A Survey Paper on Speech Enhancement Algorithms

Priyanka V. Patil[1] Vikram A. Mane[2]
[2]Assistant Professor
[1,2]Department of Electronics & Telecommunication Engineering
[1]Shivaji University, Kolhapur [2]Annasaheb Dange College of Engineering, Ashta Affiliated to Shivaji University, Kolhapur, India

Abstract— Speech enhancement aims to improve speech quality by using various algorithms. The objective of enhancement is improvement in intelligibility and/or overall perceptual quality of degraded speech signal using audio signal processing techniques. Speech extraction and enhancement is widely required in so many applications like Hearing Aid design, Music and audio Processing etc. The speech enhancement methods are studied in a large number of investigations. Still so many speech recognition systems perform poor in noisy environmental condition. It has been always remained challenge of extracting and enhancing the speech from the noisy speech. The objective of this study is a comparative study between different methods of speech enhancement. This paper tries to give an idea about the previous researches and their findings about the enhancement of required speech from noisy speech.

Key words: Minimum Mean Square Error (MMSE), Short-Time Spectral Amplitude (STSA), Signal to Noise Ratio (SNR), Noise-To Mask Ratio (NMR), Diagnostic Rhyme Test (DRT), Semantically Unpredictable Sentences (SUS), Mean Squared Error (MSE), Ideal Binary Mask (IBM)

I. INTRODUCTION

Speech is a primary form of communication between human beings. It existed since human civilization. Humans can separate a particular sound from a mixture of many sources but such a task remains a major challenge for machines. Machine requires a particular algorithm to separate speech from noise. In designing of speech enhancement algorithms, it is necessary to achieve hearing protection or increase listening comfort. Speech enhancement technique is useful to increase the efficiency of speech communication, reduce listener fatigue. Improving quality, however, might not necessarily lead to improvement in intelligibility. In fact, in some cases improvement in quality might be accompanied by a decrease in intelligibility. This is due to the distortion imparted on the clean speech signal resulting from excessive suppression of the noisy signal. Here, it is necessary to improve quality of speech with increase the speech intelligibility and to make the corrupted speech more pleasant to the listener.There are various researches are going on for the study of speech enhancement. In this paper overview of various works by researchers are given.

II. COMPARATIVE STUDY OF ALGORITHMS USED FOR ENHANCEMENT OF SPEECH

There are different algorithms are used for enhancement of speech. The algorithms used for enhancement of a speech are as follows:

A. Suppression of Acoustic Noise in Speech using Spectral Subtraction

The speech with background noise can degrade the performance of digital voice processors used for applications. The performance of digital voice systems must be maintained at a level near that measured using noise-free input speech. To ensure continued reliability, the effects of background noise can be reduced by using noise-cancelling microphones, internal modification of the voice processor algorithms. In year 1979, Steven Boll has studied the method of the Suppression of Acoustic Noise in Speech Using Spectral Subtraction [1]. The block diagram of this technique is as shown in fig 1.These researchers have developed noise suppression technique, implemented a computationally efficient algorithm, and tested its performance in actual noise environments.

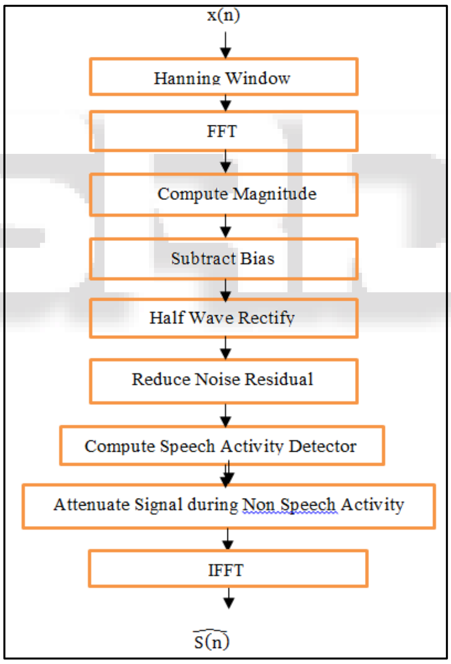


Fig. 1: Block diagram Suppression of Acoustic Noise in Speech Using Spectral Subtraction.

