A Review of Speech Signal Enhancement Techniques

Devyani S. Kulkarni
Department of Computer Science & IT,
Dr. B.A.M.U.
Aurangabad

Ratnadeep R. Deshmukh
Department of Computer Science & IT,
Dr. B.A.M.U.
Aurangabad

Pukhraj P. Shrishrimal
Department of Computer Science & IT,
Dr. B.A.M.U.
Aurangabad

ABSTRACT
Speech is the most natural and the most effective way of communication between human. During the speech communication, the signals contains some noise so when processing the digital speech signals; speech signal enhancement is very important step. The field speech processing is an applied area of signal processing. The motive of speech enhancement is to enhance the understandability and comprehensibility of speech signal. There are numbers of techniques proposed using which speech signal enhancement is performed. The objective of this paper is to provide an overview of speech enhancement algorithms which are used for enhancement of speech signal.

General Terms
Speech Signal Enhancement, Background Noise, Speech Signal Degradation.

Keywords
Speech Signal Enhancement, Speech Degradation, Speech Communication, Filtering Techniques, Speech Signal processing.

1. INTRODUCTION
In human being the interaction is using vocal communication i.e. voice. This is the motive for the researchers to carry out research in the domain of Digital Speech Signal Processing. The field Digital speech processing is a sub domain of Digital signal processing. Each signal associated with a speech communication always contains a noise. The purpose behind speech enhancement is to enhance the understandability and comprehensibility of speech signal [1]. For achieving a good performance of Speech enabled system it is necessary to have speech signals without noise, high quality and clarity. Every time it is very difficult to have speech signals without any background noise [2]. In a natural environment there is always some amount of ECHO. Acoustically echo less room are generally used for capturing the Echoless Speech [3]. During a study it was observed that the signals are affected by background noise and it affects the accuracy of the system. To increase the accuracy of the system we need to filter the background noise from speech signal acquired. The aim of speech signal enhancement techniques is reducing background noise.

In digital speech signal processing the speech enhancement is having great impact. With the help of mathematical approach and simulation there are many techniques using which speech signal enhancement is performed [4].

In this paper an overview of speech enhancement algorithms used for enhancement of digital speech signal are presented. The paper is organized as follows the section 2 explains what is meant by speech enhancement; section 3 describes types of noise because of which the speech signal can be degraded and its removal techniques. Section 4 describes the categorization of speech signal enhancement techniques followed by conclusion.

2. SPEECH ENHANCEMENT
Speech enhancement is a step in the digital speech signal processing having an objective of increasing the quality of speech signal i.e. to enhance the clarity, intelligibility, understand ability and comprehensibility of speech signal with the help of some algorithm/filter. There are various reasons which leads to degradation of speech signal due to background noise which are captured during the recording like reverberation, babble etc. For specific type of speech enabled applications like speaker recognition, mobile applications, hearing aids, VoIP etc. clean and noise free speech signals are required. The speech enhancement can be achieved by various methods. According to the type of degradation and the noise in the acquired speech signal the approach to speech enhancement varies.

The fig 1 shows the Basic steps of speech enhancement system [5].

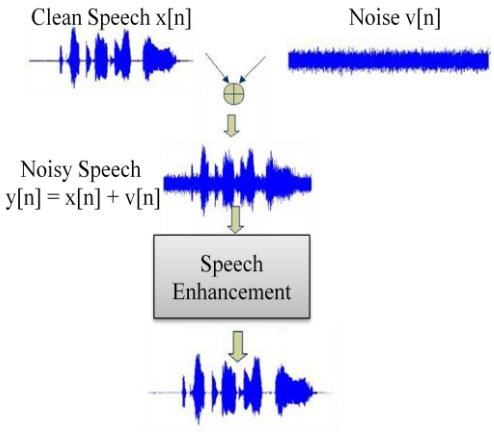

Fig 1: Basic steps of speech enhancement system [5]

3. TYPES OF NOISE AND ITS REMOVAL TECHNIQUES
In this section the review different types of noise removal techniques is described. The speech signal can be degraded because of the noise such as be periodic noise, wide band noise, and interfering speech.

