

Speech Enhancement by using Ideal Binary Mask

Priyanka V. Patil¹, Vikram. A. Mane²

¹(Department of Electronics and Telecommunication, Shivaji University, Kolhapur)

²(Assistant professor, Department of Electronics and Telecommunication, AnnasahebDange College of Engineering and Technology, Ashta affiliated to Shivaji university, Kolhapur, India)

Abstract:

Communication via speech is one of the essential functions of human beings. Humans possess varied ways to retrieve information from the outside world or to communicate with each other. Three most important sources of information are speech, images and written text. For many purposes, speech is more efficient and convenient way. Speech not only conveys linguistic contents, but also communicates other useful information like the mood of the speaker. In speech communication systems the quality and intelligibility of speech is most important for easy and accurate exchange of information. In realistic listening situations, speech is often contaminated by various types of background noise. The presence of background noise degrades the quality as well as intelligibility of speech. This degradation of speech by noise creates problems not only for just interpersonal communication but more serious problems in applications in which decision or control is made on the basis of speech signal. So, there is need to develop a new method for speech enhancement which will improve quality as well as intelligibility of the speech. This paper presents method of speech enhancement by using Ideal Binary Mask. This is a simple method for enhancing the speech with the objective to eliminate the additive noise present in speech signal and restore the speech signal to its original form with improved quality as well as intelligibility of speech.

Keywords—Ideal Binary Mask (IBM), Signal to Noise Ratio (SNR), Fast Fourier Transform (FFT), Inverse Fast Fourier Transform (IFFT).

I. INTRODUCTION

Speech is a primary form of communication between human beings. People can express their ideas, emotions, thoughts and knowledge about any subject with the help of speech. It also helps us to understand emotion and thoughts of the others. Speech communication plays important role in human life. The quality and intelligibility are the most important things of speech signal for easy and accurate exchange of information. Intelligibility refers to the 'understandability' of speech, which is helpful to understand the intention of the speaker and the response of the listener and also to communicate effectively in everyday situations. Most of the time, speech is corrupted by different noises. These noises reduce the quality and intelligibility of speech signal. Enhancement algorithms increase the quality of speech, but unable to increase the speech intelligibility. In fact, in some cases improvement in quality might be lead to decrease in intelligibility [11]. A simple and effective method for speech enhancement is Ideal Binary Mask (IdBM) method which has the

Ideal binary mask (IBM) identifies speech dominated and noise dominated units, a binary mask is computed and applied to the noisy input spectrum to get the noise-suppressed spectrum. Ideal binary mask (IdBM) defined by comparing the local signal-to-noise ratio (SNR) of each TF unit against a fixed threshold. T-F units with local SNR higher than the threshold are defined as target-dominated units, while others as masker-dominated unit. This binary mask simply keeps the speech dominated (SD) units while zeroing out the noise-dominated (ND) units [1]. A high increase in intelligibility can be obtained by applying the target binary mask to noisy speech.

This method of speech enhancement cancels the interference of noise and other speech from targeted speech using binary Ideal Mask technique and improves SNR of speech. This method helps to increase the intelligibility of the speech and to make the corrupted speech more pleasant to the listener. It is also useful to increase the efficiency of speech communication and to

reduce listener fatigue. It helps to achieve hearing protection or increase listening comfort [6].

II. SPEECH ENHANCEMENT TECHNIQUE

The quality and intelligibility are the most important things of speech signal for easy and accurate exchange of information. The proposed method of speech enhancement by using Ideal Binary Mask helps to increase the quality as well as intelligibility of corrupted speech by noise. Most of the time speech is corrupted by noise. Noise is the unwanted part of the signal. Unwanted frequency signal is present in various degrees around whole environments.

The noisy speech signal is a combination of clean speech $S(t)$ and background noise $N(t)$ [1] as shown in fig.1. The noisy speech signal $X(t)$ is described as,

$$X(t) = S(t) + N(t)$$

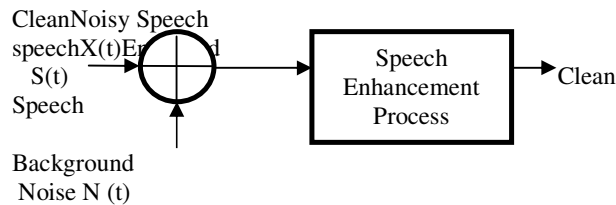


Fig1. Generation of noisy speech

A. Sources of Noise –

The background noise is added to the desired speech signal which is transmitted to the listener. Noise can be characterized as any disturbance that tends to obscure a desired signal. It is usually caused by different sources. Based on the nature and properties of the noise sources, noise can be classified as Babble noise, Car Ignition Noise, Helicopter noise, Natural Noise, Kids Cheering Voice, Transportation Noise, Random noise, Tractor Noise.

1) Babble noise: Babble means to talk without a particular goal. When everyone is babbling without paying attention to their neighbors, like the kid then the created voice is called as babble noise.

Persons talking face to face are always surrounded by too many other people. Always background noise is present by talking of them continuously. It is considered as a multi-talker babble noise. Babble noise will be overlapping on the speech conversation. It is actually very challenging to recognize speech signals in presence of background babble noise for the normal listeners and hearing impaired persons. Clean speech signals are corrupted by background noise of multi-talker babble is shown in fig 4 and 5. Interfering of babble noise decreases speech intelligibility and quality.

2) Car Ignition Noise: At the time of some of the conversation it is possible that person may be out of his home. The road traffic vehicles always generate different noises. Generally all the vehicles are travelled by specific types of engine. The car ignition voice, engine voice may also interrupt the speech. These engines generate noise in form spark plug, motors and burn fuel exhaust systems of large heavy vehicles. Here, the car ignition voice has taken as a noise in speech. Nature of car ignition noise added with clean speech is shown in fig.8 and 9. This noise degrades the speech quality and intelligibility.