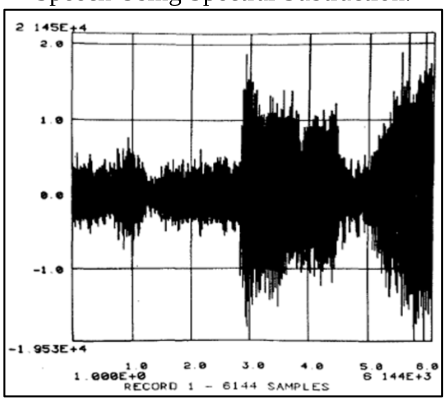


Fig. 2: Time waveform of helicopter speech. “Save your”.

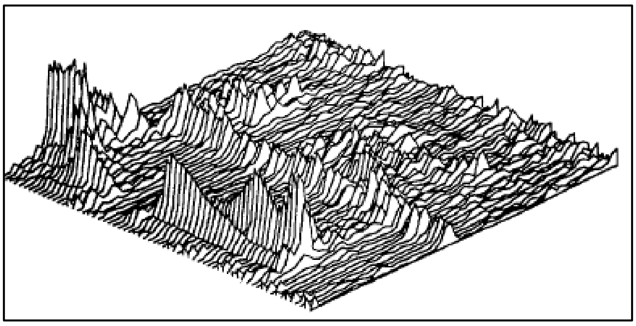


Fig. 3: Short-time spectrum of helicopter speech.

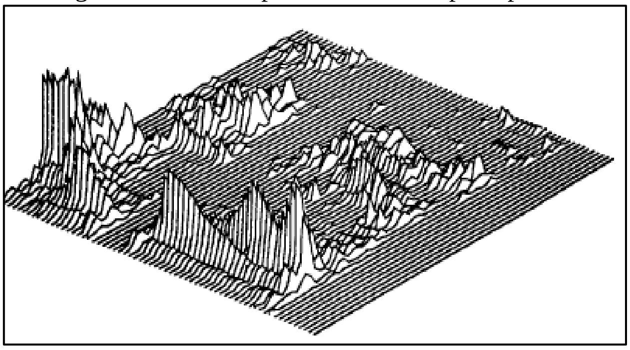


Fig. 4: Short-time spectrum using bias removal, half-wave rectification, residual noise reduction, and non speech signal attenuation (helicopter speech).

When we use this technique for speech enhancement, we have to estimate the current noise spectrum using the average noise magnitude measured during non-speech activity. After that we can estimate the magnitude frequency spectrum of the underlying clean speech by subtracting the noise magnitude spectrum from the noisy speech spectrum [1].

The methods for speech enhancement need to be eliminate a "musical residual noise" with an unnatural structure. But, the system for suppression of acoustic noise in speech using spectral subtraction cannot eliminate this noise. It is difficult to suppress noise without decreasing intelligibility and without introducing speech distortion and residual noise. In very noisy situations, the enhanced speech can be even more disturbing to a human listener than the original corrupted speech signal.

B. Speech Enhancement Using a- Minimum Mean Square Error Short Time Spectral Amplitude Estimator

Yariv Ephraim and David Malah have developed one algorithm for Speech Enhancement Using a Minimum Mean Square Error (MMSE) Short-Time Spectral Amplitude Estimator in 1984. This is an algorithm for enhancing speech degraded by uncorrelated additive noise when the noisy speech alone is available. The basic approach taken in this method is to optimally estimate the short-time spectral amplitude (STSA) and complex exponential of the phase of the speech signal rather than optimally estimating the STFT itself. Since the STSA of a speech signal is more important than its waveform in speech perception. The STSA and the complex exponential cannot be estimated simultaneously in an optimal way.

