OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Speech Enhancement Based On Spectral Subtraction For Speech Recognition System With Dpcm

A.T. Rajamanickam, N.P.Subiramaniyam, A.Balamurugan*, and T.S.Yuvarani

Department of Electronics and Communication Systems
Nehru Arts and Science College, Coimbatore- 641 105, India

ABSTRACT: This paper presents a speech enhancement algorithm using a one-microphone for automatic speech recognition system. Speech signal received in an enclosed room is distorted by reflections from walls and other objectives. This distortion effect named as "reverberation" degrades the fidelity and intelligibility of input speech in acoustic systems such as hand-free conference telephones and automatic speech recognition. In this project, we consider the importance effect of reverberation on speech signal which is referred to as overlap masking, i.e. the energy of the previous phonemes is smeared over time, and overlaps following phonemes. To reduce this effect, we introduced a one-microphone speech dereverberation algorithm based on spectral subtraction. After processing of spectral subtraction, a residue reverberation still fills some of the silent gaps right after high-intensity speech sections. Therefore, we employ a Voice Activity Detector (VAD) using spectral entropy and then attenuate these silent gaps. After the process the signal will be encoded by the DPCM coding.

KEYWORDS: Voice Activity Detector, reverberation, DPCM encoding, Spectral Subtraction.

I. INTRODUCTION

Signal Processing

Digital Signal Processing is distinguished from other areas of computer science by the unique type of data it uses: signals. In most cases, these signals originate as sensory data from the real world: seismic vibrations, visual images, sound waves, etc. DSP is the mathematics, the algorithms and the techniques used to manipulate these signals after they have been converted into a digital form. This includes a wide variety of goals, such as: enhancement of visual images, recognition and generation of speech, compression of data for storage and transmission, etc.

Audio Processing

The two principal human senses are vision and hearing. Correspondingly, much of DSP is related to image and audio processing.

DSP can provide several important functions during mix down, including: filtering, signal addition and subtraction, signal editing, etc. One of the most interesting DSP applications in music preparation is artificial reverberation.

Speech Generation

Speech generation and recognition are used to communicate between human and machines. Two approaches are used for computer generated speech: digital recording and vocal tract simulation. In digital recording, the voice of a human speaker is digitized and stored, usually in a compressed form. During playback, the stored data are uncompressed and converted back into an analog signal. This is the most common method of digital speech generation used today.

Vocal tract simulators are more complicated, trying to mimic the physical mechanisms by which human create speech.

Speech Recognition

Acoustic-phonetic recognition is based on distinguishing the phonemes of a language. First, the speech is analyzed and a set of phoneme hypotheses are made.

These hypotheses correspond to the closest recognized phonemes in the order that they are introduced to the system. Next, the phoneme hypotheses are compared against stored words and the word that best matches the hypothesis is picked.

Existing System

In existing system, a multi microphone for signaling input. That is more than one microphone used in a seminar hall or room. When several microphones are placed in a room, it will get the signal easily from all the directions

After removing the noise signal using spectral subtraction, some of the silent gaps will be present in a signal.

Proposed System

In this system, we are using a single microphone system [2]. So reverberation in signal will occur more. That is very much higher than multi microphone system. That are eliminated by spectral subtraction and the silent gaps also be removed by the Voice Activity Detector [3].

After processing the signal, the output signal is encoded using DPCM encoding at transmitter and decoding the process at the receiver.

Problem Definition

Reverberation is an acoustical distortion which degrades the fidelity and intelligibility of speech signal in a speech recognition system. This Paper presents a speech enhancement algorithm using a one-microphone for automatic speech recognition system. The proposed algorithm is based on a simple spectral subtraction.

Overview

The spectral subtraction method is a well-known noise reduction technique. Most implementations and variations of the basic technique advocate subtraction of the noise spectrum estimate over the entire speech spectrum. However, in real world noise is mostly colored and does not affect the speech signal uniformly over the entire spectrum.