3) Helicopter noise: In most of the big cities because of air ways also very huge amount of traffic is generated. Engine noise of low flying air crafts, charted plain, helicopters, air force plain have added very cumbersome noise in speech signal. Helicopters generate substantial noise levels, especially at low frequency. However, the low frequency content masks the speech, the speech recognition and intelligibility reduces. Fig.6 and 7 shows the noisy speech which are affected by Helicopter noise.

4) Natural Noise: Natural noise is a noise produced by natural sources. This includes the sounds of any living organism, from insect larvae to the largest living mammal on the planet, whales,

and non-biological sources. Non-biological sound includes the effects of water by a stream or waves at the ocean, the effects of wind in the trees or grasses, and sound generated by the earth. Here, birds chirp with water flowing voice has taken as a natural noise which causes to degradation of speech quality and intelligibility. Fig 10 and 11 show the waveforms for noisy signal with natural noise.

5) Kids Cheering Voice: When some kids plays on the playground then other kids cheer for encouraging those kids who playing on playground. This cheering voice creates large sound and act as a noise. The noise of this cheering has taken as a kids cheering voice for the experiment. This voice reduces quality and intelligibility of speech. Fig.12 and 13 represents the noisy signal with a kids cheering voice.

6) Transportation Noise: Fig 14 and 15 show the basic nature waveforms of noisy speech in Transportation Noise. Generally road traffic vehicles are most important type of noise in the social life. Transportation noise relates to noise generated by road vehicles, aircrafts, trains, vessels. This transportation noise affects the communication signals. Transportation noise leads to degradation of the quality and intelligibility of speech signal.

7) Random noise: Random noise is usually electric or acoustic signal that consists of equal amounts of all frequencies. Noise consist a large number of transient disturbances with a statistically random time distribution. This noise affects the speech signal when add in it as shown in fig 16 and 17. It reduces quality and intelligibility of speech signal and become hard to understand the particular speech signal.

8) Tractor Noise: Tractor is a vehicle having a powerful gasoline or diesel motor and usually large, heavily treaded rear tires, used especially for pulling farm implements or machinery. The tractor

creates a high level noise when it mixes with a speech as shown in fig 18 and 19. It reduces quality as well as intelligibility of speech.

This noisy environment reduces the speaker and listeners ability to communicate. There is need to remove all unwanted background noise and to improve the speech. Hence, speech enhancement described in a proposed work is necessary to avoid the degradation of speech quality and intelligibility. Speech enhancement is performed to reduce noise without distorting speech. To reduce the effect of different noises, speech enhancement using Ideal Binary Mask tries to remove as much noise as possible from the mixture of the target speech and the interfering sound, with the objective of increasing speech intelligibility and improving the speech quality of the processed signal.

III. SPEECH ENHANCEMENT BY USING IDEAL BINARY MASK

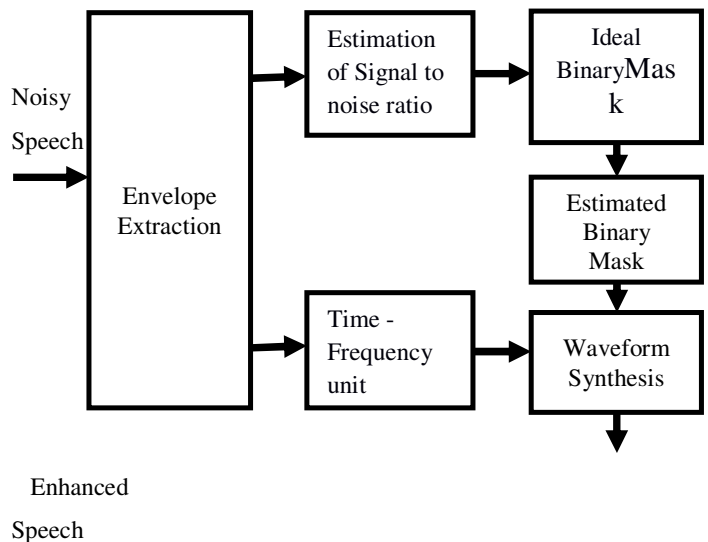


Fig2. Block Diagram Speech Enhancement Method by using Ideal Binary Mask.

A block diagram of the proposed work is shown in Fig.2. To illustrate a speech signal with sampling rate of 8 kHz in .wav format is considered. A different type of background noise has taken for creating the different input noisy

speech for analysis. Envelope extraction is done by taking short windows and processing them. The short window of signal like this is called frame. Here the total samples of noisy speech are divided into frame size of 32ms ($8000 \times 32 \times 0.001 = 256$ samples) each by using hanning windowing method.

A long noisy speech signal is multiplied with a hanning window function of finite length, giving finite length version of the original signal. The speech is first windowed for FFT analysis. The processing takes place within an FFT based short-time spectral analysis modification- synthesis framework. Fast Fourier Transform (FFT) algorithm computes the Fourier transform of a sequence. Fourier analysis converts a signal from its original domain (often time) to a representation in the frequency domain. Here zero padding is used, if there is need to increase the length of signal for exactly division of speech signal into frames. The proposed Experiment is performed on the basis of signal to noise ratio (SNR) of each frame of noisy speech. The SNR is calculated for each frame.

