© 2017 IJEDR | Volume 5, Issue 2 | ISSN: 2321-9939

Minimization of Noise in Speech Signal Using Mel-Filter

Y.Lakshmi Manasa[1] , P. Rama Krishna[2]
Research Scholar, ECE, Gudlavalleru Engineering College, Vijayawada, India[1]
Assistant Professor, ECE , Gudlavalleru Engineering College, Vijayawada, India [2]

***Abstract* :. Speech signals are usually processed in digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signal. The noise signal does not affect uniformly the speech signal over the whole spectrum .In order to deal with speech improvement in such situations a new spectral subtraction algorithm is proposed for reducing noise from noise corrupted speech. The spectrum is divided into frequency sub-bands based on a nonlinear multiband bark scale. Among the numerous techniques that were developed, the optimal MEL filter can be considered as one of the most fundamental noise reduction approaches, which has been delineated in different forms and adopted in various applications. It shows that in the single-channel case the a posteriori signal-to-noise ratio (SNR) defined by the MEL filter is greater than or equal to the a priori SNR defined by the Gabor filter, indicating that the MEL filter is always able to achieve noise reduction.**

***Keywords* : Speech Signal, Noise Reduction, Mel filter, SNR.**

I. INTRODUCTION

We are living in a environment where the noise is more and ubiquitous, speech signals are generally immersed in acoustic ambient noise and can be recorded in pure form. Therefore, it is essential for speech processing and communication systems to apply effective noise reduction/speech enhancement techniques in order to extract the desired speech signal from its corrupted observations. Noise reduction techniques have a broad range of applications, from hearing aids to cellular phones, voice-controlled systems, multiparty teleconferencing, and automatic speech recognition systems. The choice between using and not using a noise reduction technique may have a significant impact on the functioning of these systems. In multiparty conferencing, for example, the background noise picked up by the microphone at each point of the conference combines additively at the network bridge with the noise signals from all other points. The loudspeaker at each location of the conference therefore reproduces the combined sum of the noise processes from all other locations. Clearly, this problem can be extremely serious if the number of conferees is large, and without noise reduction, communication is almost impossible in this context. Noise reduction is a very challenging and complex problem due to several reasons. First of all, the nature and the characteristics of the noise signal change significantly from application to application, and moreover vary in time. It is therefore very difficult to develop a versatile algorithm that works in diversified environments. Secondly, the objective of a noise reduction system is heavily dependent on the specific context and application. Speech enhancement is a noise suppression technology It has important significance for solving the problem of noise disturbance and it can improve the quality and intelligibility of voice communications. The purpose of speech enhancement is to restore the original signal from noisy observations corrupted by various noises. Most of the existing speech enhancement algorithms only change the magnitude spectrum of the noisy speech. The modified magnitude then recombined with the unchanged phase spectrum to produce a modified complex spectrum, which is the estimated clean speech spectrum. These algorithms are called magnitude spectrum based methods. First of all, the nature and the characteristics of the noise signal change significantly from application to application, and moreover vary in time. It is therefore very difficult to develop a versatile algorithm that works in diversified environments. Secondly, the objective of a noise reduction system is heavily dependent on the specific context and application

II. RELATED WORK

Lots of research has been done in the area of speech processing. In last few years various efficient methods have been proposed for reduction of noise in speech signal. Designing a machine that mimics human behavior, particularly the capability of speaking naturally and responding properly to spoken language, has intrigued engineers and scientists for centuries. Since the 1930s, when Homer Dudley of Bell Laboratories proposed a system model for speech analysis and synthesis , the problem of automatic speech recognition has been approached progressively, from a simple machine that responds to a small set of sounds to a sophisticated system that responds to fluently spoken natural language and takes into account the varying statistics of the language in which the speech is produced[5]. Some noticeable work in area of speech processing is as follows :