To improve the system performance, we employ a method of Voice Activity Detection (VAD) using spectral entropy [3]. VAD also known as speech activity detection or speech detection is a technique used in speech processing in which the presence or absence of human speech is detected. The main uses of VAD are in speech coding and speech recognition. It can facilitate speech processing, and can also be used to deactivate some processes during non-speech section of an audio session.

Distortion effect named as "reverberation" degrades the fidelity and intelligibility of input speech in acoustic systems such as hand-free conference telephones and automatic speech recognition. Therefore to improve the performance of speech recognition system, it is necessary to investigate the application of signal processing techniques to the speech enhancement.

Here, we consider the importance effect of reverberation on speech signal which is referred to as overlap masking. To reduce this effect, we introduced a one-microphone speech dereverberation algorithm based on spectral subtraction.

Spectral subtraction has been used widely in speech enhancement [2]. After processing of spectral subtraction, a residue reverberation still fills some of the silent gaps right after high-intensity speech sections. Therefore, to further improve system performance by reduction of this residue reverberation, we employ a Voice Activity Detector (VAD) using spectral entropy and then attenuate these silent gaps. After the process the signal will be encoded by the DPCM coding.

Block Diagram

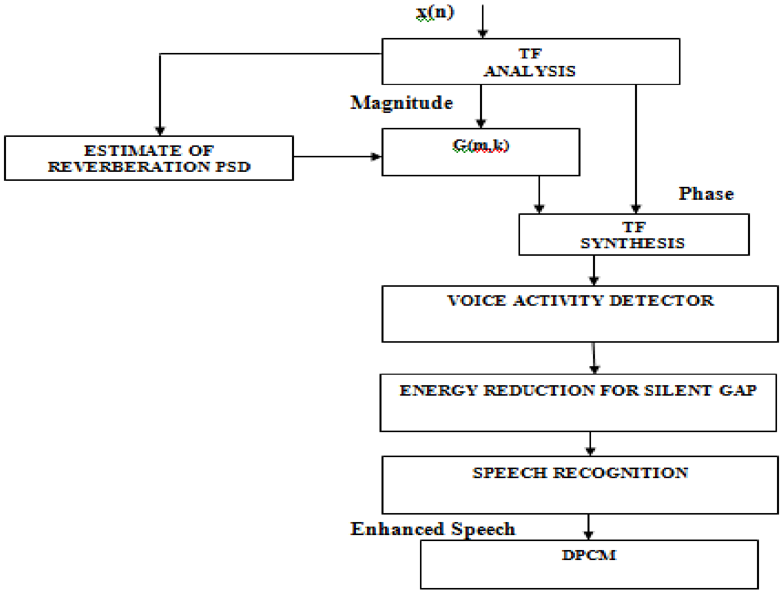


Figure: Block Diagram of Speech Enhancement Algorithm

The block diagram of the speech enhancement algorithm is shown. Prior to speech recognition, input speech signal is pre-processed by spectral subtraction and reverberation reduction for silent gap with VAD [2]. The received speech signal $x(n)$ is decomposed into a Short-Time Fourier Transform (STFT)[1]. The analysis window of time domain is Hamming window and overlap between two successive windows is set to 50%. Then the Power Spectral
Density (PSD) of the reverberation is estimated by autocorrelation function of received signal $x(n)$.

The square root of this estimate is then subtracted from magnitude spectrum of the reverberated signal that yielding an estimate of the magnitude spectrum of the dereverberated signal. This is in practice realized by a short-term spectral attenuation, equivalent to spectral subtraction. One problem of a result from spectral subtracted speech signal is that residue reverberation still fills some of the silent gaps right after high-intensity speech sections.

Therefore it is necessary to employ the VAD techniques to identify and then attenuate these silent gaps. In this paper we used VAD using feature of spectral entropy which performs better in terms of correct decision for silent gaps than typical feature of energy threshold.

Voice Activity Detection

The basic function of a VAD algorithm is to extract some measured features or quantities from the input signal and to compare these values with thresholds, usually extracted from the characteristics of the noise and speech signals. Then, voice-active decision is made if the measured values exceed the thresholds.

