Frequency Dependent Noise Flooring Parameter (FDNFP) for Speech Enhancement Using Linear Prediction

Dr.Prof.Mr.D.B.Kadam[1] Prof.Mr.V.S.Suryawanshi[2] Mayuri Ghuli[3] Dipanjali Hate[4] Payal More[5]
[1]HOD [2]Assistant professor [3,4,5]U.G. Scholar
[1,2,3,4,5]Department of Electronics and Tele Communication Engineering
[1,2,3,4,5]P.V.P.Institute of Technology Budhgaon.

Abstract— Speech enhancement aims to improve speech quality by using various algorithms. It may sound simple, but at least it should have the terms clarity and intelligibility, pleasantness, or compatibility with some other method in speech processing. An enhancement method for single-channel speech degraded by additive noise is proposed. A spectral weighting functions derived by constrained optimization to suppress noise in the frequency domain. Two design parameters are included in the suppression gain, namely, the frequency-dependent noise flooring parameter (FDNFP) and the gain factor. The FDNFP controls the level of admissible residual noise in the enhanced speech. Enhanced harmonic structures are incorporated into the FDNFP by time-domain processing of the linear prediction residuals of voiced speech. Further enhancement of the harmonics is achieved by adaptive comb filtering derived using the gain factor with a peak-picking algorithm. Moto of this paper is to reduce the background noise added in the original speech signal so that listener will able to listen the noiseless speech along with clear voice. Here, we have implemented this paper for previously recorded historical speech signal. Advantage of this paper is that we can get clear voice from the noisy one and can store the speech for long time without destroying.
Key words: Speech Enhancement, FDNFP

I. INTRODUCTION

Speech enhancement involves processing of speech signals for human listening or as preparation for further processing before listening. The enhancement process aims to improve the quality of the overall speech; to increase the speech intelligibility in order to reduce the listener fatigue. Speech enhancement is closely related to speech restoration. The enhancement of single-channel speech degraded by additive noise has been extensively studied in the past and remains a challenging problem because only the noisy speech is available. Techniques have been proposed to exploit the harmonic structure of voiced speech for enhancing the speech quality. Here we propose a new method that enhances the harmonics of voiced speech without ascribing to any underlying speech models.

The harmonic speech structure obtained through short-time Fourier analysis is enhanced by applying a combination of time and frequency domain-based criteria, which are applicable for white as well as for colored additive noise conditions. Two design parameters are introduced into the proposed suppression gain, namely the frequency-dependent noise-flooring parameter (FDNFP) and the gain factor. The FDNFP shapes the residual noise in the frequency domain such that the harmonic structure of clean speech is preserved. To further enhance the harmonics of voiced speech, adaptive comb filtering is performed.

II. LITERATURE REVIEW

Several studies are reported in the area of speech enhancement, this chapter includes the literature reviewed different methods for speech enhancement technology.

A. J. Jensen And J. H. L. Hansen, "Speech Enhancement Using A Constrained Iterative Sinusoidal Model," Ieee Trans. Speech Audio Process., Vol. 9, No. 7, Pp. 731–740, Oct. 2001.:

This paper presents a sinusoidal model based algorithm for enhancement of speech degraded by additive broad-band noise. In order to ensure speech-like characteristics observed in clean speech, smoothness constraints are imposed on the model parameters using a spectral envelope surface (SES) smoothing procedure. The aim of the proposed enhancement scheme is to improve the quality of the enhanced speech signal, by exploiting this knowledge of the signals origin to apply speech production constraints in the enhancement process. A measureable improvement in objective speech quality was observed compared to more traditional speech enhancement methods such as spectral subtraction.

B. Y. Hu and P. C. Loizou, "Incorporating a psycho acoustical model in frequency domain speech enhancement," IEEE Signal Process. Lett., vol. 11, no. 2, pp. 270–273, Feb. 2004.:

This paper presents various ideas for removing noise from audio signals requires a non-diagonal processing of time-frequency coefficients to avoid producing "musical noise". Audio signals are often contaminated by background environment noise and buzzing or humming noise from audio equipments. Audio denoising aims at attenuating the noise while retaining the underlying signals. They have concentrated on the coefficient processing as opposed to the choice of representations. Numerical experiments are performed with short-time Fourier transforms that are most commonly used in audio processing. After his performance, he concludes that non-diagonal time-frequency estimators are more effective than diagonal estimators to remove noise from audio signals because they introduce less musical noise.