In the 1950s, Bell Laboratories developed the first effective speech recognizer for numbers. In 1952, Davis, Biddulph, and Balashek of Bell Laboratories built a system for isolated digit recognition for a single speaker [1], using the formant frequencies measured (or estimated) during vowel regions of each digit. In the late 1960's, Atal and Itakura independently formulated the fundamental concepts of Linear Predictive Coding (LPC, which greatly simplified the estimation of the vocal tract response from speech waveforms.By the 1980s, two distinct types of commercial products were available. The first speakerindependent recognition of small vocabularies[4]. It was most useful for telephone transaction processing. The second, offered by Kurzweil Applied Intelligence, Dragon Systems, and IBM, focused on the development of large-vocabulary voice recognition systems so that text documents could be created by voice dictation.Over the past two decades, voice recognition technology has developed to the point of real-time, continuous speech systems that augment command, security, and content creation tasks with exceptionally high accuracy[7]. Other systems developed under DARPA's SUR program included CMU's Hearsay(-II) and BBN's HWIM , Neither Hearsay-II nor HWIM (Hear What I Mean) met the DARPA program's performance goal at its conclusion in 1976. However, the approach proposed by Hearsay-II of using parallel asynchronous processes that simulate the component knowledge sources in a speech system was a pioneering concept.

III. PROPOSED METHOD BLOCK DIAGRAM

A filter bank is an array of band-pass filters which separates the input signal into no of components, each one carrying a single frequency sub-band of the original signal in signal processing. One application of a filter bank is a graphic equalizer, which can attenuate the components differently and recombine them into a modified version of the original signal. The process of decomposition performed by the filter bank is called analysis (meaning analysis of the signal in terms of its components in each sub-band)the output of analysis is referred to as a sub band signal with as many sub bands as there are filters in the filter bank. The reconstruction process is called synthesis, meaning reconstitution of a complete signal resulting from the filtering process.

Fig .1: Block Diagram

In digital signal processing, the term filter bank is also commonly applied to a bank of receivers. The difference is that receivers also down-convert the sub bands to a low center frequency that can be re-sampled at a reduced rate. The same result can sometimes be achieved by under sampling the band pass sub bands. Another application of filter banks is signal compression when some frequencies are more important than others. After decomposition, the important frequencies can be coded with a fine resolution. Small differences at these frequencies are significant and a coding scheme that preserves these differences must be used. On the other hand, less important frequencies do not have to be exact. A coarser coding scheme can be used, even though some of the finer (but less important) details will be lost in the coding. The vocoder uses a filter bank to determine the amplitude information of the sub bands of a modulator signal (such as a voice) and uses them to control the amplitude of the sub bands of a carrier signal (such as the output of a guitar or synthesizer), thus imposing the dynamic characteristics of the modulator on the carrier.

The mel scale, named by Stevens, Volkmann, and Newman in 1937,is perceptual scale of pitches judged by listeners to be equal in distance from one another. Above about 500 Hz, increasingly large intervals are judged by listeners to produce equal pitch increments. The name mel comes from the word melody to indicate that the scale is based on pitch comparisons. One approach to simulating the subjective spectrum is to use a filter bank, one filter for each desired Mel frequency component. The filter bank has a triangular band pass frequency response, and the spacing as well as the bandwidth is determined by a constant mel-frequency interval. The information carried by low frequency components of the speech signal is more important compared to the high frequency components. In order to place more emphasize on the low frequency components, mel scaling is performed.

A popular formula to convert f hertz into m mels is

mel(f)= 2595*log10(1+f/700)

A Channel can take many forms a connection between initiating and terminating nodes of a circuit. A path for conveying electrical or electromagnetic signals, usually distinguished from other parallel paths. A storage which can communicate a message over time. In a Communication system, the physical or logical link that connects a data source to a data link. The input voice speech signal is given, in that noise is present. The noise and the speech signal is given to the Mixing System. Mixing System is used to mix up both the signals and is given for the identification of noise .The noise is identified and it is given to the MEL-Filter. MEL filter is used to reduce the noise in the speech signal .By using MEL filter we have to re construct the speech signal. Then after removing the noise by using MEL filter we have to display the images. Here MEL filter is used to convert the time signal in to frequency form. By using the Discrete short time Fourier transform we have to reconstruct the signal. Mel filter banks are non-uniformly spaced on the frequency axis, so we have more filters in the low frequency regions and less number of filters in high frequency regions.