The ideal binary mask is computed on the basis of signal-to-noise ratio (SNR) by comparing with local SNR criterion. The Ideal Binary Mask is equal to 1 when the SNR is above a threshold value and 0 when the SNR is lower than this threshold value. Once the mask obtained then for getting the processed signal, the ideal binary mask is multiplied with noisy speech. After this all process, we get processed speech signal. Due to this binary masking when every frame is compared and estimated it gives sharper output and fine clean signal [10]. Inverse FFT of processed speech signal is taken to convert it into the time domain signal. At the output of inverse FFT, algorithm gives the processed signal in the form of frames with time domain. For getting continuous output signal, overlap and add method is used here. This method discards the zero padding which had previously added to signal to increase its length. After all this procedure the algorithm gives the continuous and enhanced speech signal.

IV. ALGORITHM FOR SPEECH ENHANCEMENT BY USING IDEAL BINARY MASK

Modules:

- Noisy speech
- Vector to frame conversion
- Fast Fourier Transform
- Ideal Binary Mask
- Inverse Fast Fourier Transform
- Frame to Vector conversion.

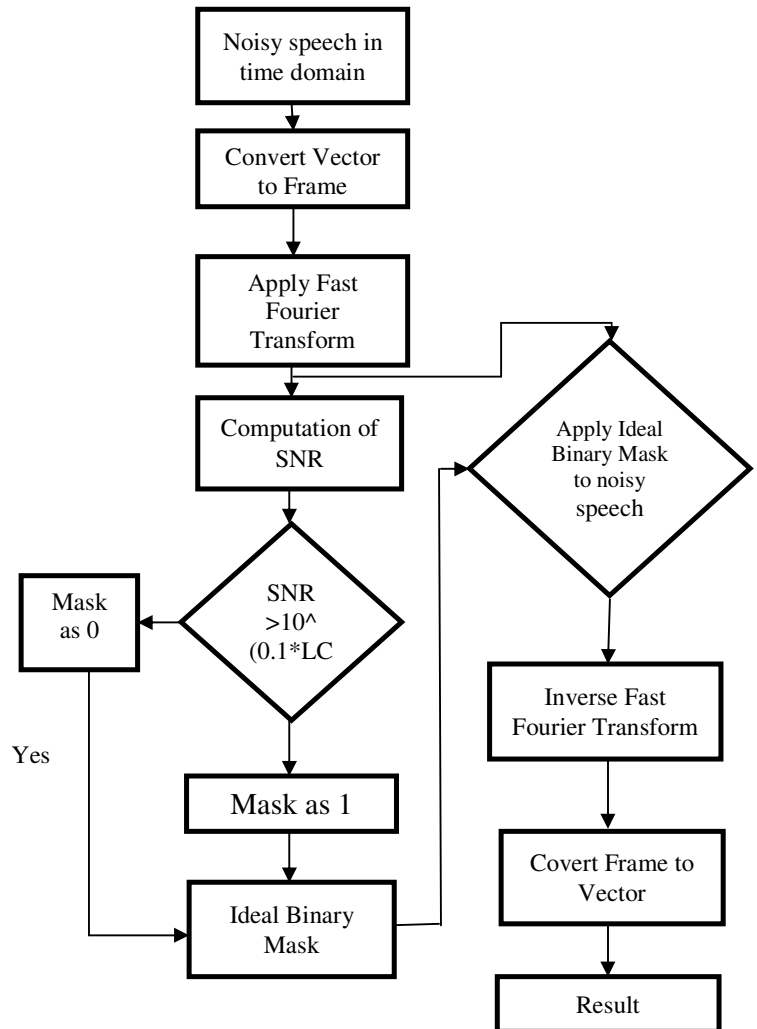


Fig3. Flow Chart of Speech enhancement method using Ideal Binary Mask

1) **Noisy Speech:** This is speech which we are applying as input to the speech enhancement method. Noisy speech is a combination of pure clean speech and background noise. The presence of background noise in speech significantly reduces the quality and intelligibility of speech. Here, we

have taken noisy speech samples at different background noises for testing of speech enhancement method by using Ideal Binary Mask. The background noises includes: multi-talker babble noise, speech in speech noise, car ignition noise, environmental noise, kids cheering voice, helicopter noise, traffic jam noise, random noise, tractor noise, factory noise.

2) Vector to Frame conversion: For assessment of the algorithm, the noisy speech of 8000 Hz is considered. Speech or audio signal contains sound amplitude that varies in time. Vector is a time vector of noisy speech signal in which each element corresponds to the time of each sample. The input signal is a very long data sequence. The noisy speech signal is segmented into small fixed size data frames before further process. Vector to frame conversion splits signal into overlapped frames using indexing and it creates a matrix of input vector. The noisy speech signal is divided into frames of 32ms with frame shift of 4ms. The created matrix consist rows of segments of length 256, taken at every 32 samples along the input vector and windowed using the hanning window.

Speech processing to convert vector into frames is done by taking short windows and processing them. The short window of signal like this is called frame. Here the hanning window is an optional analysis window function to be applied to each frame. Hanning window coefficients,

$$W(n) = 0.5[1 - \cos(2\pi n/N)] \quad , \quad 0 \leq n \leq N$$

Where, the window length is $L = N+1$.

A long signal of speech is multiplied with a hanning window function of finite length and gives finite length version of the original signal. For making the exact division of signal samples into frames, the conversion performs zero padding. Zero padding is a method of adding the zeroes in the signal to increase its length; it will not give any additional information about the signal.

3) Fast Fourier Transform: Fast Fourier Transform is a highly efficient algorithm for transformation of signal from time domain into frequency domain. The speech is first windowed for FFT analysis. FFT process the frames one after

another and processed frames are placed systematically to form output signal.

A fast Fourier transform (FFT) algorithm computes the Fourier transform of a noisy speech signal. FFT converts a signal from its original domain (often time or space) to a frequency domain (spectrum). The functions $Y = \text{fft}(x)$ implements the transform function for vector X of length N by

$$X(k) = \sum_{j=1}^N x(j) W_N^{(j-1)(k-1)}$$

where, $W_N = e^{(-2\pi i)/N}$.