C. S. Gustafsson, P. Jax, and P. Vary, "A novel psycho acoustically motivated audio enhancement algorithm preserving background noise characteristics," in Proc. IEEE Int. Conf. Acoustic., Speech, Signal Process.(ICASSP), 1998, pp. 397–400.:

They have proposed an algorithm for reduction of noise in audio signals. In contrast to several previous approaches he do not try to achieve a complete removal of the noise, but instead his goal is to preserve a pre-defined amount of the original noise in the processed signal. This is accomplished by exploiting the masking properties of the human auditory system. Simulation results confirm that no audible artifacts are left in the processed signal, while speech distortions are comparable to those caused by conventional noise reduction techniques.

D. C. Li and S. V. Anderson, "Inter-frequency dependency in MMSE speech enhancement," in Proc. 6th Nordic Signal Process. Symp.,2004, pp. 200–203.:

From this paper the comprehensive Linear Minimum Mean Squared Error (LMMSE) approach for parametric speech enhancement is developed. The proposed algorithms aim at joint LMMSE estimation of signal power spectra and phase spectra, as well as exploitation of correlation between spectral components.

III. PROPOSED METHOD

In this method the information about speech enhancement process is proposed. Actually for this purpose different methods are available. Here the enhancement is performed by using Frequency Dependent Noise Flooring Parameter. (FDNFP). For this FDNFP process we have to perform LP analysis, Calculation of pitch period, Windowing & Averaging. LPC is generally used for speech analysis and resynthesis. It is used as a form of voice compression by phone companies, for example in the GSM standard. It is also used for secure wireless, where voice must be digitized, encrypted and sent over a narrow voice channel.

IV. SPEECH ENHANCEMENT

In this section we are going to have the introduction about speech enhancement which includes- what actually speech enhancement is? , why it is so required? Actually speech enhancement improves the quality of speech signal. In speech communication, the speech signal is always accompanied by some noise. In most cases the background noise of the environment where the source of speech lies, is the main component of noise that adds to the speech signal. Though the obvious effect of this noise addition is to make the listening task difficult for a direct listener, there are many more far reaching negative effects when we process the degraded speech for some other applications. A related problem is processing degraded speech in preparation for coding by a bandwidth compression system. Hence speech enhancement not only involves processing speech signals for human listening but also for further processing prior to listening. Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility, or degree of listener fatigue. In our work, we study one such enhancement technique to enhance the quality of speech in the presence of additive, broadband acoustic noise.

The interfering noise generally degrades the quality and intelligibility of speech. While the term intelligibility refers to the recognisability of actual content of speech, quality refers to the aspect of the speech that determines the ease with which one can understand the 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. Hands free voice control is one such application. The different contexts in which the enhancement techniques are used are as follows: -

1) Enhancement of speech degraded by additive noise
2) 2.Making speech immune to degradation by processing speech before noise affects it.
3) Enhancement of speech degraded by reverberation.

V. FUTURE SCOPE

There is always a scope of improvement in any design and implementation of the speech enhancement techniques, supported by the advancements in the tool, technology, and thinking. A few of the suggested future plans or work extensions are listed below:

– Real time implementations:

In this implementation of speech enhancing method, on real time application like simultaneously processing of noise removal while communication is going on and removing noise from audio database like speech, music.

This method can be useful in applications like auditorium hall where we can eliminate the background noise when some interference is occurring such as ringing of mobile phone, person coming late in hall saying may I come in. By that time this system becomes useful to clear out the noise and we can get the pure speech of speaker.