Therefore, we have to use an optimal MMSE STSA estimator and combine it with an optimal MMSE estimator of the complex exponential of the phase which does not affect the STSA estimation. The latter constrained complex exponential estimator is found to be the complex exponential of the noisy phase. The MMSE STSA estimator and the Wiener STSA estimator which results from the optimal MMSE STFT estimator are nearly equivalent at high SNR. On the other hand, the MMSE STSA estimator results in significantly less MSE and bias when the SNR is low. This fact supports approach to optimally estimate the perceptually important STSA directly from the noisy observations rather than deriving it from another estimator (e.g., from the Wiener one). The a priori SNR is a key parameter of the STSA estimator.

When we use different estimators for the a priori SNR then it results in different STSA estimations. In this algorithm a "decision-directed'' method is used for estimating the a priori SNR. This method is useful when it is applied to either the MMSE or the Wiener STSA estimator. Then we have to combine this estimator with the MMSE STSA estimator which takes into account the uncertainty of signal presence in the noisy observations to obtain the best speech enhancement. The proposed approach results in a significant reduction of the noise and provides enhanced speech with colorless residual noise [3].

C. Speech Enhancement Based on Audible Noise Suppression

Dionysis E. Tsoukalas, John N. Mourjopoulos and George Kokkinakis in year 1997 studied the method of Speech Enhancement Based on Audible Noise Suppression. This enhancement approach is based on the definition of an audible noise component of the STSA. The performance of the proposed technique was evaluated using objective measures such as the SNR and the noise-to mask ratio (NMR). Furthermore, the technique was assessed by the diagnostic rhyme test (DRT) and the semantically unpredictable sentences (SUS) test. From these tests, it was found that, at very low SNR's (5dB), significant improvements could be achieved by the proposed method. This technique could achieve speech reconstruction for arbitrary low SNR's given the correct sparse data. This speech enhancement System which uses audible noise suppression technique gives smaller but significant intelligibility gains [4].

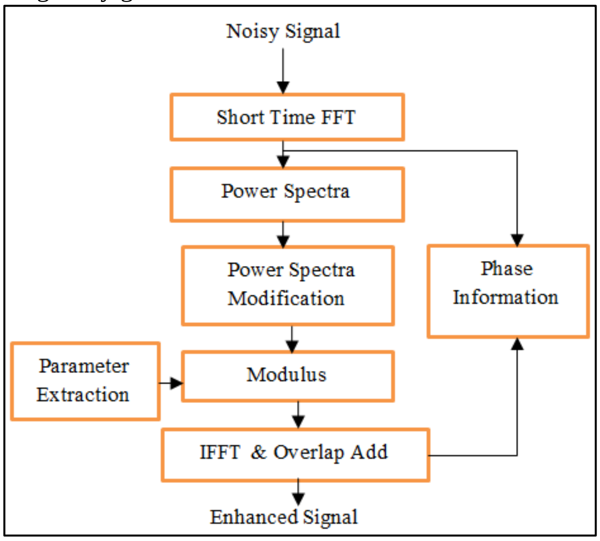


Fig. 5: General block diagram for the ANS technique

As shown in fig 5, the input for this technique is a noisy speech. After providing the input to the algorithm the short-time windows of the noisy speech will transform into the frequency domain using the short-time fast Fourier transform (STFFT).The power spectrum of the noisy speech is obtained and the phase information is extracted. The power spectrum of the noisy speech is processed using the nonlinear law given by in conjunction with the previously estimated parameters and per CB. The modulus of the modified power spectrum is transformed back into the time domain using the short-time inverse Fast Fourier transform (FFT) and the original (noisy signal) phase information. The enhanced speech is reconstructed using the overlap-add method.

An enhancement scheme presented in this paper is based on the utilization of a well-known auditory mechanism, noise masking. But, when no additional information is available on the nature of noise degradation, then this system cannot provide the speech enhancement.