4) Ideal Binary Mask: Ideal binary mask (IBM) identifies speech dominated and noise dominated units, a binary mask is computed and applied to the noisy input spectrum to get the noise-suppressed spectrum. The ideal binary mask is returned in MASK, while the true instantaneous spectral signal-to-noise ratio is returned in SNR. The ideal binary mask is computed from an oracle (true) signal-to-noise ratio (SNR) by thresholding with local SNR criterion specified in LC. The units with local SNR higher than the threshold are defined as target-dominated units, while others as masker-dominated unit. Speech dominated unit labeled as 1 and noise dominated unit labeled as 0. The noisy speech signal and estimated binary mask is combined to produce the enhanced speech signal.

This binary mask method simply keeps the speech dominated (SD) units while zeroing out the noise-dominated (ND) units of noisy speech. The synthesized enhanced speech is returned in Y.A high increase in intelligibility can be obtained by applying the Ideal binary mask to noisy speech. Accurate estimation of the ideal binary mask is important for the algorithms aimed at improving speech intelligibility.

5) Inverse Fast Fourier Transform: An inverse Fourier transform converts the frequency domain components back into the original time domain signal. After this conversion the algorithm provides enhanced output speech signal in time domain.

The function $y = \text{ifft}(X)$ implement inverse transform given for vectors of length N by,

$$x(j) = \left(\frac{1}{N}\right) \sum_{k=1}^N X(k) W_N^{-(j-1)(k-1)}$$

where, $W_N = e^{(-2\pi i)/N}$

$y = \text{ifft}(X)$ returns the inverse discrete Fourier transform (DFT) of vector X, computed with a fast Fourier transform (FFT) algorithm.

6) Frame to Vector conversion: The frame to vector conversion is important to convert frames of signal into vector form. After this process we get continuous enhanced signal at the output. The algorithm use weighted overlap-and-add synthesis method for conversion. It converts frames in matrix into continuous signal using weighted overlap-and-add procedure, specified by synthesis window. Overlap and add method segments the input data into block of length L. If M is length of impulse response then M-1 zeros are added to each block and FFT is computed. Each block has input data of length L with zero padding of length M-1. Block is always terminated with M-1 zeros, the last M-1 points from each output block must be overlapped and added to first M-1 points of next block. Hence this method is called as overlap-add method. At the end of conversion, overlap and add method combines the entire blocks to get the continuous output signal.

V. RESULTS

The different samples of clean speech and noises are collected with 8000 HZ frequency. The different types of noise are used to generate noisy speech data in experiment. These noises: Babble noise, Car Ignition Noise, Helicopter noise, Natural Noise, Kids Cheering Voice, Transportation Noise, Random noise, Tractor Noise. The experiment for speech enhancement is performed on different noisy speech to analyse how well a speech enhancement algorithm works with different types of noise. The main goal of speech enhancement is to eliminate the additive noise present in speech signal and restore the speech signal to its original form. Therefore, we get enhanced speech as output after applying the ideal binary mask method of speech enhancement. Performance is assessed on the basis of speech quality and intelligibility. SNR of clean, noisy and enhanced speech is measured to

determine the quality of processed speech. Intelligibility is difficult to measure by any mathematical algorithm. Therefore listening experiment is conducted to measure speech intelligibility. Some Results of speech enhancement using Ideal Binary Mask are as shown in below. The different samples of signals are used for testing the results of experiments.

The nature of clean speech, noisy speech corrupted by babble noise and processed speech is as shown in below. In figures below first signal is the clean speech, second is noisy speech because of adding the noise in the clean speech and last is the estimation in the form of processed signal. The spectrograms of clean, noisy and processed signals are in 2nd column.

A. Results for Babble noise

Babble noise creates serious problem to recognize a particular clean speech. Babble noise will be overlapping on the speech conversation and reduces intelligibility.

Fig 4 and 5 shows the results for different clean speech which are affected by babble noise and their SNRs for different signals.