FDNFP Block Diagram

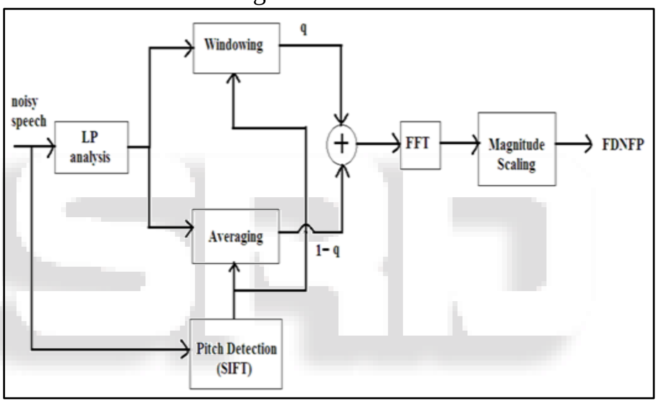

Fig. 1: Computation of the FDNFP using LP analysis

In many real world applications the speech enhancement is necessary for getting clear speech signal instead of speech signal with added background noise. Enhancing of speech degraded by noise, or noise reduction, is the most important field of speech enhancement, and used for many applications such as mobile phones, VOIP, teleconferencing systems, speech recognition, and hearing aids. The proposed method improves speech quality by suppressing the noise in the frequency domain with the use of a spectral weighting function. Because, voiced speech is quasi-periodic in nature, its magnitude spectrum exhibits peaks and valleys separated by harmonics of the fundamental frequency. The harmonic structure of clean voiced speech is often corrupted by the additive noise spectrum. Many classical noise reduction methods that use multiplicative spectral gains, e.g. short-time spectral amplitude modification estimations, fail to recover the harmonic structure because they do not take advantage of the properties of this redundancy in voiced speech. In order to impose a harmonic envelope upon the FDNFP, we propose here an approach to extract the harmonic structure of voiced speech in the time domain.

The motivation for time domain processing is to preserve the correlation between both spectral amplitudes and phases when restoring the harmonics. Because the phase coherence in voiced speech is a significant source of correlation and corresponds to energy localization in the time domain, we retrieve the harmonic information from noisy speech by enhancing the excitation peaks in the linear prediction residuals. For voiced speech, a linear prediction (LP) analysis is performed on the noisy speech. In our implementation, the classical autocorrelation method is used to derive the LP parameters. The model order is set to 15. The LP residual signal is processed in parallel by two different methods to enhance the excitation peaks.

The first method attenuates the signal amplitudes between excitation peaks by windowing the LP residual signal with a Kaiser Window series. The duration of each window is set to be equal to the pitch period. The centers (peaks) of the windows are aligned in time with the peaks of excitation pulses. The purpose of windowing is to enhance the amplitude contrast between peaks and valleys of the excitation pulses. The motivation for this averaging is based on the fact that while the LP bursts of voiced speech are quasi-periodic, the additive noise tends to be random and uncorrelated. By averaging the LP residuals over several pitch periods, the periodic components will therefore be enhanced while the uncorrelated random components will be suppressed.

In the second method, the LP residuals are averaged over the pitch period.

$$\mu_a(n)=\frac{1}{M}\sum_{i=0}^{M-1}\mu(n+iP) \qquad \text{where, } n=0, 1, 2\ldots\ldots\ldots P-1.$$

Where, μ_a (n)andμ (n) are the averaged and noisy LP residuals, respectively. M is the largest integer number of pitch periods in the current analysis frame. P is the number of samples in one pitch period. n Is the time sample index, and i is the pitch epoch index. The duration μ_a (n),of the averaged LP residual, is exactly one pitch period. Then, 〖μ〗_a (n), is repeated during the whole analysis frame.

The motivation for this averaging is based on the fact that while the LP bursts of voiced speech are quasi-periodic, the additive noise tends to be random and uncorrelated. By averaging the LP residuals over several pitch periods, the periodic components will therefore be enhanced while the uncorrelated random components will be suppressed.

In order to provide the necessary pitch information for the aforementioned windowing and averaging process, a pitch detection algorithm is run in parallel to determine the pitch period of the current frame. Here we use the relatively simple SIFT (Simple Inverse Filter Tracking) method [14] for pitch determination. The final processed LP residual with enhanced periodicity is obtained by,

$$u_h(n)= qu_w(n)+ (1\text{-}q)\, u_e(n)$$

where, n=0, 1, 2………L-1.