PROPOSED METHODOLOGY

The flow chart shown in fig 2 explains the procedure to reduce the noise present in speech signal by using MEL Filter.

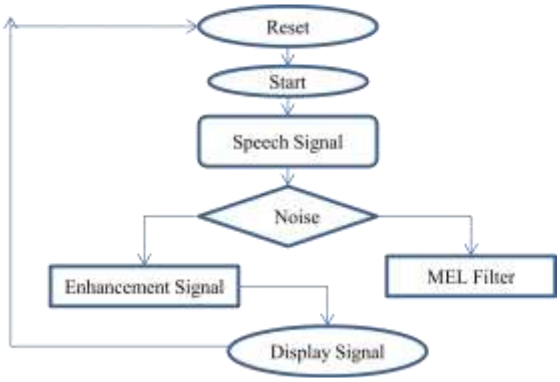

Fig.2. Flow Chart

Firstly the speech signal is taken in that we have to recognize whether the noise is present or not. If there is no noise display the output enhancement signal, if noise is present mel filter is used to reduce the noise and after reduction display the signal.

IV. EXPERIMENTAL RESULTS

Male Voice Signal :

Here Fig 3,4 shows the male voice signal "HAI FRIENDS" is transmitted in that noise is present.

Fig : 3 Male Voice Signal

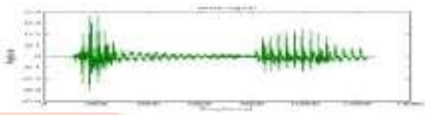

Fig :4 Noise Signal

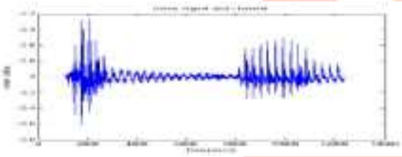

Fig : 5 Noise Signal with Channel

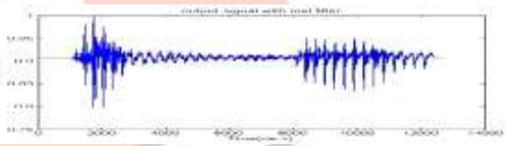

Fig :6 Output Speech Signal with MEL-Filter

Fig: 5 shows that the noise signal with voice signal is given to the channel. The mixing of these signals is given to the mel filter for reduction of noise.Fig:6 is the speech signal after reduction of noise.

Female Voice Signal Results :-

Here Fig 7,8 shows the Female voice signal "HAI " is transmitted in that noise is present

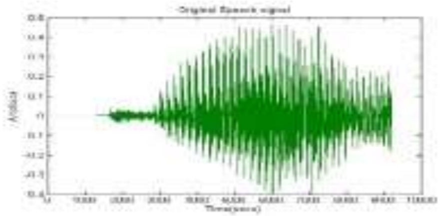

Fig :7 Female Voice Signal

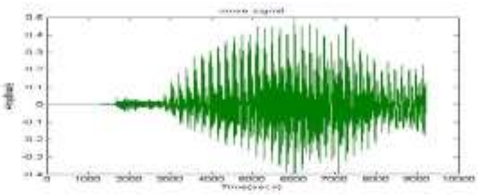

Fig :8 Noise Signal

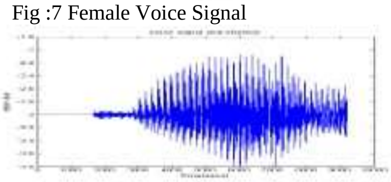

Fig :9 Noise Signal with Channel

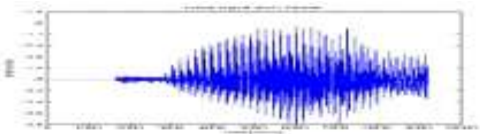

Fig :10 Output Speech Signal with MEL-Filter

Fig: 9 shows that the noise signal with voice signal is given to the channel. The mixing of these signals is given to the mel filter for reduction of noise. Fig: 10 is the speech signal after reduction of noise.