D. A Modified Speech Enhancement Algorithm Based on the Subspace

The another algorithm for a speech enhancement based on the subspace is developed by Hairong Jia,Xueying Zhang and Chensheng Jin in 2009 [5]. The whole speech signal space will be divided into   noise subspace and signal plus noise subspace through decomposition space, the optimal linear estimator will be structured by the noise eigenvalue matrix and the noisy speech eigenvalue matrix in the subspace. How to gain their true eigenvalue matrix are the key. In the traditional subspace, the eigenvalue matrix of noise are attained by eigen-decomposing to covariance matrix of noise, but covariance matrix of noise is estimated by using variance in the silence sequent, it cannot instead the whole noise, and lead to residual noise. In addition, the noisy speech eigenvalue matrix minus noise’s equal directly speech’s, the result increases the residual noise.

To solve the question, the modified method tracks real time noise eigenvalue matrix in the subspace domain by applying statistical information in the whole time, and corrects speech eigenvalue matrix making use of the principle of winner filtering, then produces the better true eigenvalue matrix of speech. At last, speech will be restored. We can see the comparison between the modified Speech Enhancement Algorithm Based on the Subspace and the Tradition’s from Spectrogram in fig .In the fig the dark color denote the spectrum of speech, the tint color express the spectrum of noise.

Fig 6(a) shows the clean speech which is not affected by the noise and fig 6(b) shows the noisy speech which is the combination of clean speech and noise.

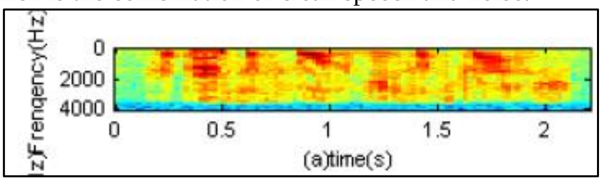
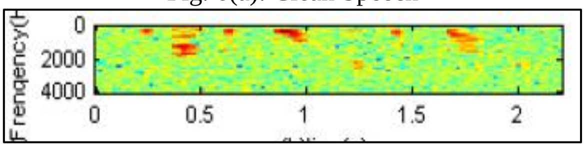


Fig. 6(a): Clean Speech



(b) time (s)
Fig. 6(b): Noisy Speech.

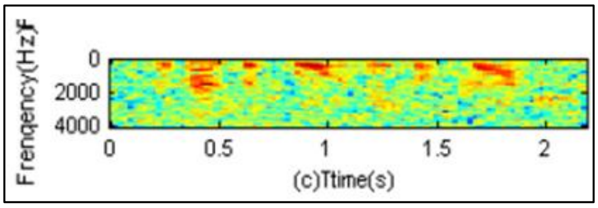


Fig. 6(c): Enhanced Speech of Traditional Algorithm

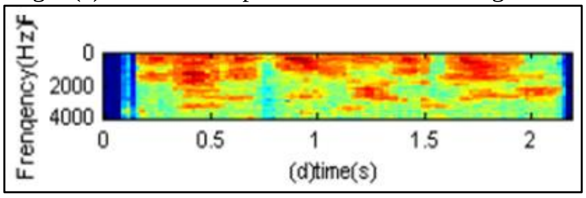


Fig. 6(d): Enhanced Speech of Modified Algorithm

After applying the traditional algorithm on noisy speech, we get the result as shown in fig3(c). Spectrogram of clean speech and processed speech is different. It means that at the end of the traditional algorithm we will not get the exact same speech as clean speech. But, when we apply the modified algorithm which is based on the subspace on the noisy speech then we will get a result spectrogram as shown in fig3 (d).From the fig d we can say that the spectrum of enhanced speech using modified algorithm is very clear by eliminating the spectrum of noise, it is the same as the speech in the fig 3(a).The speech enhancement based on the subspace is the better than the traditional algorithm.

E. Speech Enhancement Using Ideal Binary Mask

Up to now, we can see that none of the existing algorithms are designed to improve speech intelligibility. The statistical-model based algorithms derive the magnitude spectra by minimizing the mean squared error (MSE) between the clean and estimated (magnitude or power) spectra. The MSE metric, however, pays no attention to positive or negative differences between the clean and estimated spectra. A positive difference between the clean and estimated spectra would suggest attenuation distortion, while a negative spectral difference would suggest amplification distortion. If we can somehow manage or control these two types of distortions, then we should expect to receive large gains in intelligibility [6]. The block diagram of the system is as shown in fig 4.