Where q is a smoothing factor, and u_w (n) is the window-enhanced LP residuals. u_e (n) Is obtained by periodically extending u_a (n)in over the entire duration of the analysis frame. u_h (n)Is the final LP residual with enhanced periodicity. Because the averaging-enhanced residuals may not be as accurate as windowing-enhanced residuals, due to shimmer for example, the parameter is set to 0.8. u_h (n) is then transformed to the frequency domain, and its magnitude spectrum is normalized to 0 dB by its maximum magnitude.

Finally, the FDNFP is obtained by scaling the normalized spectrum of u_h (n) to some comfort noise level. In our implementation, the normalized spectrum of u_h (n) is scaled down by 5 dB for strongly voiced speech demonstrate the process of obtaining the FDNFP varying system. It has become the predominant technique for estimating the basic speech parameters such as pitch, formants, and spectra. The importance of linear prediction lies in its ability to provide extremely accurate estimates of speech parameters as well as its relative speed of computation.

Pitch Detection:

Pitch detection is very important for many speeches processing algorithm. Speech recognition system of tonal language use pitches tracking for tone recognition, which is important in disambiguating the myriad of homophones. Pitch is also crucial for prosodic variations in text-to-speech systems and spoken language systems. The fundamental frequency (F0) is the main cue of the pitch. However, it is difficult to build a reliable statistical models involving fundamental frequency F0 because of pitch estimation errors and the discontinuity of the F0 space. Thus, a reliable pitch detection algorithm (PDA) is a very important component in many speech processing systems. For pitch detection Simplified Inverse Filtering Technique (SIFT) algorithm is used.

Windowing & Averaging:

After being portioned into frames each frame is multiplied by a window function prior to the spectral analysis to reduce the effect of discontinuity introduced by the before process by attenuating the values of the samples at the beginning and at the end of each frame, commonly used window is called Hamming. The purpose of windowing is to enhance the amplitude contrast between peaks and valleys of the excitation pulses.

Windowing: multiply the full waveform s (n) by a window w (n) (in time domain):

$$X[n] = w[n]s[n]$$

Simply cutting out a short segment (frame) from s (n) is a rectangular window — causes discontinuities at the edges of the segment Instead; a tapered window is usually used

E.g. Hamming (_ = 0.46164) or Hanning (_ = 0.5) window.

By averaging the LP residuals over several pitch periods, the periodic components will therefore be enhanced while the uncorrelated random components will be suppressed. Averaging is a commonly used procedure for estimating the noise power spectrum.

Spectral Subtraction:

Spectral subtraction is historically one of the oldest simple algorithm to implement easily and a minimal complexity of a speech enhancement. The spectral subtraction is based on the theory that the enhanced speech can be acquire by subtracting the estimated spectral factors from the continuum of the input noisy signal.

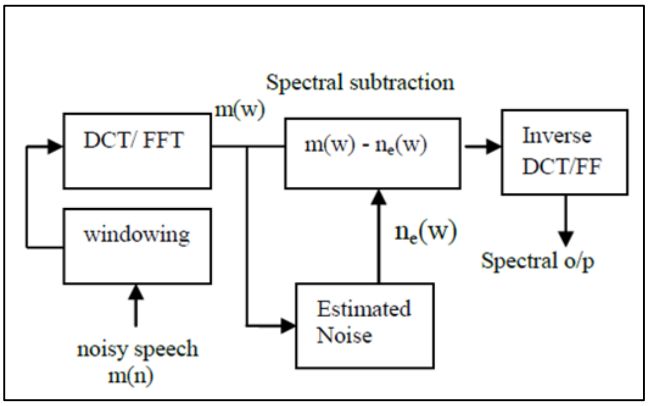

Fig. 2: Block diagram of Spectral subtraction

Software Design:

Flowchart

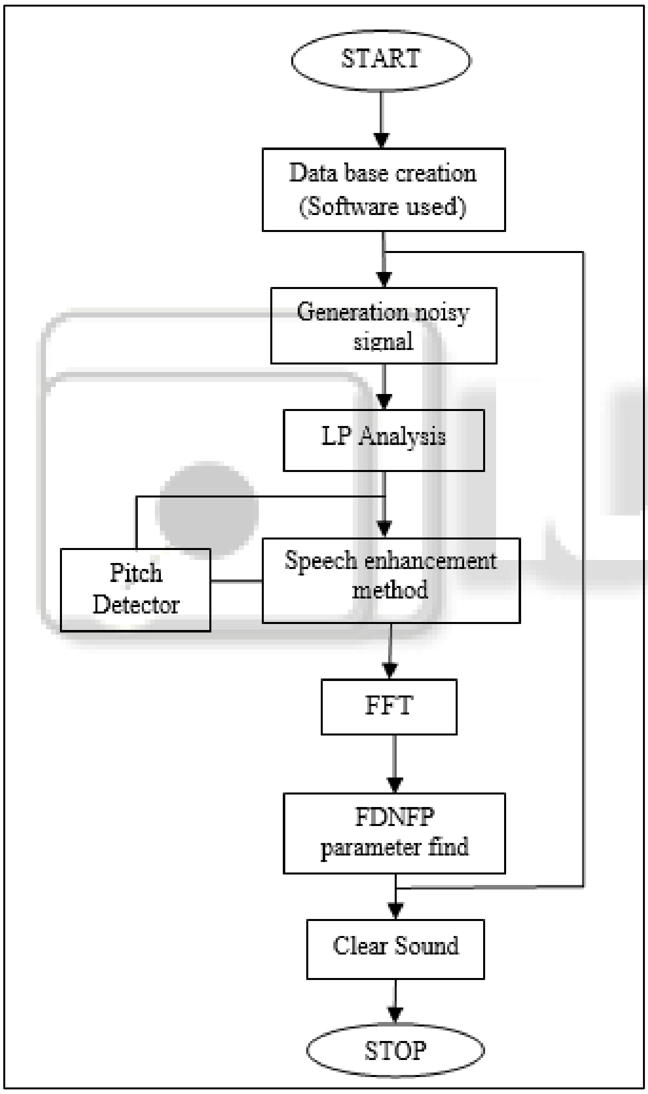

Fig. 3:

Database Creation:

The very first step in our project is database. Wave type database is an annotated and time aligned continuous speech database of English language. It includes recordings of 10 Historical speeches (male and female). Thus the database captures phonetic, acoustic, intra-speaker and inter-speaker variations in English speech. In this database creation we are using 2 software's which are include MATLAB, COOL EDIT.

In MATLAB software MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. In COOL EDIT, cool edit pro is an audio recording and editing software program. Standard features include cutting, pasting, cropping and combining audio files in order to create that exact song or special effect that was stuck in your head all day long. Cool edit tool is also used while creating a database. This tool provides various facilities, using this tool we have to amplify the weak recorded signal and this tool is also used for data of multiple speakers will be mixed to form the data corresponding to meeting discussion.

Generation of Noisy Signal:

We have taken here input signal having Fs=16000 Hz form WAVE type database.

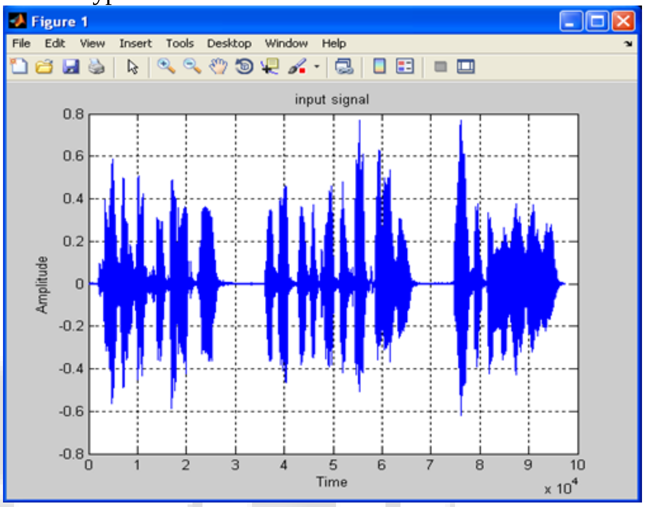

Fig. 4: Input Signal

Now By using above input we can obtain the noisy signal. For these two types of noise is used 1) White Gaussian Noise 2) Multitalker babble noise. These noises are added into the above input signal. Following figure shows the noisy signal from this We can analyze the difference between input signal and noisy signal.

