Development of algorithm for voice operated switch for digital audio control systems

Gokula Krishnan [1], Ram Singh[2], Satyanarayana[3]

[1](M.Tech, DSCE, ECE Department, SNIST, Hyderabad, INDIA)
[2](Scientist 'E', RCMA, DRDO, Hyderabad, INDIA)
[3](Assistant professor, ECE Department, SNIST, Hyderabad, INDIA)

ABSTRACT: *VOS (voice Operated Switch) is a switch that operates when sound over a certain threshold is detected. It is usually used to turn on a transmitter or recorder when someone speaks and turn it off when they stop speaking. It is used instead of a push-to-talk button on transmitters or to save storage space on recording devices. Unlike push-to-talk (PTT) operation, VOS is automatic. The user can keep his hands free while talking. But VOS also has some significant disadvantages that explain why PTT is still common. The Present VOS technology works based on the comparison of the energy estimate in speech band and the noise band. The bandwidth of speech band is from 300Hz to 3.3 KHz and for noise band it is from 3.3 KHz to 6.6 KHz. The major disadvantage with this technique is false operation of VOS due to noise interference or heavy breathing or a side conversation. Hence, a DSP based algorithm is proposed to be developed for DACS (Digital Audio Control System) to improve the performance and the noise suppression.*

Keywords: *Envelope dynamics, Noise estimation, Speech pause detection, SNR (Signal to Noise ratio), VAD (Voice activity detection).*

I. INTRODUCTION

Now a day's speech recognition is becoming more and more popular technology in the society. The innovative technologies in mobile telecommunication, robust speech recognition and digital hearing aids are a strongly dynamic force in the development of real-time noise reduction algorithms. The number of publications on single-microphone noise reduction algorithms indicates an unbroken interest in this research field over the past two or three decades. A decisive point for these kind of algorithms is the synchronized estimate of the object speech spectrum and the interfering noise spectrum in particular. Cockpit voice recorder in aircraft records many cockpit voices, such as speaker voices, noises and background sounds. In these communications, the speech signal which is transmitted is corrupted by additive acoustical noises. These noises may be from non-stationary source and hence there may be situations where only noise segment or non-speech segment is present, hence it is necessary to update the noise spectrum estimate as often as possible to sustain an effective noise reduction. This can be done, whenever target speech is absent, which means that the input signal consists of only noise. Computational and memory requirements should be as low as possible since these algorithms may be ported on a digital circuit.

Speech detection is a process where speech segments be reliably separated from non-speech and activates VOS when only speech segments are detected. The required characteristics of an ideal endpoint detector are: reliability, robustness, accuracy, adaptation, implicitly, real-time processing and know a priori knowledge of the noise.

Earlier, the VOS technology works based on the comparison of the energy estimate in speech band and the noise band. We calculate the signal to noise ratio for both the bands i.e. speech band and noise band and based on a threshold the operation of VOS will take place. The major disadvantage with this technique is false operation of VOS due to noise interference or heavy breathing or a side conversation or any other background noise such as helicopter noise or babble noise etc. The techniques used in this algorithm are discussed in the later sections.

Hence VAD (Voice activity detector) algorithms are developed and employed to make voice operated switch work efficiently and this also increases the switching speed of the voice operated switch.

VAD techniques are designed using various methods. Most of them use heuristically chose statistical properties of speech parameters like: energy, pitch, entropy etc. Therefore, the performances of different VAD are different and varying according to the level and type of signal-to-noise ratio (SNR). As a result, the performances of different speech based systems are significantly sensitive to the employed VAD technique. Therefore, VAD algorithm should be carefully chosen while designing a speech based system.

Several VAD algorithms have been standardized for specific applications. The most commonly mentioned includes ITU-T Recommendation G.729 Annex B [1] and ETSI AMR Option 2 VAD [2], which aim to design simple features that can be implemented in embedded applications efficiently, such as audio recording and transmission on mobile phones. In these systems, speech are often recorded with close-talk microphones, which ensures the sound level of speech is always much higher than background noise. Back then, VAD algorithms often dealt with only little or no noise corruption in speech coding applications and with separate recording utterances in speech recognition systems. Up to recently, advances in various speech applications require the detection of human speech in a continuous real-time fashion, and is often corrupted by a wide variety classes of noise. Algorithms for VAD had grown accordingly over the years. Most of the speech activity detectors are based on either time domain or frequency domain approach. Sangwan, Chiranth [3] have compared various VAD algorithms. The core of any VAD proposed consists of two parts: a 'feature extraction' and a 'speech/ non-speech decision mechanism.