Algorithm Overview

The typical design of a VAD algorithm is as follows

1. There may first be a noise reduction stage, e.g. via spectral subtraction.
2. Then some features or quantities are calculated from a section of the input signal.
3. A classification rule is applied to classify the section as speech or non-speech - often this classification rule finds when a value exceeds a threshold.

The Process Of Echo Cancellation

An echo canceller is basically a device that detects and removes the echo of the signal from the far end after it has echoed on the local end's equipment. In the case of circuit switched long distance networks, echo cancellers reside in the metropolitan Central Offices that connect to the long distance network. These echo cancellers remove electrical echoes made noticeable by delay in the long distance network.

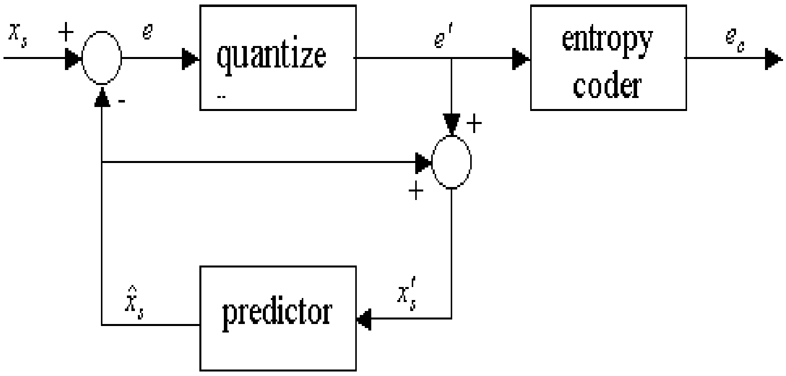
An echo canceller consists of three main functional components:

- Adaptive filter.
- Doubletalk detector.
- Non-linear processor.

Enhancement Of Noisy Speech

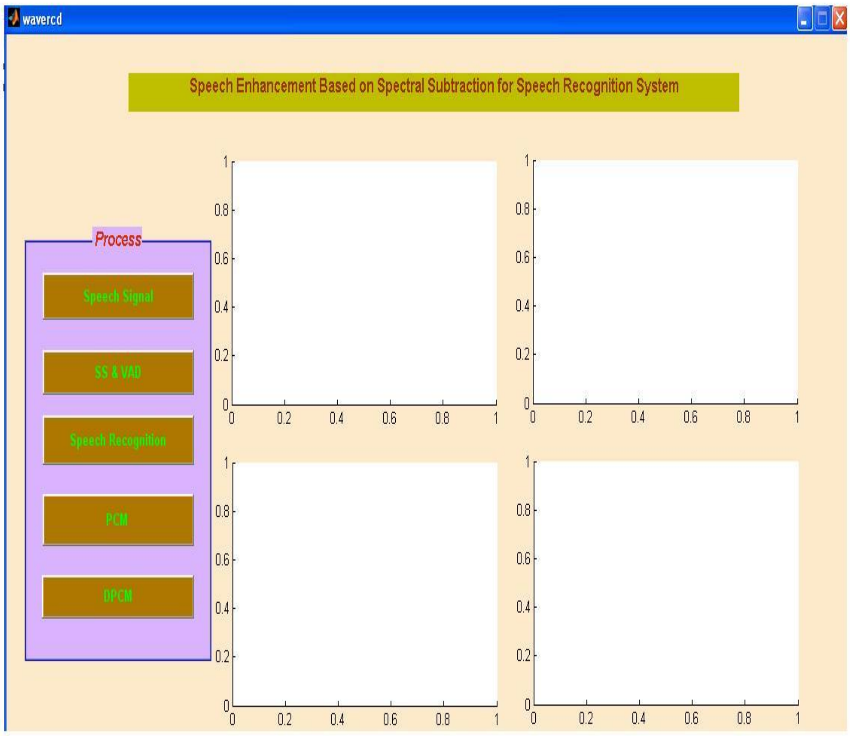
One of the accepted conventional techniques for noise suppression is spectral subtraction, in which the noise power spectrum is estimated in intervals between speeches and subtracted from a power spectrum of the signal [2]. The enhanced signal is then reconstructed by an overlap-add inverse Fourier transform using the modified magnitude and the original noisy phase of the signal spectrum.