A. Periodic Noise and its Removal Techniques

Stationary filters, adaptive filters, or transform domain filters are used for removing the periodic noise. First approach is stationary in which a bank of notch filters such as twin T-filters can be used as a comb filter for removal of periodic noise. Second is adaptive filters, in which a forward

prediction error filter can be used as an inverse filter which will filter out periodic noise. Third one is transform domain in which periodic noise spectrum can be observed and manipulated. The periodic components can be identified by inspection of the spectrum.

B. Wide Band Noise and its Removal Techniques

Spectral Subtraction method (SS) and adaptive cancellation are used for removal of Wide band noise. In spectral subtraction method, estimated noise spectrum is subtracted from the spectrum of the noisy speech. And with the help of adaptive cancellation the noise correlated with signal can be removed. The correlated signal may be obtained as the estimated channel in the absence of signal. Adaptive filter whose impulse response must be such that the filtered channel noise matches the signal noise may be tuned to remove noise. The coefficients are updated until output reaches minima.

C. Interfering Speech and its removal techniques

When two speech signals are interfering Speech enhancement techniques are not useful. If we are able to identify different pitches the voices of different speakers can be isolated. We must track voiced segments In order that pitch separation works. For recovering desired speaker's harmonics a comb filter can be used provided pitch values are already known. In order to isolate voice of different speakers a transform domain technique can also be used. Assuming that pitch values of speakers are known, we may find Discrete Fourier Transform (DFT) of the mixed signal and track the harmonics of the fundamental frequencies of the two speakers. We have to simply take IDFT of the isolated DFTs to get individual speaker's voices if we can isolate the DFT outputs [6].

4. SPEECH ENHANCEMENT METHODS

There are so many different methods used for speech enhancement some of them are as follows. They can be divided in to two basic categories as: Single Channel Enhancing Techniques and Multi-Channel Enhancing Techniques.

a) Single Chanel Enhancement Techniques

This technique is a common for real time applications such as mobile communication, hearing aids etc. as generally there is no second channel present. This method gives the limited performance as it improves the quality of noisy signal at the cost of some intelligibility. Also as compare to multichannel system this system is easier and cost effective. Generally this system uses different statistics of speech and unwanted noise [7].

1. Spectral Subtraction Method

It is one of the basic methods used for speech enhancement. In the spectral subtraction it is assumed that a signal is formed by two additive components. The speech contains noise can be expressed as

$$y(t) = s(t) + d(t) \text{---------- (1)}$$

Where t is time, $s(t)$ is the uncorrupted speech signal, $d(t)$ is the additive noise signal and $y(t)$ is the corrupted speech signal available for processing. The observed signal is split into overlapping frames using the application of a window function and implemented in the short-time Fourier transform (STFT) magnitude domain. Also in the frequency domain this can be represented as

$$\mathrm{Y}(\omega) = \mathrm{S}(\omega) + \mathrm{D}(\omega) \text{---------- (2)}$$

The estimation of power spectrum of noisy speech can be done as:

$$|\mathrm{Y}(\omega)|^2 = |\mathrm{S}(\omega)|^2 + \delta_{\mathrm{n}}(\omega) \text{ ---------- (3)}$$

Where $\delta_{\mathrm{n}}(\omega)$ are the statistical average values of $|\mathrm{D}(\omega)|^2$ during non-speech period, so eq. (4) - (5) shows the enhanced speech signal amplitude.

$$|\widehat{\mathrm{S}}(\omega)| = [|\mathrm{Y}(\omega)|^2 - \mathrm{E}(|\mathrm{D}(\omega)|^2)]^{1/2} \text{-------- (4)}$$

$$= [|\mathrm{Y}(\omega)|\text{^}2 - \delta_{\mathrm{n}}(\omega)]^{1/2} \text{---------- (5)}$$

Combined with the phase of the noisy signal to synthesize the signal again

$$\mathrm{S}(\omega) = \left|\widehat{\mathrm{S}}(\omega)\right| e^{j\omega g[y(\omega)]} \text{---------- (6)}$$

The reverse short-time Fourier transform is performed to transform the signals into time domain. Traditional spectral subtraction calculation assessing uproarious vitality throughout no speech stage, in any case, it can't upgrade noise throughout speech stage. Additionally the method obliges a VAD that may not work extremely well under low SNR.