It is necessary to update the noise spectrum estimate as often as possible to sustain an effective noise reduction. Different algorithms have been proposed which continuously update the noise estimate and hence avoid the need for explicit speech pause detection. Martin [4] uses the minimum of the sub-band signal power within a time window of about 1S as an estimate of the noise power in the respective sub-band. Paul proposed a continuous noise estimation scheme similar to Martin's which is computationally more efficient. This scheme was, however, not systematically tested.

Hirsch and Ehrlicher proposed an algorithm [5] which is based on the observation that the most commonly occurring spectral magnitude value in clean speech. Hence, having noisy speech their algorithm measures the distribution density function of the spectral magnitude and density determines the maxima which are then used as an estimate of the respective noise magnitude. These kind of algorithms which avoid speech pause detection for noise estimation are supposed to cope better with non-stationary (i.e., fluctuating) noise, since they are generally faster in their adaptation to changing noise levels even during speech activity. On the other hand, the continuous update of the noise estimate (Independently in the sub-bands) is susceptible to erroneously capture speech energy. This, however, leads inevitably to speech deterioration in a subsequent noise reduction process. Fischer and Stahl proposed a spectral subtraction noise reduction algorithm with a continuous noise spectrum up- dating scheme. They found that the corruption of the noise estimate by speech is too large to be further considered and conclude that voice activity detection plays an important role and cannot be fully omitted, Recently Nemer et al. proposed to use the kurtosis (fourth-order statistics) of the noisy signal to continuously estimate speech and noise energies. The examples presented used noisy speech signals with positive signal-to-noise ratios (SNRs) and yield promising results, but further research is required to extend these results to negative SNR s and different classes of noise, respectively. Dendrinos and Bakamidis [6] presented an algorithm for determining the starting and ending points of speech segments in colored-noise environments through singular value decomposition based on some thresholds which have been determined experimentally. Good performance was proved for SNRs higher than 0 db. However, the complexity of the algorithm makes a real-time implementation difficult.

Weiwu Jiang, Wai Kit Lo [7] presented a novel VAD algorithm using MVSS (Maximum Values of Sub-band SNR) as a decision feature. Here the given the spectrum of a speech utterance transformed by DFT, it first divides the whole spectrum into several sub-bands. Secondly, sub-band SNRs are estimated and maximum values of sub-band SNR (MVSS) are extracted as detection features. The background noise estimation and final VAD decision are made by comparing feature value with an estimated threshold. During initialization, the estimated noise spectrum and threshold are calculated by assuming that speech always follows an initial period of noise.

As the basic requirement to develop or modify an algorithm is to identify the noise i.e. the noisy speech sample is processed by voice operated switch to allow the presence of audio in case of speech and mute when only noise is present.
Based on the above algorithms a modified algorithm is presented and compared with the conventional energy based algorithm.

This paper is organized as it starts with a review of the conventional energy based algorithm and proposed algorithm. Next section reports the expected results of the algorithms under various background noise conditions. Finally, conclusions are presented.

II. ALGORITHMS

The algorithms that were implemented for voice operated switch are discussed in this chapter. In all the mentioned algorithms the plain speech is mixed with various noises like white noise, pink noise, brown noise, helicopter noise etc. with 0db, 3db, 5db, 10db, 15db SNR levels (Signal to Noise Ratio). It will be called as noisy speech from now, which are used in the experiments.

2.1 Conventional energy based algorithm

This algorithm works based on the comparison of the energy estimate of the noisy speech. In general the bandwidth of the speech band is from 300Hz to 3.3 KHz and for noise band it is from 3.3 KHz to 6.6 KHz. Therefore using the filters the noisy speech is divided into two bands, that is the speech band and the noise band by assuming the cut off frequency of 3000 Hz. The Butterworth filter was preferred over other filters based on the parameters suitable to the algorithm.