1. Results for 1st clean speech with babble noise

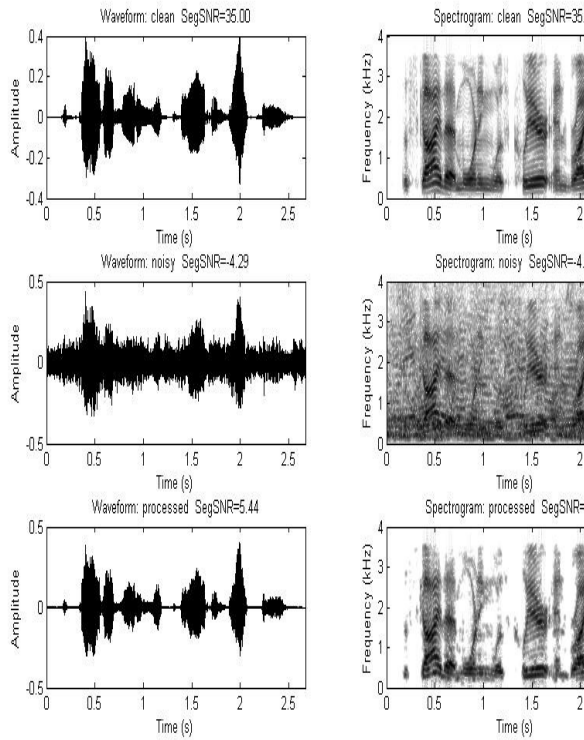


Fig4. Results for 1st clean speech with babble noise

2. Results for 2nd clean speech with babble noise

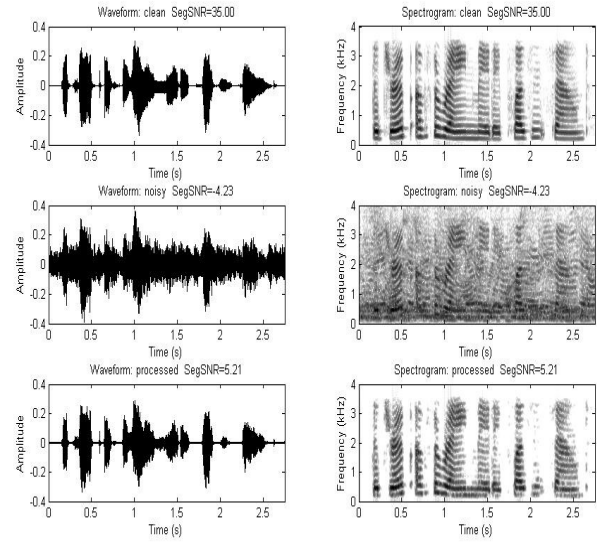


Fig5. Results for 2nd clean speech with babble noise

B. Results for Helicopter noise

Fig 6 and 7 shows the results for different clean speech which are affected by helicopter noise and SNRs for different signals.

1. Results for 1 clean speech with Helicopter noise

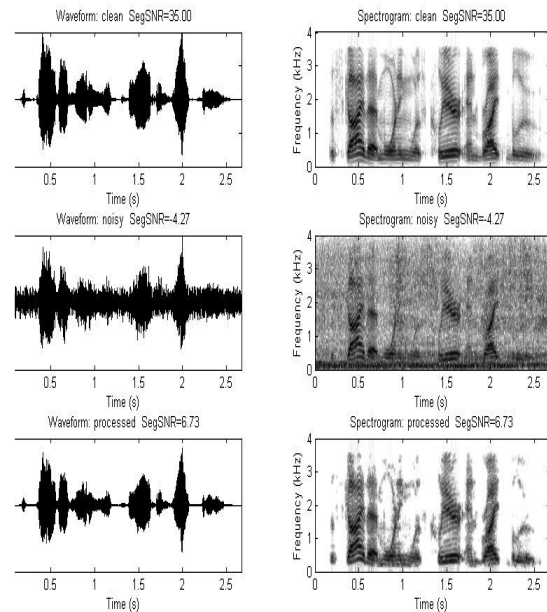


Fig6. Results for 1st clean speech with Helicopter Noise
2. Results for 2nd clean speech with Helicopter noise

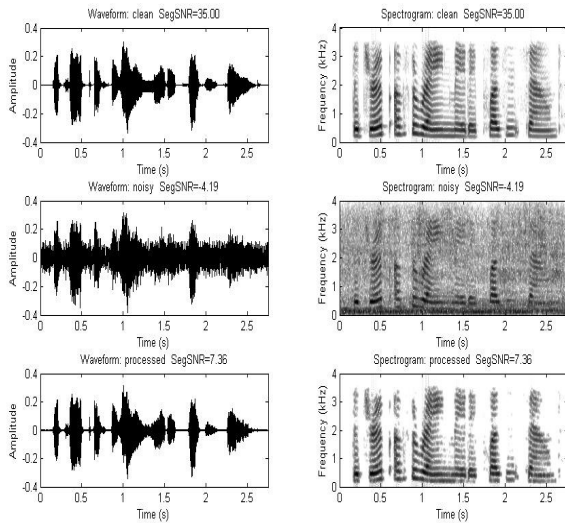


Fig7. Results for 2nd clean speech with Helicopter noise

C. Car ignition noise

This is also one of the types of noise which degrades the speech quality and intelligibility. Here, car ignition noise has taken for experiment test. Fig 8 and 9 shows the results for different clean speech which are affected by car ignition noise and SNRs for different signals.

1. Results for 1st Clean speech with Car ignition noise –

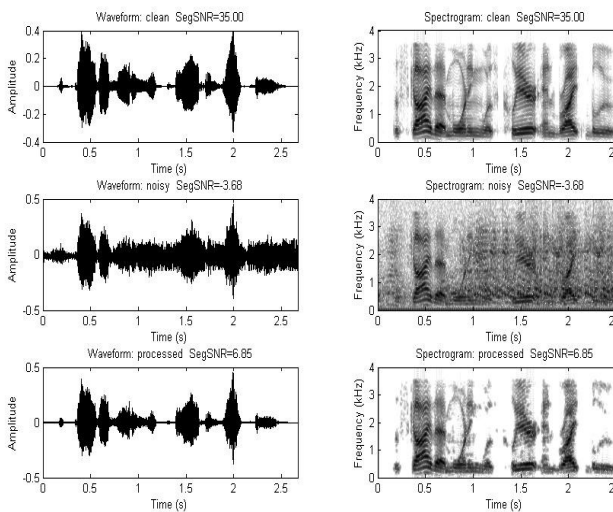


Fig8. Results for 1st Clean speech with Car ignition noise

2. Results for 2nd clean speech with Car ignition noise

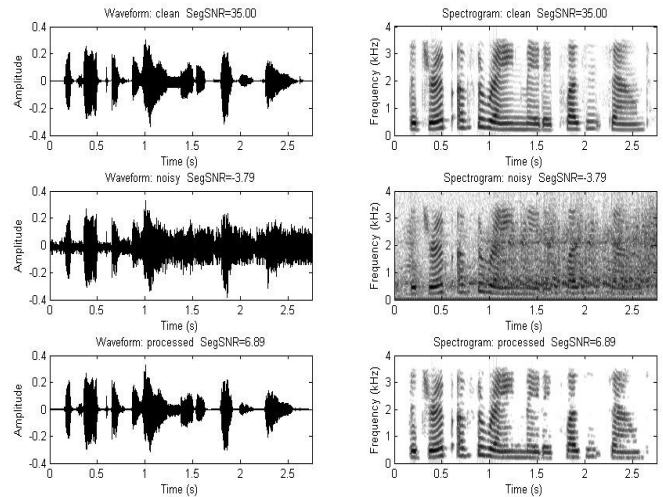


Fig9. Results for 2nd clean speech with Car ignition noise

D. Natural noise

Fig 10 and 11 Show the results for different clean speech which are affected by Natural noise and SNRs for different signals.