Interference Of Noisy Speech Signal:

Additive white Gaussian noise (AWGN) is a basic noise model used in Information theory. This noise is added in the input signal at SNR value 15.The modifiers denote specific characteristics:

- Additive because it is added to any noise that might be intrinsic to the information system.
- White refers to the idea that it has uniform power across the frequency band for the information system. It is an analogy to the color white which has uniform emissions at all frequencies in the visible spectrum.

Gaussian because it has a normal distribution in the time domain with an average time domain value of zero.

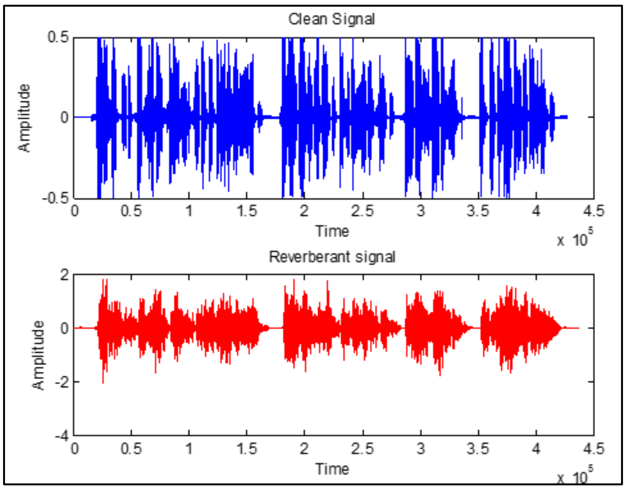

Fig. 5: Noisy Speech Signal of Human (White Gaussian Noise)

Fdnfp Parameter:

Now finally FDNFP parameter that is noise reduced signal is obtained by performing FFT by overlap-add method for both noisy signals. Now from following figure we observe the similarities between the input signal and obtained noise reduced signal.

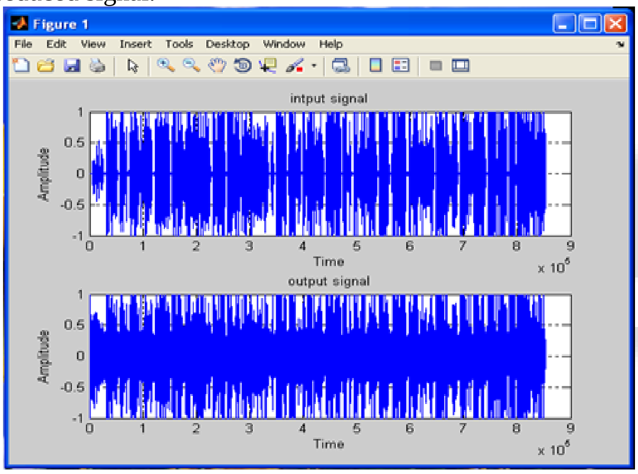

Fig. 6: Output Signal for White Gaussian Noise

VI. Conclusion

In this thesis, we have proposed a speech enhancement method which aims at emphasizing harmonics. The harmonics are enhanced by processing the degraded speech in both the time and frequency domains. In contrast to many other state-of-the-art methods, the proposed algorithm allows a low level of residual noise in the enhanced speech. The noisy speech is enhanced in the frequency domain by a spectral weighting function, which contains two design parameters. One of the design parameters, namely, the frequency-dependent noise-flooring parameter (FDNFP), is used to emphasize the harmonics of voiced speech as well as to control the frequency-dependent level of admissible residual noise. For voiced speech, the periodicity in the linear prediction residual signal was detected and enhanced and then transformed to the frequency domain to be used as the FDNFP. The magnitudes of the FDNFP are scaled to some small values in order to suppress the level of residual noise in the enhanced speech. The other design parameter is the dominant term in the spectral gain function.

