IMPROVEMENT OF RECORDED SPEECH USING SPECTRAL SUBTRACTION

Pruthvi Ninganur[1],Sonal Chopade[2],Kruthika Urde[3]

E&TC,DR.D.Y.PATIL KNOWLEDGE CITY,PUNE,MAHARASHTRA,INDIA.

***Abstract*—Speech enhancement aims to improve speech quality and intelligibility by using various techniques and algorithms. Speech signal is always accompanied with some background noises. Speech processing and communication systems are to apply effective noise reduction techniques in order to extract the desired speech signal from its corrupted speech signal. That is, removal of background noise in the noisy speech. Some of noise reduction techniques are used in the speech processing like spectral subtraction, ceptral mean subtraction, blind equalization ,Adaptive wiener filtering ,Kalman filtering etc., are used various enhancement situations. Among this spectral subtraction is oldest one of the first algorithm proposed for removal of background noise. It is a single channel speech enhancement method for enhancement of speech degraded locale noise. The locale noise can disturb our conversation in a noisy environment like auditorium, street, market etc. This paper presents the performance of spectral subtraction algorithm is evaluate of a speech by signal to noise ratio value.**

Keywords— Spectral Subtraction, SNR, Noise Estimation

INTRODUCTION

Speech is a natural and basic way for humans to convey message and thoughts. Speech frequency normally ranges between 3 Hz to 4 KHz depending upon the character. However the human beings have an audible frequency range of 20 Hz to 20 KHz. The most common problem in speech processing is the effect of meddling of noise in the speech signals. The noise masks the speech signal reduces the quality and the speech is greatly affected by presence of backdrop noise.This make the listening task difficult for straight listeners and gives poor performance in some of speech processing like speech recognition ,speech coder and speaker identification etc.. Noise shrinking or speech enrichment algorithm is to improve the performance of communication systems when their input or output signals are corrupted by noise signal.

The main objective of speech enhancement is to improve one or more perceptual aspects of speech such as class or clearness The quality is a subjective measure that indicates the naturalness of the perceived speech and intelligibility is expected by the percentage of words that can be correctly identified by listeners . The performance measures the excellence and intelligibility is very tough to satisfy at the same time. This paper presents speech enhancement method using spectral subtraction algorithm with their performance evaluation.

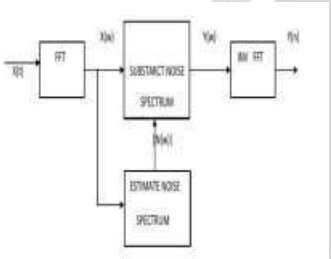
Speech enhancement techniques can be classified into, singlc channel, dual channel or multi-channel enhancement. Although the performance of multichannel speech enhancement is better than that of single channel enhancement, the single channel speech enhancement is still a significant field of research interest because of its simple implementation and ease of computation. In single channel applications, only a single microphone is available and the characterization of noise statistics is extracted during the periods of pauses, which requires a stationary assumption of the background noise. The estimation of the spectral amplitude of the noise data is easier than estimation of both the amplitude and phase. It is revealed that the short time spectral amplitude (STSA) is more important than the phase information for the quality and intelligibility of speech. Based on the STSA estimation, the single channel enhancement technique can be divided into two classes. The first class attempts to estimate the short-time spectral magnitude of the speech by subtracting a noise

The noise is estimated during speech pauses of the noisy speech. The second class applies a spectral subtraction filter (SSF) to the noisy speech, so that the spectral amplitude of enhanced speech can be obtained. The design principle is to select appropriate parameters of the filter to minimize the difference between the enhanced speech and the clean speech.

1.1 **SPECTRAL SUBTRACTION METHOD**

Algorithm and Implementation:

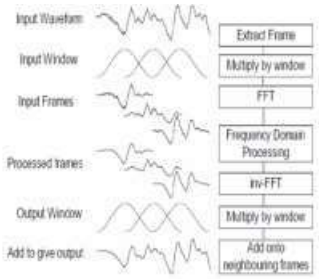
Many different algorithms have been proposed for speech enhancement: the one that we will use is known as spectral subtraction. This technique operates in the frequency domain and makes the assumption that the spectrum of the input signal can be expressed as the sum of the speech spectrum and the noise spectrum. The procedure is illustrated in the diagram below and contains two tricky parts:
 estimating the spectrum of the background noise
 subtracting the noise spectrum from the speech



Block diagram of noise subtraction in spectral domain

1.1.2 Overlap-add processing

To perform frequency-domain processing, it is necessary to split the continuous time domain signal up into overlapping chunks called frames. After processing, the frames are then reassembled to create a continuous output signal. To avoid spectral effects, we multiply the frame by a window function before performing the FFT and again after performing the inverse-FFT.