V. SNR,PSNR,MSE VALUES

Signal-to-noise ratio is a measure used in science and engineering that compares the level of a desired signal to the level of background noise. It is defined as the ratio of signal power to the noise power, often expressed in decibels

$$\mathrm{SNR} = 10\log \frac{\text{Signal Power}}{\text{Noise Power}}$$

Peak signal-to-noise ratio, often abbreviated PSNR, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. To compute the PSNR, the block first calculate the mean squared error using the following equation

$$\mathrm{MSE} = \sum_{M,N} \frac{[I1(m,n) - I2(m,n)]}{M*N}$$

$$\mathrm{PSNR} = 10\frac{\log_{10} R^2}{MSE}$$

COMPARISON **:**

NAME	MALE VOICE	FEMALE VOICE
SNR	37.96dB	29.26 dB
PSNR	18.59dB	16.18 dB
MSE	0.985	0.856

VI. CONCLUSION

An implementation of employing mel filtering to speech processing had been developed. As has been previously mentioned, the purpose of this approach is to reconstruct an output speech signal by making use of the accurate estimating ability of the mel filter. True enough, simulated results from the previous chapter had proven that the Mel filter indeed has the ability to estimate accurately. PSNR,SNR values is increased and it is compared with the existing method.

REFERENCES

[1] W. Bastiaan Kleijn, R. C. Hendriks,"A Simple Model of Speech Communication and its Application to Intelligibility Enhancement" IEEE SIGNAL PROCESSING LETTERS, VOL. 22, NO. 3, MARCH 2015.

[2] Sayf A. Majeed, Hafizah Husain, Salina Abdul Samad, Tariq F. Idbeaa," mel frequency cepstral coefficients (mfcc)feature extraction enhancement in the application of speech recognition"in Journal of Theoretical and Applied Information Technology, VOL. 79, NO. 1,September 2015.

[3] J. B. Crespo and R. C. Hendriks, “Multizone speech reinforcement,” IEEE/ACM Trans. on Audio, Speech, Lang. Process., vol. 22, no. 1, pp. 54–66, 2014.

[4] J. Taghia and R. Martin, “Objective intelligibility measures based on mutual information for speech subjected to speech enhancement processing,” IEEE Trans. Audio, Speech Lang. Process., vol. 22, pp. 6–16, Jan. 2014.

[5] R. Houben, J. Koopman, H. Luts, K. Wagener, A. van Wieringen, H. Verschuure, and W. Dreschler, “Development of a Dutch matrix sentence test to assess speech intelligibility in noise,” Int. J. Audiol., pp. 1–4, 2014.

[6] C. Taal, J. Jensen, and A. Leijon, “On optimal linear filtering of speech for near-end listening enhancement,” IEEE Signal Process. Lett., vol. 20, no. 3, pp. 225–228, 2013.

[7] P. N. Petkov, G. E. Henter, and W. B. Kleijn, “Maximizing phoneme recognition accuracy for enhanced speech intelligibility in noise,”IEEE Trans. Audio, Speech Lang. Process., vol. 21, no. 5, pp. 1035–1045, 2013.

[8] R. C. Hendriks, T. Gerkmann, and J. Jensen, DFT-Domain based single-microphone noise reduction for speech enhancement-a Survey of the State of the Art. San Rafael, CA, USA: Morgan & Claypool, 2013.

[9] T. C. Zorilă, V. Kandia, and Y. Stylianou, “Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression,” in ISCA Interspeech, Portland, OR, USA, 2012.

[10] “American national standard methods for calculation of the speech intelligibility index,” ANSI/ASA S3.5-1997 (R2012) 2012.