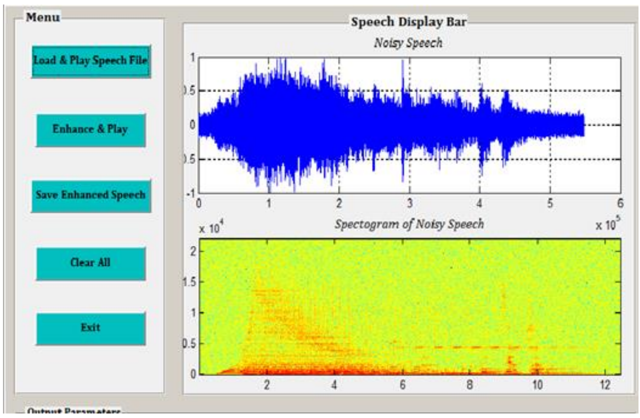

Fig. 7: Noisy Speech Signal

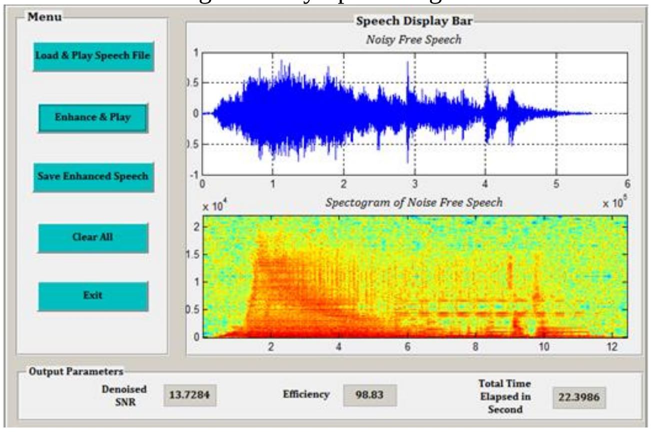

Fig. 8: Clear Speech Signal

VII. Result

We have surveyed some aspects of the speech enhancement problem and present state of the art solutions to this problem. In particular, we have identified the difficulties inherent to speech enhancement so we have mainly focused on speech signals degraded by additive noise. Previously, Speech technology had many problems that means speech signal was not clear to any anyone .So, we have overcome that problem by applying the technology of speech enhancement. Now we are able to get the clear voice from distracted speech signal.

References

[1] J. Hardwick, C. D. Yoo, and J. S. Lim, “Speech enhancement using the dual excitation model,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 1993, pp. 367–370.

[2] S. Dubost and O. Cappe, “Enhancement of speech based on non-parametric estimation of a time varying harmonic representation,” in Proc.IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 2000, pp.1859–1862.

[3] M. E. Deisher and A. S. Spanias, “HMM-based speech enhancement using harmonic modeling,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), 1997, pp.

[4] J. Jensen and J. H. L. Hansen, “Speech enhancement using a constrained iterative sinusoidal model,” IEEE Trans. Speech Audio Process., vol. 9, no. 7, pp. 731–740, Oct. 2001.

[5] Y. Hu and P. C. Loizou, “Incorporating a psycho acoustical model in frequency domain speech

enhancement," IEEE Signal Process. Lett., vol. 11, no. 2, pp. 270–273, Feb. 2004.
[6] S. Gustafsson, P. Jax, and P. Vary, "A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics,"in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.(ICASSP), 1998, pp. 397–400.
[7] C. Li and S. V. Anderson, "Inter-frequency dependency in MMSE speechenhancement," in Proc. 6th Nordic Signal Process. Symp.,2004, pp. 200–203.
[8] T. F. Quatieri and R. G. Danisewicz, "Cochannel speaker separation by harmonic enhancement and suppression," IEEE Trans. Acoust., Speech, Signal Process., vol. 38, no. 1, pp. 56–69, Jan. 1990.