General steps of overlap-and-add processing on streaming data

2. Principle of Spectral Subtraction Method

Consider a noisy signal which consists of the clean speech degraded by statistically independent additive noise

$$y[n] = s[n] + d[n]$$

where y[n], s[n] and d[n] are the sampled noisy speech, clean speech, and additive noise, respectively. It is assumed that additive noise is zero mean and uncorrelated with the clean speech. Because the speech signal is non-stationary and time variant, the noisy speech signal is often processed on a frame-by-frame. Their representation in the short-time Fourier transform (STFT) domain is given by

$$Y(\omega, k) = S(\omega, k) + D(\omega, k)$$

Where k is a frame number. Throughout this paper, it is assumed that the speech signal is segmented into frames, hence for simplicity, we drop k. Since the speech is assumed to be uncorrelated with the background noise, the short-term power spectrum of y[n]has no cross-terms. Hence

$$|Y(\omega)|^2 = |S(\omega)|^2 + |D(\omega)|^2$$

The speech can be estimated by subtracting a noise estimate from the received signal.

$$|\widehat{S}(\omega)|^2 = |Y(\omega)|^2 - |\widehat{D}(\omega)|^2$$

The estimation of the noise spectrum is obtained by averaging recent speech pauses frames:

$$|\widehat{D}(\omega)|^2 = \frac{1}{M}\sum_{j=0}^{M-1}|Y_{SP_j}(\omega)|^2$$

where M is the number of consecutive frames of speech pauses (SP). If the background noise is stationary, converges to the optimal noise power spectrum estimate as a longer average is taken. The spectral subtraction can also be looked at as a filter, by manipulating such that it can be expressed as the product of the noisy speech spectrum and the spectral subtraction filter (SSF) as:

$$|\widehat{S}(\omega)|^2 = \left(1 - \frac{|\widehat{D}(\omega)|^2}{|Y(\omega)|^2}\right)|Y(\omega)|^2$$
$$= H^2(\omega)|Y(\omega)|^2$$

where H(ω) is the gain function and known spectral subtraction filter (SSF). The H(ω) is a zero phase filter, with its magnitude response in the range of $0 \le H(\omega) \le 1$

$$H(\omega) = \left\{\max\left(0, 1 - \frac{|\widehat{D}(\omega)|^2}{|Y(\omega)|^2}\right)\right\}^{1/2}$$

To reconstruct the resulting signal, the phase estimate of the speech is also needed. A common phase estimation method is to adopt the phase of the noisy signal as the phase of the estimated clean speech signal , based on the notion that short-term phase is relatively unimportant to human ears. Then, the speech signal in a frame is estimated as

$$\widehat{S}(\omega) = |\widehat{S}(\omega)|e^{j<Y(\omega)} = H(\omega)Y(\omega)$$

The estimated speech waveform is recovered in the time domain by inverse Fourier transforming S (ω) using an overlap and add approach.
The spectral subtraction method, although reducing the noise significantly, it has some severe drawbacks. Thus, it is clear that the effectiveness of spectral subtraction is heavily dependent on accurate noise estimation, which is a difficult task to achieve in most conditions. When the noise estimate is less than perfect, two major problems occur, remnant noise with musical structure and speech distortion

3. Speech Objective quality measures

The objective comparison of three single channel speech enhancements is carried by evaluating performance of parameters such as, Mean Square Error (MSE), Normalized Mean Square Error (NRMSE), Signal to Noise Ratio (SNR), and Root Mean Square Error. It is based on mathematical comparison of the original and processed speech signal.

3.1 Signal to Noise Ratio (SNR)

It is most widely used and popular method to measure the quality of speech. It is ratio of signal to noise power in decibels.

$$SNR_{dB} = 10 \log_{10}((\sigma x)^2 / (\sigma_d)^2)$$

Where (σx)2 is the mean square of speech signal and (σd)2 is the mean square difference between the original and where N is length of input speech signal, x(n) is input speech signal and r(n) is reconstructed speech signal

4. Conclusion

In this paper, a comparison and simulation study of different forms of spectral subtractive-type algorithms for suppression of additive noise is presented. In particular, algorithms based on short time. Fourier transforms are examined and the limitations of spectral subtraction method are discussed briefly.

. 5.References

1] Speech Enhancement by Spectral Subtraction Method Kaladharan N Assistant professor

2] Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study Navneet Upadhyaya,∗ and Abhijit Karmakarb

3] http://dsp-book.narod.ru/304.pdf

4]http://www1.icsi.berkeley.edu/ftp/pub/speech/papers/keele96-usu.pdf