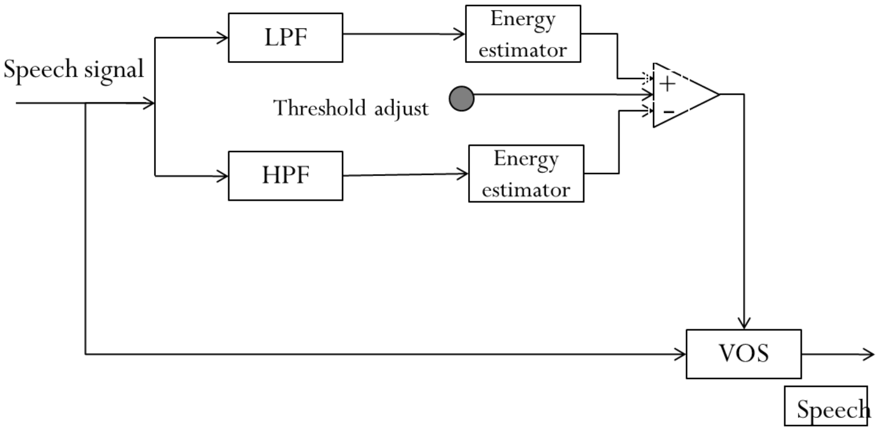

Fig1: Conventional energy based algorithm block diagram

The energies of both the speech band and noise band are calculated by considering a window or frame length of 160 samples. Later the Signal to Noise Ratio of the each frame is determined. The threshold is calculated based on the previous experiments.

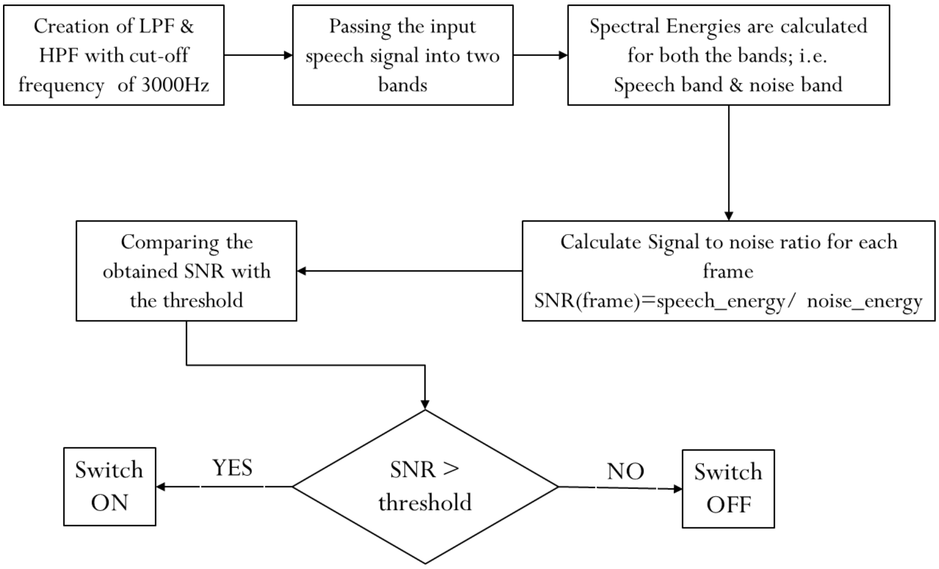
Creation of LPF & HPF with cut-off frequency of 3000Hz
Passing the input speech signal into two bands
Spectral Energies are calculated for both the bands; i.e. Speech band & noise band
Calculate Signal to noise ratio for each frame SNR(frame)=speech_energy/ noise_energy
Comparing the obtained SNR with the threshold
SNR > threshold
YES
NO
Switch ON
Switch OFF

Fig2 : Conventional energy based algorithm - flowchart

The operation of the Voice operated switch is based on the obtained SNR when compared with the assumed threshold. i.e., when the SNR is greater than threshold then switch gets activated (VOS=1). When the SNR is less than threshold then the switch deactivates (VOS=0). The noisy speech sample is processed by voice operated switch to allow the presence of audio in case of speech (VOS=1) and mute when only noise (VOS=0) is present. The simulation and results are shown in the next section.

The performance of the conventional energy based algorithm was not as anticipated and the result was not satisfactory, will be discussed in further section.