1. Results for 1st clean speech with Natural noise

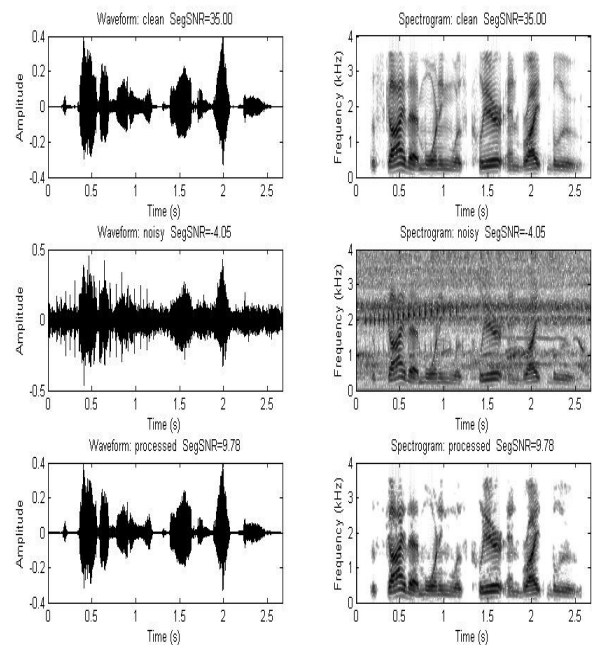


Fig10. Results for 1st clean speech with Natural noise

2. Results for 2nd clean speech with Natural noise

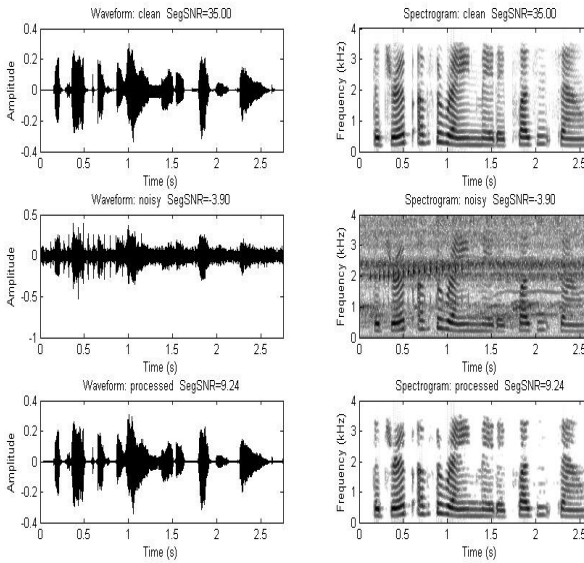


Fig11. Results for 2nd clean speech with Natural noise

E. Kids cheering voice

Fig 12 and 13 show the results for different clean speech which are affected by kids cheering voice and SNRs for different signals.

1. Results for 1st clean speech with Kids cheering voice

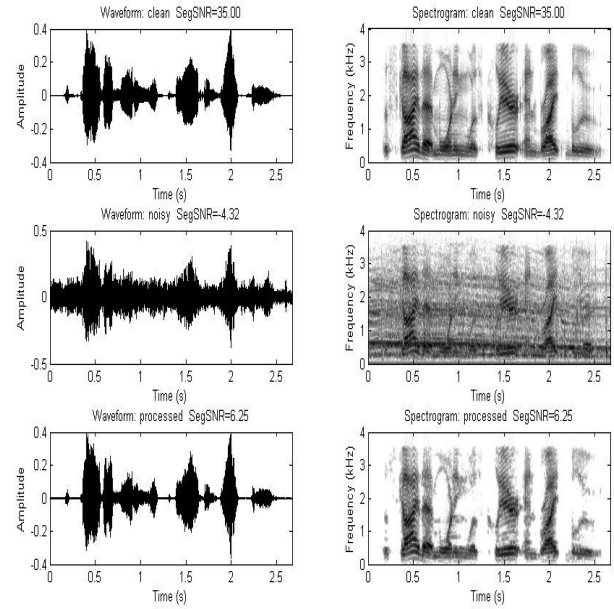


Fig12. Results for 1st clean speech with Kids cheering voice

2. Results for 2nd clean speech with Kids cheering voice

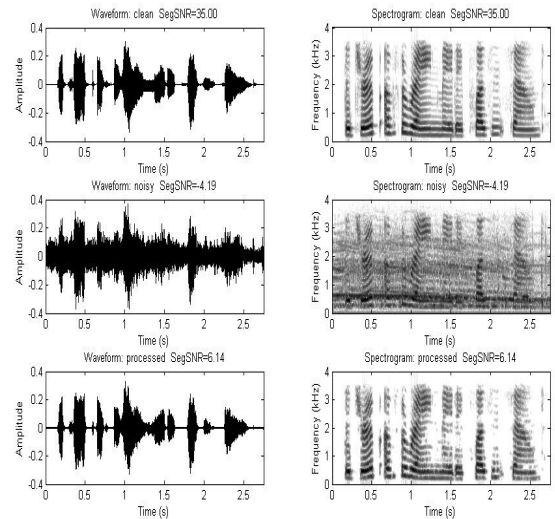


Fig13. Results for 2nd clean speech with Kids cheering voice

F. Transportation Noise

Generally road traffic vehicles are most important type of noise in the social life. Transportation noise relates to noise generated by road vehicles, aircrafts, trains, vessels. This transportation noise affects the communication

signals. Transportation noise leads to degradation of the quality and intelligibility of speech signal. Fig 14 and 15 Show the results for different clean speech which are affected by transportation noise and SNRs for different signals.