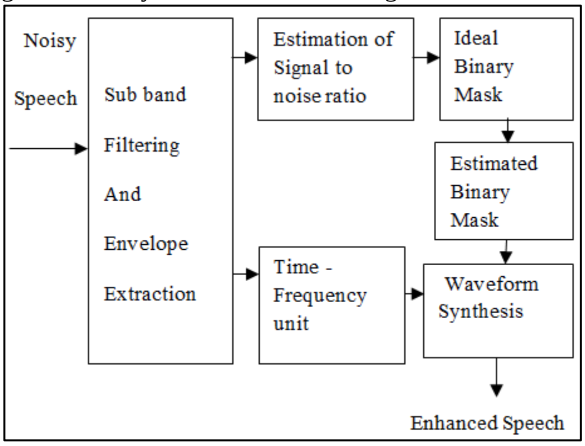


Fig. 4: System Block Diagram for speech enhancement using Ideal Binary Mask

Speech signals are often corrupted by noise, therefore an effective system for speech enhancement and noise suppression is necessary in many applications. There is one simple method for speech enhancement is Ideal Binary Mask [6]. Recently, the concept of speech enhancement using Ideal Binary Mask (IBM) has received attention. With the help of IBM method it is possible to increase the speech quality and intelligibility.

For that first we have to collect the noisy speech data base which consisting speech samples captured by noises and interference from other speech. This algorithm will decompose the input speech into Time-Frequency units with the help of sub band filter. After decomposing of the speech, it needs to calculate the signal to noise ratio. The next step is Applying Ideal Binary Mask to noisy speech. The ideal binary mask (IBM) identifies speech dominated and noise dominated units in a T-F representation of noisy speech according to their signal to noise ratio. Speech dominated unit labeled as 1 and noise dominated unit labeled as 0.Time-Frequency unit and estimated binary mask will combine to produce the enhanced speech waveform. Finally, algorithm will give enhanced speech [7].

III. Conclusion

By the literature review it is seen that all algorithms are developed for improving the quality of speech. The speech enhancement algorithms perform important role in an applications like Hearing aid devices and Audio processing devices to get clear and enhanced speech as a result. Increment in the intelligibility of the speech will make the corrupted speech more pleasant to the listener. Most of the algorithms try to improve only the quality of the speech not the intelligibility. In some cases improvement in quality might be accompanied by a decrease in intelligibility. There is need to improve intelligibility of speech.

References

[1] Steven Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, IEEE transactions on acoustics, speech, and signal processing, vol. assp-27, no. 2, April 1979.

[2] M. Berouti, R. Schwartz, and R. Makhoul, “Enhancement of speech corrupted by acoustic noise,” IEEE International Conference on Acoustics, Speech and Signal Processing, 1979.

[3] Yariv Ephraim and David Malah,“Speech Enhancement Using a- Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator”, IEEE transactions on acoustics, speech, and signal processing, vol. assp-32, no. 6, December 1984.

[4] Dionysis E. Tsoukalas, John N. Mourjopoulos, “Speech Enhancement Based on Audible Noise Suppression” IEEE transactions on speech and audio processing, vol. 5, no. 6, November 1997.

[5] Hairong Jia Xueying Zhang Chensheng Jin,”A Modified Speech Enhancement Algorithm Based on the Subspace” Second International Symposium on Knowledge Acquisition and Modeling, 2009.

[6] Philipos C. Loizou, Senior Member, IEEE, and Gibak Kim,”Reasons Why Current Speech-Enhancement Algorithms Do Not Improve Speech Intelligibility and Suggested Solutions” IEEE transactions on audio, speech, and language processing, vol. 19, no. 1, January 2011.

[7] Raphael Koning, Nilesh Madhu, and Jan Wouters “Ideal Time–Frequency Masking Algorithms Lead to Different Speech Intelligibility and Quality in Normal-Hearing and Cochlear Implant Listeners” IEEE transactions on biomedical engineering, vol. 62, no. 1, january 2015.