2.2 Power envelope based algorithm

An algorithm is proposed which detects speech pauses by adaptively tracking minima and maxima in a noisy signal's power envelope both for the broadband signal and for the high-pass and low-pass filtered signal.

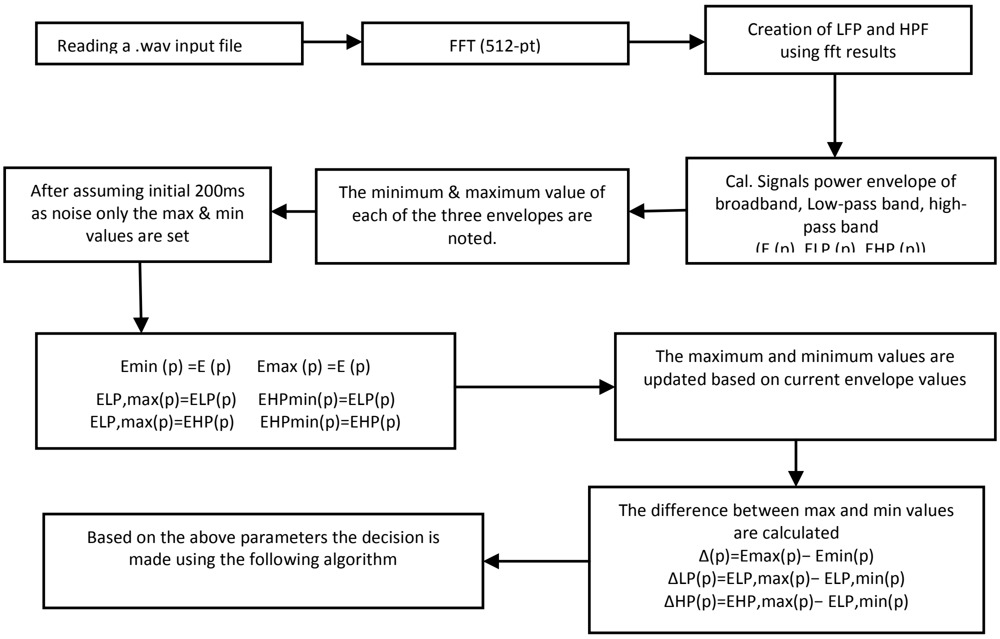
Reading a .wav input file
FFT (512-pt)
Creation of LFP and HPF using fft results
Cal. Signals power envelope of broadband, Low-pass band, high-pass band (E (p), ELP (p), EHP (p))
The minimum & maximum value of each of the three envelopes are noted.
After assuming initial 200ms as noise only the max & min values are set
Emin (p) =E (p) Emax (p) =E (p)
ELP,max(p)=ELP(p) EHPmin(p)=ELP(p)
ELP,max(p)=EHP(p) EHPmin(p)=EHP(p)
The maximum and minimum values are updated based on current envelope values
The difference between max and min values are calculated
Δ(p)=Emax(p)– Emin(p)
ΔLP(p)=ELP,max(p)– ELP,min(p)
ΔHP(p)=EHP,max(p)– ELP,min(p)
Based on the above parameters the decision is made using the following algorithm

Fig3: Power envelope based algorithm- Flowchart

The speech pause detection algorithm calculates the signal's temporal power envelope E (p) by summing up the squares of the spectral components of the input signal in each short-time frame
: $E(p) = \sum_k |X\ (p, \omega k)|^2$……………. (1)

Here, X (p, ωk) denotes the spectral component of the noisy input signal at frequency ωk at time frame. In addition a low-pass band envelope and a high-pass band envelope are calculated:
: $E_{LP}(p) = \sum_l |X\ (p, \omega k)|^2$……… (2)
: $E_{HP}(p) = \sum_m |X\ (p, \omega k)|^2$…….. (3)

Where "l" runs over all spectral components up to the cut-off frequency, and "m" runs over the remaining spectral components. In order to slightly smooth the envelopes, E(p), $E_{LP}(pp)$and $E_{HP}(pp)$are averaged over a few frames by a recursive low-pass filter of first order with a release time constant ȚE no smoothing is per- formed in case of an increase in energy (i.e., attack time zero) to avoid smearing over onsets. The algorithm tracks the minimum and maximum value of each envelope and uses these for the speech pause decision as described by the following scheme.