2. Spectral Subtraction with Over subtraction Model: (SSOM)

In order to come down with the musical noise effect SSOM procedure was introduced. The perception of musical noise can be reduced using this. This Method does the subtraction of an overestimate of the noise power spectrum and present the resultant spectral components from going below a preset minimum spectral floor value.

3. Non-Linear Spectral Subtraction: (NSS)

This method is based on combination of the two ideas first one is The use of an extended noise and an over subtraction model and second is Non-linear implementation of the subtraction process, considering that the subtraction process must depend on the SNR of the frame, to go to apply less subtraction with high SNRs and vice versa [8].

b) Multi Chanel Enhancement Techniques

The systems which are of this kind are more complex one as compare to single channel systems. This systems takes advantage of available multiple signal inputs to the system and uses noise reference in adaptive noise cancellation device. These systems can do better for non-stationary noises than single channel systems by considering the spatial properties of the noise source and the signal, also limitations inherent to single channel systems [9].

1. Adaptive Noise Cancellation

This method is one of the powerful speech enhancement techniques.Which is based on the auxiliary channels availability, which is known as reference path, where a correlated sample or reference of the contaminating noise is present. Following an adaptive algorithm, this reference input will be filtered in order to subtract the output of this filtering process which is in the main path, where noisy speech is present. The *adaptive noise cancellation* (ANC) cancels the primary unwanted noise *r*(*n*) with is help of introducing a cancelling anti-noise of equal amplitude but opposite phase by using a reference signal. The reference signal generated is derived from one or more sensors located at points which are near the noise and interference sources at the point where the interest signal is weak or undetectable [10].

2. Multisensor Beamforming

A multiple-input and single-output (MISO) application is a *Beamforming,* which consists of multichannel advanced multidimensional (space-time domain) filtering techniques which enhances the desired signal and also suppress the noise signal. In beamforming, the arrangements of two or more microphones are in an array of some geometric shape. Then a *beamformer* is used to filter the sensor outputs and amplifies or attenuates the signals depending on their *direction of arrival* (DOA). The hidden idea of this method is based on the assumption that the contribution of the reflexions is small, and the direction of arrival of the desired signal is known. Then, from the correct alignment of the phase function present in each sensor, enhancement of the desired signal can be done by rejecting all the noisy components not aligned in phase.

The speech enhancement can also be done in both time domain and transform domain as follows [11].

a) Time Domain Method

1. Winer Filtering

Lim and Oppenheim in December 1979 suggested the wiener filter for speech enhancement as an improvement to spectral subtraction. This method is popularly used in so many signal enhancement methods. The basic of Winer filter is getting an estimate of the clean signal from that corrupted by additive noise is. With minimizing the Mean Square Error (MSE) between the desired signal s (n) and the estimated signal ˆs (n) we obtained the estimate.

Solution to this optimization problem in the frequency domain gives the following filter transfer function:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_v(\omega)} \quad \text{---------- (7)}$$

Where $P_s(\omega)$and $P_v(\omega)$are the power spectral densities of the clean and the noise signals, respectively. This formula can be derived considering the signal *s* and the noise *v* as uncorrelated and stationary signals. The SNR is defined by [13]:

$$SNR = \frac{P_s(\omega)}{\hat{P}_s(\omega)} \quad \text{---------- (8)}$$

This definition can be integrated to the Wiener filter equation as follows

$$H(\omega) = \left[1 + \frac{1}{SNR}\right]^{-1} \text{---------- (9)}$$

The fixed frequency response at all frequencies and the requirement to estimate the power spectral density of the clean signal and noise prior to filtering is the drawback of the Wiener filter [12].