1. Results for 1st clean speech with Transportation Noise

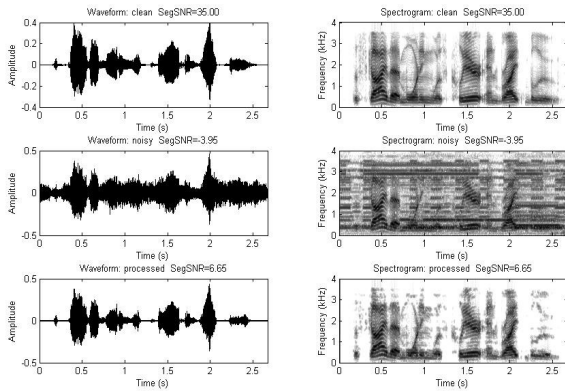


Fig14. Results for 1st clean speech with Transportation Noise

2. Results for 2nd clean speech with Transportation Noise

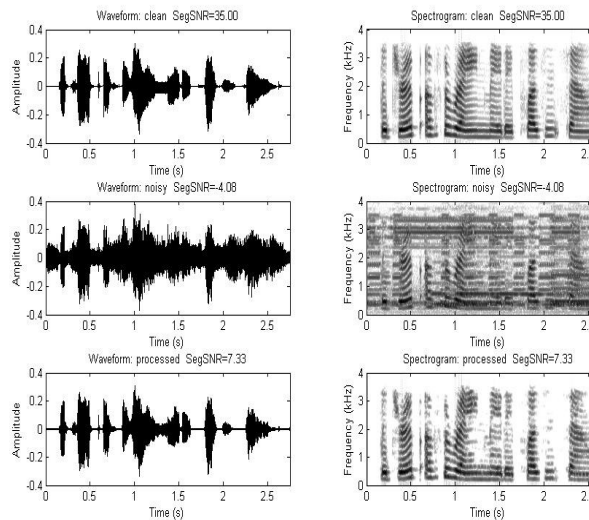


Fig15. Results for 2nd clean speech with Transportation Noise

G. Random noise

This noise affects the speech signal when add in it. It reduces quality and intelligibility of speech signal and become hard to understand the particular speech signal. Fig 16 and 17 Show the results for different clean speech which are affected by Random noise and SNRs for different signals.

1. Results for 1st clean speech with Random noise

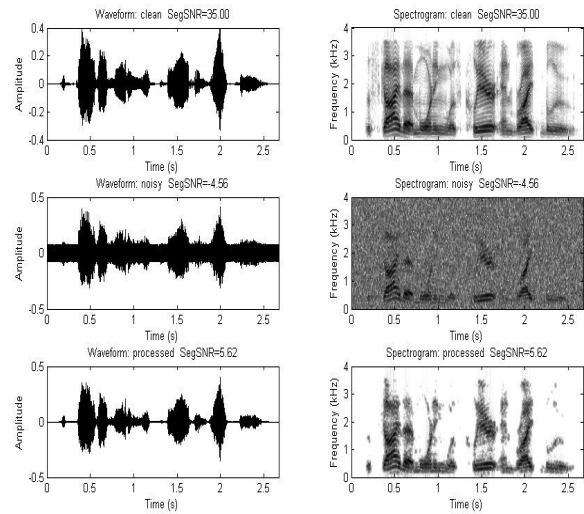


Fig16. Results for 1st clean speech with Random noise

2. Results for 2nd clean speech with Random noise

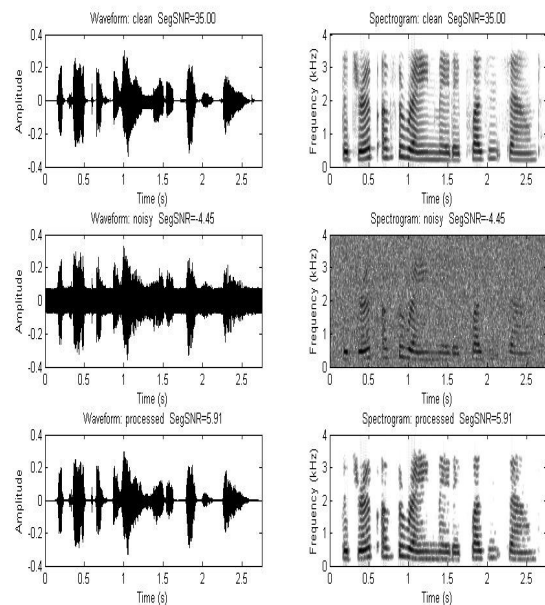


Fig17. Results for 2nd clean speech with Random noise

H. Tractor noise

The tractor creates a high level noise when it mixes with a speech. It reduces quality as

well as intelligibility of speech. Fig 18 and 19 Show the results for different clean speech which are affected by tractor noise and SNRs for different signals.

in analysis of algorithm. On the basis of these SNRs, we can compare the quality of clean speech and speech after processing.

