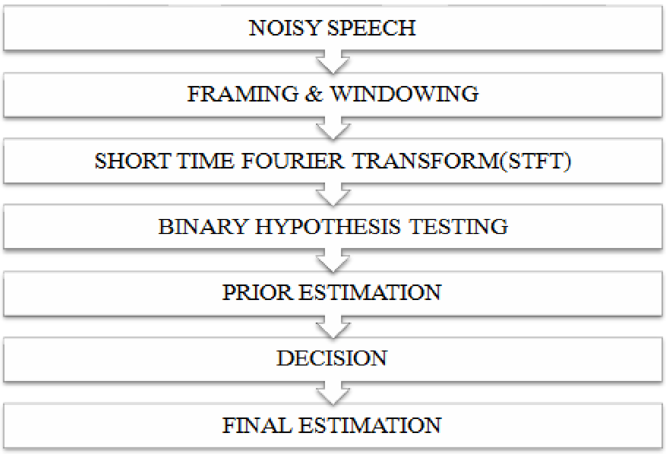

fig.5.block diagram of proposed system

Figure 5 is the block diagram representation of the proposed system. Here the speech signal were sampled at 8kHz, segmented into frames of 256 samples. Applying the window function and then finding the fourier transform .The experiment considers observations in the time-frequency domain after short time Fourier transform (STFT).

In order to take into account the presence and absence of speech, the general framework involves a two-state model, specified by two hypotheses $H0$ and $H1$ for the absence and presence of speech signal, respectively. Specifically, $H0$ models the case where speech is absent or present with little interest, whereas hypothesis $H1$ models the case where speech is present[5].

IV. APPLICATIONS

Speech enhancement system is used to suppress the noise in a noisy speech signal. For robust speech recognition, such a system is used as a preprocessor to a speech recognizer. These techniques helps in removing interference caused by other speakers. Speech enhancement is also used in hearing aids for a better experience. The other applications include the voice communication etc.

V. CONCLUSION

In this paper, an overview of various speech enhancement methods are reviewed. It includes the denoising techniques and algorithms used for the reduction of noise and hence enhance the speech signal. The implementation steps for the proposed system are also discussed. The proposed system can be implemented using MATLAB.

V1.EXPECTED RESULT

The main idea is to take the presence and absence of speech in each time-frequency domain into account so as to improve speech quality and intelligibility in noisy environments. Here it is the combined framework for speech enhancement based on the optimal combination of Neyman-Pearson detectors and Bayesian speech estimators of speech in noise. In contrast to the optimal Bayesian joint detection/estimation, the Neyman-Pearson test may performs the detection without prior knowledge of the speech presence probability.

REFERENCES

[1]"Speech Enhancement by Spectral Subtraction Method "Kaladharan N ,Assistant professor ,Department of Electrical Engineering ,Annamalai University, Annamalai Nagar International Journal of Computer Applications (0975 – 8887) Volume 96– No.13, June 2014.

[2]"Speech enhancement with an adaptive Wiener filter",Marwa A.Abd El-Fattah.Moawad I.Dessouky. Alaa M.Abbas.Salaheldin M.Diab. El-Sayed M.El-Rabaie.

[3]"A signal Subspace Approach for Speech Enhancement",Yariv Ephraim, Fellow, IEEE, and Harry L.Van Trees,Life Fellow, IEEE.

[4]"Speech Enhancement using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimatotr", Yariv Ephraim,student member,IEEE, and David Malah, Senior member, IEEE.

[5]" Optimal Bayesian Speech Enhancement by Parametric Joint Detection and Estimation"
Van-Khanh Mai, Dominique Pastor , (Member, IEEE), and Abdeldjalil Aissa-El-Bey,(Senior Member, IEEE).