Differential Pulse Code Modulation

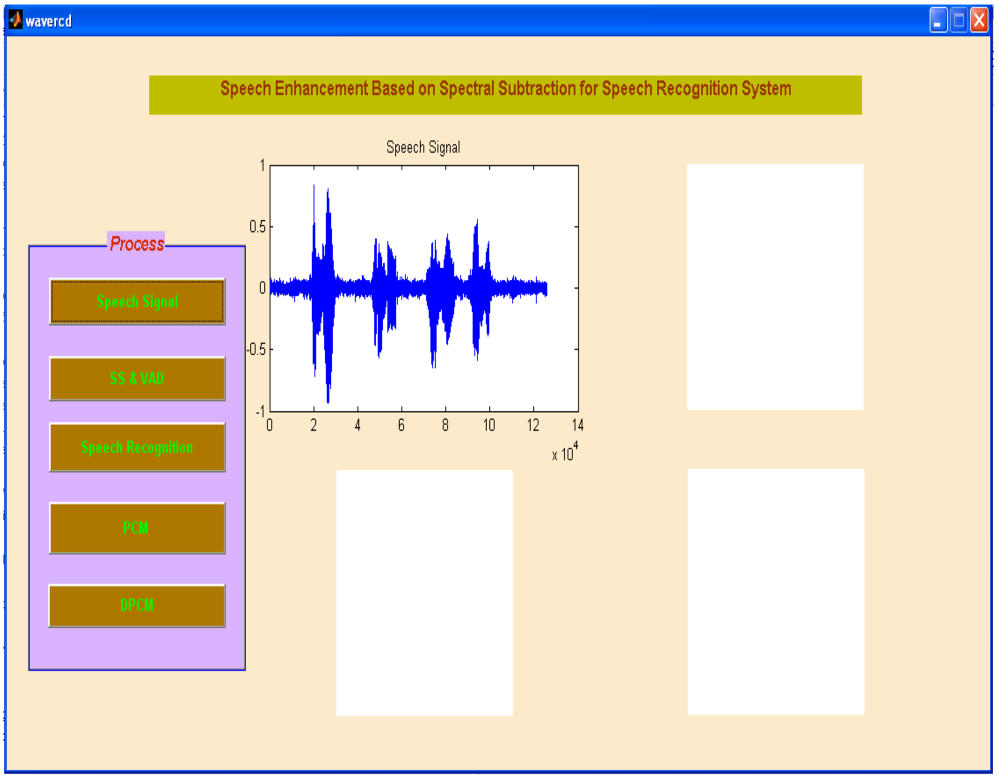


Differential pulse code modulation (DPCM) is method of converting analog to digital signal in which analog signal is sampled and then difference between actual sample value and its predicted value is quantized and then encoded forming digital value. Concept of DPCM is coding a difference. It is based on the fact that most source signals shows significant correlation between successive samples so encoding uses redundancy in sample values which implies lower bit rate.

