Speech Enhancement Techniques using Wiener Filter and Subspace Filter

Ankeeta A. Dhande  
Department of Electronics and Communication Engineering  
Priyadarshini bhagwati college of engineering  
Nagpur, India

Dr. N. K. Choudhari  
Department of Electronics and Communication Engineering  
Priyadarshini bhagwati college of engineering  
Nagpur, India

Abstract

In the speech enhancement method by using the wiener filter and subspace filter. Because of uses advantages in reduction in noise with the subspace speech enhancement technology and stable characteristics of the wiener filter. These proposed enhancements of speech method has a better performance. It can be removed colored noise from noisy speech signal. The proposed enhancement of multi-channel speech signal can be obtain a better speech recovery result as compared to the tradition multichannel wiener filter and the subspace filter.

Keywords: Enhancement speech; wiener filter; subspace filter

I. INTRODUCTION

Speech is most important factor of communication for human. Speech can be defined may be delivering thoughts and ideas with the help of vocal sound. Speech captured by microphones in the hearing aids are always corrupted by additive noise [5]. Speech needs to be clean off irrelevant contents. However, removed the irrelevant information. The object of this paper is to enhancement of the speech quality signal[1-3].

Enhancement of speech has been studied of many application such as voice communication, transmitted speech signal and voice control [1]. Noise is everywhere around most of the places we feel is silent will have noise floor well below the full scale level. During the conversation on mobile phone between the person A and person B then these conversation is meaningful speech conversation. The direct sound contaminated by early and the late reflections. This paper will remove the additive noise from the signal recorded and to improving the speech quality, improving speech intelligibity and speech recognition rates [1-6].

II. CLASSIFICATION OF SPEECH ENHANCEMENT TECHNIQUE

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 Channel 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.

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

\[ Y(t) = S(t) + d(t) \] \hspace{1cm} (1)

Where is time, is the uncorrupted speech signal, is the additive noise signal and 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

\[ Y(\omega) = S(\omega) + D(\omega) \] \hspace{1cm} (2)

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 presents 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.

**B. Multi-Channel Enhancement Techniques:**

The systems which are of this kind are more complex one as compare to single channel systems. This system 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.

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 the 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.

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 reflections 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.

### III. Estimation Based Filtering Techniques

The simplest form of speech enhancement primitive is the noise reduction from the noisy speech and is applicable for single channel based speech applications. In this type of speech enhancement techniques, algorithms are either/combinely based on the model of noisy speech and/or perceptual model of speech using masking threshold. The generalized diagram of single channel enhancement technique is shown in Fig.1.

![Fig. 1: Single channel enhancement technique](image)

One of the early papers in speech enhancement considers the problem of estimation of speech parameters from the speech which has been degraded by additive background noise. In this work they propose the two suboptimal procedures which have linear iterative implementations in order to suppress the non-linear effect on the speech parameters due to background noise. In another similar problem of enhancing the speech in presence of additive acoustic noise, spectral decomposition of frame of noisy speech was adopted. The attenuation of particular spectral component was determined based on how much the measured speech plus noise power exceeds an estimation of background noise leading an importance of proper choice of the suppression or subtraction factors. The short-time spectral amplitude (STSA) was used to model the speech and noise spectral components in.