1. Results for 1st clean speech with tractor noise

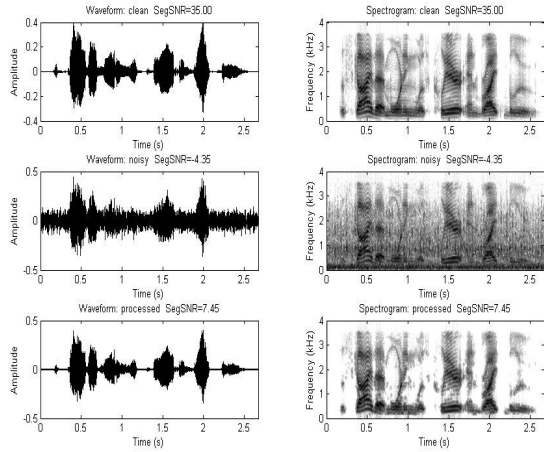


Fig18. Results for 1st clean speech with tractor noise

5.1.8.2 Results for 2nd clean speech with tractor noise

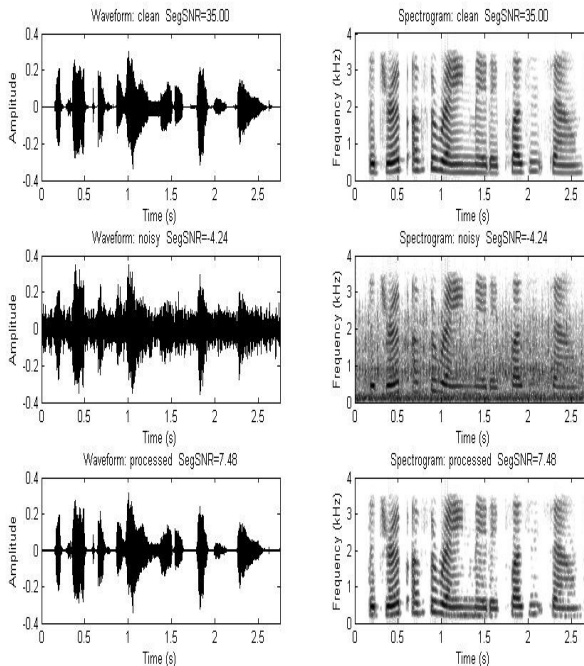


Fig19. Results for 2nd clean speech with tractor noise

Table1.Shows the SNRs for different clean speech, noisy and processed speech which are used

Table1. SNRs for different signals

VI. CONCLUSION

Speech enhancement method by using Ideal Binary Mask is performed on different speech to improve the quality as well as intelligibility of speech. Performance is assessed on the basis of speech quality and intelligibility. SNR of clean, noisy and enhanced speech is measured to determine the quality of processed speech. Intelligibility is difficult to measure by any mathematical algorithm. Therefore listening experiment is conducted to measure speech intelligibility.

On the basis of this analysis of SNR and listening test:

1. The proposed method is able to eliminate the additive noise present in speech signal and restore the speech signal to its original form.
2. Speech enhancement method in proposed method attenuates the additive background noise without introducing speech distortion.
3. The proposed speech enhancement method is suitable to increase the recognition

accuracy of speaker under real life noise

Sr. No.	Type of clean speech	Type of noise	SNR(dB)		
			Clean Speech	Noisy speech	Processed speech
1	1 st Clean Speech	Babble Noise	35	4.29	5.44
		Helicopter noise	35	4.27	6.73
		Car ignition noise	35	3.68	6.86
		Natural noise	35	4.05	9.78
		Kids cheering voice	35	4.32	6.25
		Transportation Noise	35	3.95	6.66
		Random noise	35	4.56	5.62
		Speech in speech noise	35	4.50	4.38
		Tractor noise	35	4.35	7.45
2.	2 nd Clean Speech	Babble Noise	35	4.23	5.21
		Helicopter noise	35	4.19	7.36
		Car ignition noise	35	3.79	6.89
		Natural noise	35	3.90	9.24
		Kids cheering voice	35	4.19	6.14
		Transportation Noise	35	4.08	7.33
		Random noise	35	4.45	5.91
		Speech in speech noise	35	4.43	4.44
		Tractor noise	35	4.24	7.48

conditions.

- The proposed method is able to enhance the speech quality as well as intelligibility by using Ideal Binary Mask in different noisy environments.

REFERENCES

- Raphael Koning, NileshMadhu, 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.
- NimaYousefian and Philipos C. Loizou, "A Dual-Microphone Speech Enhancement Algorithm Based on the Coherence Function", *IEEE transactions on audio, speech, and language processing*, vol. 20, no. 2, February 2012.
- DeLiang Land Guy J. Brown, "Separation of Speech from Interfering Sounds Based on Oscillatory Correlation", *IEEE transactions on neural networks*, vol. 10, no. 3, May 1999.
- Dionysis E. Tsoukalas, John N. Mourjopoulos and George Kokkinakis, "Speech Enhancement Based on Audible Noise Suppression", *IEEE transactions on speech and audio processing*, vol. 5, no. 6, November 1997.
- Steven F. Ball, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", *IEEE transactions on acoustics, speech, and signal processing*, vol. assp-27, no. 2, April 1979.
- Philipos C. Loizou 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.
- 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
- Arun Narayanan and DeLiang Wang, "A CASA-Based System for Long-Term SNR Estimation", *IEEE transactions on audio, speech, and language processing*, vol. 20, no. 9, November 2012.

9. Bastian Sauert and Peter Vary, "Near End Listening Enhancement: Speech Intelligibility Improvement in Noisy Environments", *Proceeding of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 493–496, May 2006.
10. Dr. Madhuri. A. Joshi and Priyanka S. Vispute, " Factors Affecting on Improvement of Intelligibility in Speech Enhancement", *3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS)-2015*.
11. Gibak Kim and Philipos C. Loizou, "Improving Speech Intelligibility in Noise Using Environment-Optimized Algorithms", *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 8, November 2010.
12. D. L.Wang and G. J. Brown, Eds. Hoboken, NJ:Wiley,"*Computational Auditory Scene Analysis: Principles, Algorithms and Applications*", IEEE Press, 2006.