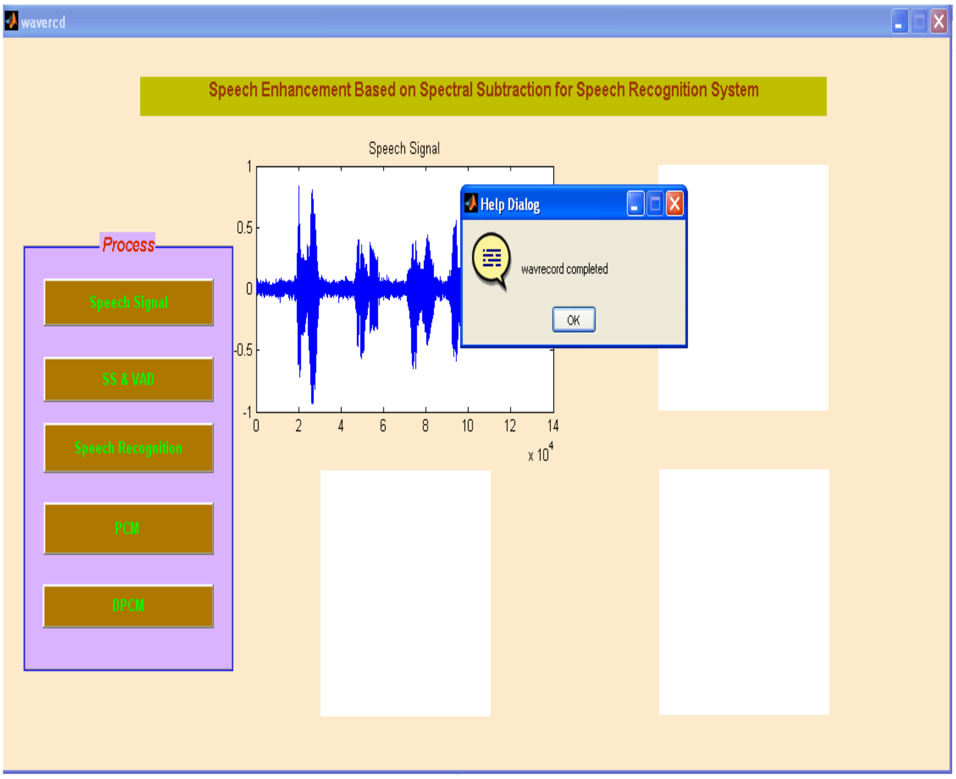
Outputs



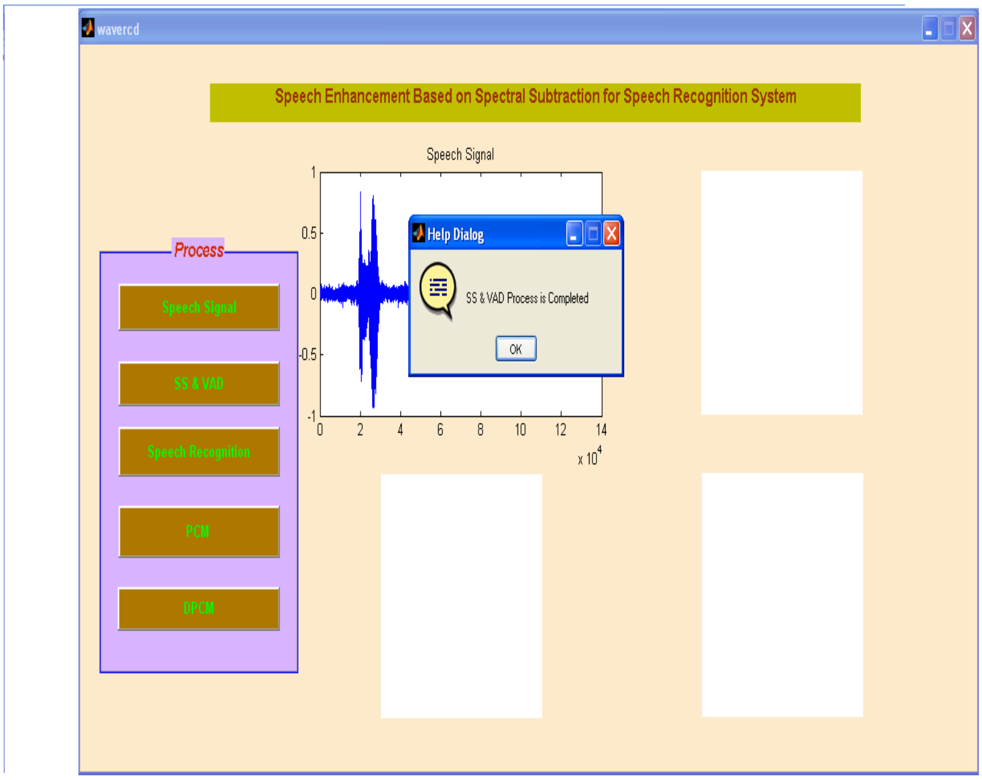
Main Window



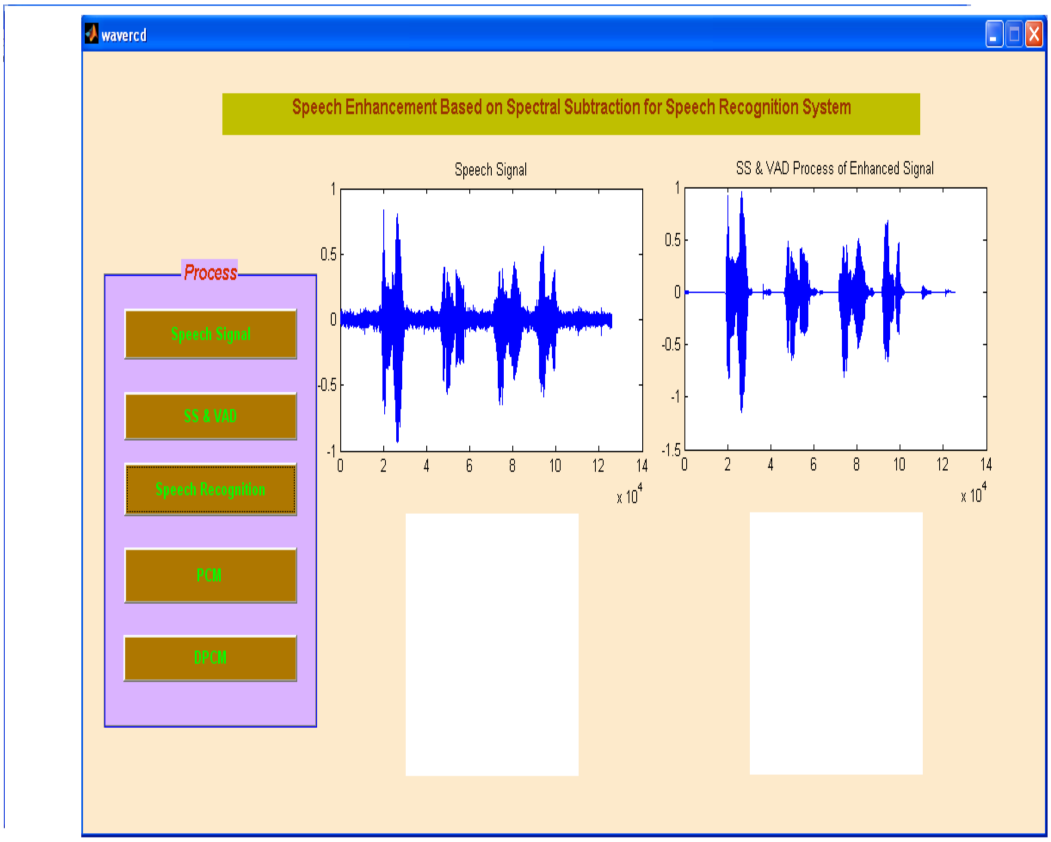
Input



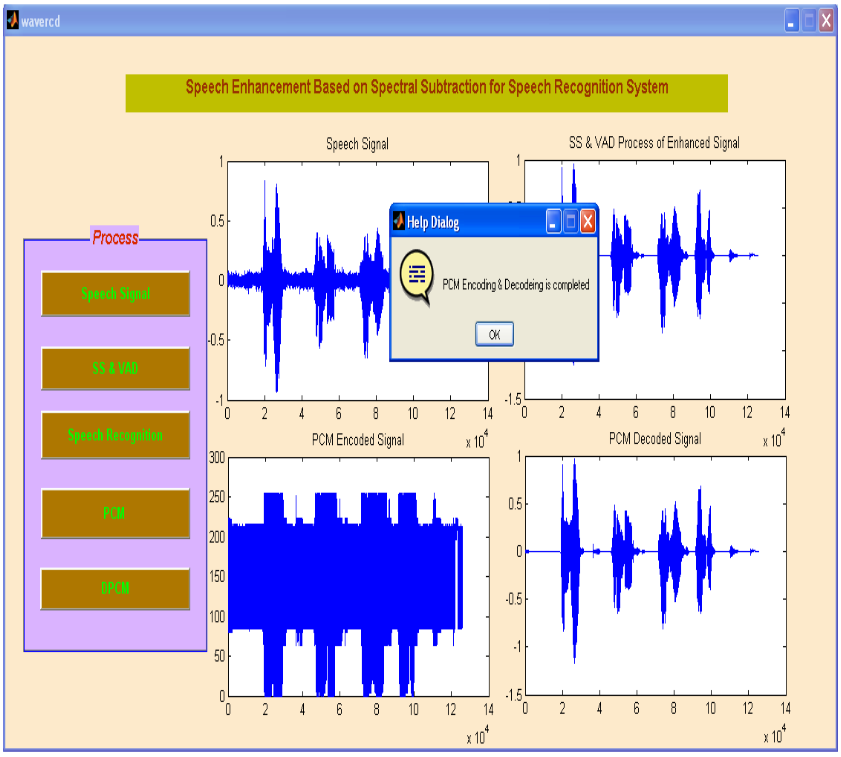
Input With Dialog



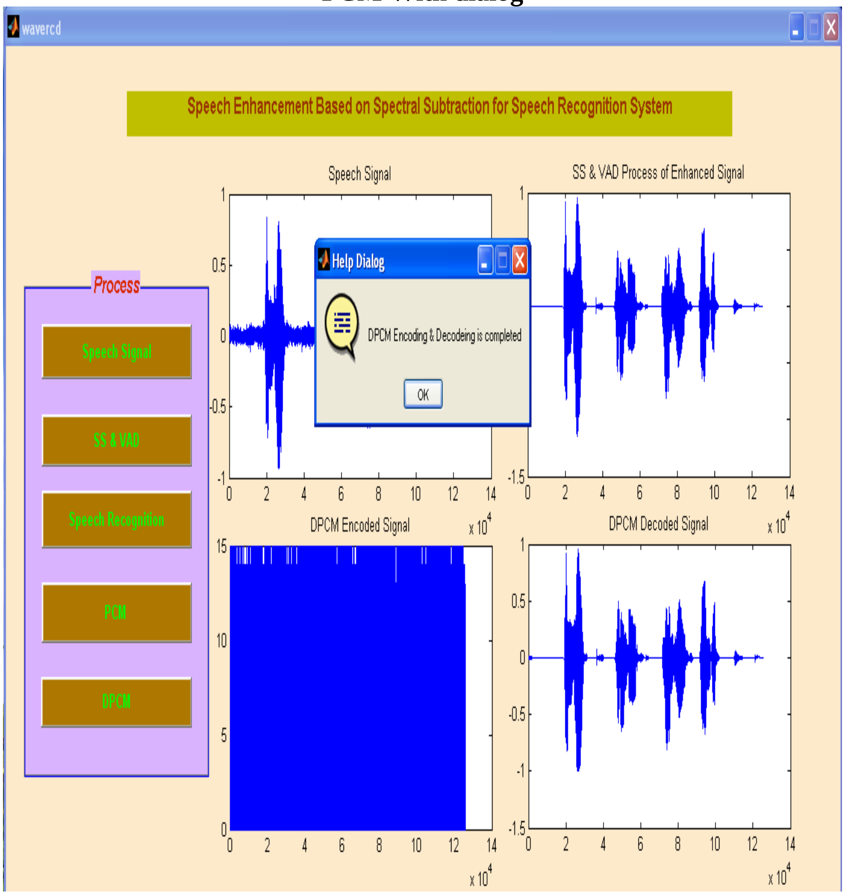
SS And VAD



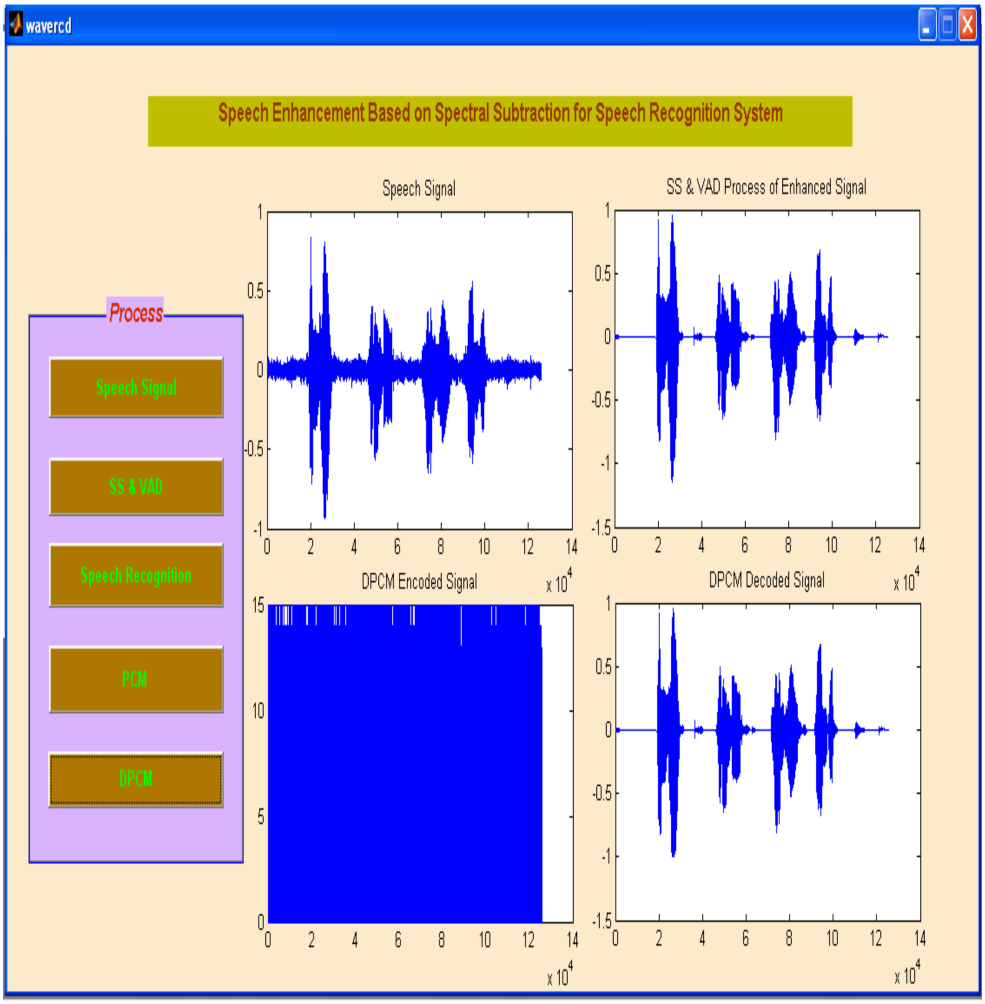
Speech Recognition



PCM With dialog



DPCM With dialog



II. CONCLUSION

The proposed dereverberation method for speech recognition system was designed using spectral subtraction and VAD algorithm. We tested this method by comparing with previous method in terms of values of Reverberation Reduction and speech recognition scores. As a result, the proposed method represents a good performance than previous method using features of energy detection.

REFERENCES

[1]. E.A.P. Habets, "Single-Channel Speech Dereverberation based on Spectral Subtraction," In Proc. ProRISC 2004, the 15th Annual Workshop on Circuits, Systems and Signal Processing, Veldhoven, Netherlands, pp. 250-254. 2004. Mingyang Wu and DeLiang Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," IEEE Trans. Speech Audio Process., Vol. 14, no. 3, pp. 774 - 784, 2006.

[2]. R. V. Prasad, R. Muralishankar and S. Vijay, "Voice Activity Detection for VoIP-An Information Theoretic Approach," in proc. IEEE Int. Conf. Telecommunications, pp. 1-6, 2006.