The parametric estimation techniques, where parameters of underlying model, consist of small set of parameters, is determined and then numerical process is used to modify the parameters, can be contrasted by the non-parametric method which can be used as in where no model is assumed and uses non-parametric spectrum estimation techniques. In application point of view, there is work described in, where noisy speech enhancement algorithm has been discussed and implemented to compare its performance against the various levels of LPC (Linear Predictive coefficient) perturbation. Various speech enhancement techniques have been considered here such as spectral subtraction, spectral over subtraction with use of a spectral floor, spectral subtraction with residual noise removal and time and frequency domain adaptive MMSE filtering. The speech signal sued here for recognition...
experimentation was a typical sentence with additive normally distributed white noise distortion. The single channel speech enhancement algorithm at very low SNR has been presented in which uses masking properties of human auditory system. This algorithm is the subtractive type in its nature and subtraction parameter is adapted as per the levels of rough estimate of the background noise and the added musical residual noise and thus making this algorithm adaptable to noise present in every frame of speech. In another interesting research, speech was enhanced from noise along with coding using discrete wavelet packet transform decomposition. Two stages of subtractive-type algorithm used, once estimating noise and subtracting it from noisy speech to have rough estimate of speech; later, this estimate is further used to determine the time-frequency masking threshold assuming high-energy frames of speech will partially mask the input noise and hence reducing the need for a strong enhancement process. The both of these work used Noisex-92 database to evaluate the performance of their proposed algorithms. In yet another similar work, the noise autocorrelation function is estimated during non-speech activity periods and it is used in deciding the masking threshold for the speech enhancement. Here, author also uses frequency to Eigen-domain transformation to provide the upper bound estimate of residual noise to be introduced in the speech. It is believed that the time distribution of speech samples is much better modeled by a Laplacian or a Gamma density functions rather than a Gaussian density function. The same is valid for short time DFT domain, typically, frame size less than 100ms. Optimal estimators for speech enhancement in the Discrete Fourier Transform (DFT) domain is used for estimating complex DFT coefficients in the MMSE sense when the clean speech DFT coefficients are Gamma distributed and the DFT coefficients of the noise are Gaussian or Laplace distributed. When the noise model is a Laplacian density, this estimator outperforms other estimators in the sense it show less annoying random fluctuations in the residual noise than for a Gaussian density noise. In, adaptive estimation of non-stationary noise present in the speech has been presented.

IV. SPEECH QUALITY MEASUREMENTS

This paper work focuses only on objective speech quality measurements because the subjective measurements are time consuming and expensive. In industry, it is very critical to meet software deadlines often. Hence it would be handy to objectively test the software’s performance. A combination of Itakura-Saito scheme mentioned in is used.

A. Objective Speech Quality:

Objective speech quality measures are generally calculated from the original undistorted speech and the distorted speech using some mathematical formula. It does not require human listeners, and so is less expensive and less time consuming. Often, objective measures are used to get a rough estimate of the quality. These estimates are then used iteratively to “screen” subjective quality test conditions so that only the minimum necessary conditions need to be tested subjectively. Many good estimators of subjective quality have been developed, but we still need to evaluate subjective quality at some point since there are still situations where estimations fail. Some objective quality measures are highly correlated with subjective perceived quality, while others are more correlated with subjective intelligibility. In this section, we will describe a few examples of commonly used objective quality measures.

Objective speech quality measures are generally calculated from the original undistorted speech and the distorted speech using some mathematical formula. Objective speech quality measures are based on some physical measurement, typically acoustic pressure or its electrically converted level in case of speech, and some mathematically calculated values from these measurements. It does not require human listeners, and so is less expensive and less time-consuming.

B. Subjective Speech Quality:

Subjective quality measures are measures based on the subjective opinion of a panel of listeners on the quality of the speech sample. Generally, subjective quality can be classified into utilitarian and analytical measures. Utilitarian measures results in a measure of speech quality on a uni-dimensional scale, i.e., a numerical value that rate the quality of speech. This numerical value can be used to compare the speech quality resulting from varying conditions, e.g., coding algorithms, noise levels, etc. On the other hand, analytical measures try to characterize the perceived speech quality on a multidimensional scale, e.g., rough or smooth, bright or muffled. The results of this measure give a value for each of the scale, indicating how the listener perceived the quality on each scale, e.g., how rough or how smooth the listener perceived the test speech sample. In this book, we will only deal with the utilitarian measure. Subjective quality measures are based on comparison of original and processed speech data by a listener or a panel of listeners. Subjective quality can be classified into utilitarian and analytical measures.
V. RESULT

Fig. 5.1: fig1. Amplitude Vs Time waveform and fig2. Shows frequency vs time. In fig.1 shows that plot of three different signal. First is clear i/p signal, second is noisy i/p signal, and third is enhancement o/p signal, where noise is reduce from the input signal to check the enhance o/p.
Fig. 5.2: Fig2 Amplitude Vs Time Waveform And Fig3. Shows Frequency Veruse Time. In Fig.1 Shows That Plot Of Three Different Signal. First Is Clear I/P Signal, Second Is Noisy I/P Signal, And Third Is Enhancement O/P Signal, Where Noise Is Reduce From The Input Signal To Check The Enhance O/P. Red Part Shows The Speech In The Sample Signal.
VI. CONCLUSION

Various speech enhancement approaches have been approached in this paper. In these paper work focus on the improving the quality of speech signal and improving the speech intelligibility and speech recognition rates.
REFERENCES