2. Kalman Filtering

A generalization of the Winer Filter is the kalman filter. It contains a slowly varying AR model. In the Kalman filtering framework the AR model and the excitation model fit nicely, fully exploiting the capability of the Kalman filter for processing non-stationary signals in an LMMSE optimum manner. The coefficients of AR-model are estimated with the help of decision directed type Power Spectral Subtraction method which is followed by an LPC analysis. Multi- Pulse Linear Predictive Coding (MPLPC) based method is used for the robust estimation of the rapidly time-varying excitation model in the presence of noise. We can say that the Kalman filter combines all the available data measured, also the knowledge of the system and the measurement devices, for producing an estimation of the desired variables in such a way that the error is statistically minimized. One of the most basic differences between the Wiener filter and the Kalman filter is the ability of the latter to accommodate non-stationary signals [13].

3. Linear Predictive Coding

Linear predictive coding (LPC) is a tool mainly used for processing the audio signal and speech processing for representing the spectral envelop of a speech digital signal in a compressed way (using the information of linear prediction model). It start by making assumption LPC starts with the assumption that the speech signal is produced by a buzz at the end of a tube, adding, sometimes, hissing and popping sounds. This model is a good approximation to the reality. The glottis produces the buzz, which is known by his intensity (loudness) and frequency (pitch). The vocal tract generates a tube which is known by his resonances, called formants. The lips, tongue and throat generate the hisses and pops sounds. [14]

LPC does the analysis in the speech signal by using the formants, by removing their effect from the speech signal and estimating the intensity and frequency of the speech signal which are remaining. The removal of formants process is called inverse filtering. The remaining signal after the subtraction is known as residue. The numbers which describe the frequency and intensity of the buzz, the formants and the residue signal can be stored or transmitted. Determine the formants from the original signal is the fundamental problem of the LPC system. So the solution of this is to express each n every sample as a linear combination of previous samples. The coefficients of the equation represent the formants, so we use the LPC system to estimate these coefficients.

4. Transform Domain Method

a. DFT Based (STSA Methods)

This is most known method as these methods have less computational complexity as easy implementation. Uses short time DFT (STDFT) and have been intensively investigated and also known as spectral processing methods. To spectral phase For Human speech perception these methods are not sensitive. But the clean spectral amplitude must be properly extracted from the noisy speech to have acceptable quality speech at output. Hence they are known as short time spectral amplitude (STSA) based methods is the face on which they are is based [].

b. Signal Subspace Method

This method contains the use of a signal dependant transform for decomposing a noisy signal into two separate subspaces, the signal plus noise subspace, and also the noise-only subspace. This transform uses to perform this operation is the Karhuenen-Loeve transform (KLT). This assumption expects that speech can just extent the signal in accumulation to noise subspace, and the noise-just subspace. The KLT components which denote the noise just subspace are nulled, while the modules which represent the noisy signal are modified by a gain function. The enhanced signal is derived from the inverse KLT of the altered components. To improve the quality is the aim here and concurrently minimising any loss in intelligibility. The enhanced speech which is produced by the signal subspace using adaptive noise estimation (SSANE) algorithm, is of a good, natural-sounding quality and contains no audible noise. Still, this algorithm can only update the noise estimate when speech is absent, and suffers degradation in performance in many different noise types [16]. Following Fig 2. Represents the Block diagram of Subspace speech enhancement system.

Fig 2: Block diagram of Subspace speech enhancement system

5. CONCLUSION

Speech enhancement is a technique having objective of increasing the quality of speech signal.In this paper different speech enhancement techniques have discussed. We have studied different Types of Noise and Its Removal Techniques. Also we have seen Speech Enhancement Methods like Single Channel and Multi-Channel Enhancing Techniques and their sub types. Also in Time Domain Method we have seen Winer Filtering, Kalman Filtering, and Linear Predictive Coding. And in Transform Domain Method DFT Based (STSA Methods), Signal Subspace Method.

6. ACKNOWLEDGMENTS

This work is supported by University Grants Commission under the scheme Major Research Project entitled as "Development of Database and Automatic Recognition System for Continuous Marathi Spoken Language for agriculture purpose in Marathwada Region". The authors would also like to thank the Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad for providing the infrastructure to carry out the research.

7. REFERENCES

[1] Sunita Dixit, Dr. MD Yusuf Mulge, "Review on Speech Enhancement Techniques", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 3, Issue. 8, August 2014, pg.285 – 290.

[2] Chanchal Pandey, Sandeep Nemad, "Distinctive Methods for Speech Enhancement using Kalman Filtering", International Journal of Computer Applications (0975 – 8887)Volume 105 – No. 5, November 2014.

[3] P. Bravin Jose, Mrs. M. Jayasanthi, "Review on Speech Enhancement Techniques", KARPAGAM JOURNAL OF ENGINEERING RESEARCH (KJER), Volume No.: 01, Issue No.: 01. 2014.

[4] Vyankatesh Chapke, Prof. Harjeet Kaur, "Review of Speech Enhancement Techniques using Statistical Approach", International Journal of Electronics Communication and Computer Engineering, Volume 5, Issue (4) July, Technovision-2014, ISSN 2249–071X

[5] Ganga Prasad, Surender "A Review of Different Approaches of Spectral Subtraction Algorithms for Speech Enhancement" Department of Electronics, Madhav Institute of Technology & Science Gwalior, M.P. – 474005.

[6] Chaudhari, Amol, and S. B. Dhonde. "A review on speech enhancement techniques." Pervasive Computing (ICPC), 2015 International Conference on. IEEE, 2015.

[7] Pankaj Bactor, Anil Garg, "Different Techniques for the Enhancement of the Intelligibility of a Speech Signal", International Journal of Engineering Research and Development, Volume 2, Issue 2 (July 2012), PP. 57-64.

[8] Yariv Ephraim, Hanoch Lev-Ari and William J.J. Roberts "A Brief Survey of Speech Enhancement" IEEE Sig. Proc. Let., vol. 10,pp. 104-106, April 2003 s.

[9] Lu-ying SUI, Xiong-wei ZHANG, Jian-jun HUANG, Bin ZHOU "An Improved Spectral Subtraction Speech Enhancement Algorithm under Non-stationary Noise" Institute of Command Automation, PLAUST Nanjing, China,IEEE,2011.

[10] Reddy, D.R, "Speech recognition by machine: A review", Proceedings of IEEE (Volume: 64, Issue: 4) ISSN: 0018-9219.

[11] Young, S.J, "Robust continuous speech recognition using parallel model", IEEE Transactions on Speech and Audio Processing (Volume: 4, Issue: 5).

[12] Savita Hooda and Smriti Aggarwal Maharishi Markandeshwar University, Mullana (Ambala), INDIA.

[13] JOSEPH W. PICONE, SENIOR MEMBER, "Signal Modeling Techniques in Speech Recognition", PROCEEDINGS OF THE IEEE, VOL. 81, NO. 9, SEF'TEMBER 1993.

[14] Yoon. B-Y.; Tashev, I. & Acero, A. (2007) Robust Adaptive Beamforming Algorithm Using Instantaneous Direction Of Arrival With Enhanced Noise Suppression Capability. IEEE International Conference on Acoustics, Speech and Signal Processing 1:I-133–I- 136.

[15] Nandini Garg, Jyoti Gupta, "Review on Speech Enhancement using Signal Subspace method", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 5, May 2013.

[16] Barry Commins "Signal Subspace Speech Enhancement with Adaptive Noise Estimation" National University of Ireland, Galway, September 2005.