I. After an assumed 200 ms initial phase of noise only the minimum and maximum values are set as follows:
:$E_{min}(p) \equiv E_P(p)$ $E_{max}(p) \equiv E_P(p)$
:$E_{LP,min}(p) \equiv E_{LP}(p)$ $E_{LP,max}(pp) \equiv E_{LP}(p)$
:$E_{HP,min}(p) \equiv E_{HP}(p)$ $E_{HP,max}(p) \equiv E_{HP}(p)$……. (4)

This guarantees that the minimum envelope values correspond roughly with the noise energy at the beginning.

II. The minimum and maximum values are updated for each of the three envelopes in the following manner.

a) If the current envelope value is larger than the maximum value for the corresponding envelope, then the maximum value is set to the current value. Otherwise, the maximum value slowly decays. This is done by a recursive low-pass filter of first order with a release time constant τ_{decay}, which takes as input the current envelope value.

b) If the current envelope value is smaller than the minimum value for the corresponding envelope, then the minimum value is set to the current value. Otherwise, the minimum value is slowly raised. This is done by a recursive low-pass filter of first order with attack time constant τ_{rise}, which takes as input the current envelope value.

III. The differences between the maximum and the minimum values are calculated for each envelope
$\Delta(p) = E_{max}(p) - E_{min}(p)$
$\Delta_{LP}(p) = E_{Lp,max}(p) - E_{Lp,min}(p)$
$\Delta_{HP}(p) = E_{Hp,max}(p) - E_{Hp,min}(p)$…… (5)

IV. Three different criteria are introduced of which only one has to be true for making the decision that target speech is not present in the actual frame: a) the speech pause decision can be made because of a low signal dynamics in both the low-pass and the high-pass band(Syn speech pause); b) the decision can be based on the low-pass band information (LP Speech Pause); and c) it can be made upon the high-band information (HP Speech Pause). These decision criteria are derived as follows:

a) If ΔLP is smaller than some threshold ή and ΔHP (p) < ή then it is assumed that only noise is present due to the very small dynamic range of the signal (Dyn Speech Pause).

b) If a) is not true, it is checked whether ΔLP is bigger than (otherwise the dynamic range in the low-pass band is very small and it should not receive too much attention no LP Speech Pause). Now, if the difference between the current $E_{Lp}(p)$ and $E_{Lp,min}(p)$ of the low-pass band envelope is smaller than some fraction pc of ΔLP (which means that the actual envelope is near its minimum), a closer look at the high-pass band is necessary to support speech pause detection.

Case 1) ΔHP of the high-pass band is smaller than threshold in this case no additional information can be obtained from the high-pass band because of its small dynamic range. Now, if at least E (p) (the signal's envelope) lies in the lower half of its dynamic range [i.e., in the lower half between $E_{min}(p)$and $E_{max}(p)$ the current frame can be assumed to be a speech pause because of the closeness of the low-pass band energy to its minimum value (LP speech pause) otherwise, however, there is not enough support for a speech pause decision (No LP Speech Pause).

Case 2) ΔHP is bigger than two times the threshold ή .In this case, there is enough dynamic range to pay attention to the high-pass band. Thus, it is demanded that the difference between the current $E_{Hp}(p)$and $E_{Hp,min}(p)$of the high-pass envelope is smaller than two times the fraction pc of ΔHP to support the small envelope value in the low-pass band. Then a noise-only frame is assumed (LP Speech Pause). This demand is not as strict as that for the low-pass band, to account for the case that the disturbing noise has a rather high-frequency characteristic. But if this condition is not fulfilled, speech may be present in the actual frame (no LP Speech Pause).

Case 3) ΔHP Is smaller than two times the threshold ή , but bigger than. In this case, which is not as clear as Case 2, it is only demanded that EHP(p) (the high-pass Envelope) lies in the lower half of its dynamic range to support the small envelope value in the low-pass band Then it is assumed that target speech is absent (LP Speech Pause). However if this condition is not fulfilled, speech may be present in the actual frame (no LP Speech pause).

c) Condition b) accounts for the case that the disturbing noise has a rather high frequency characteristic, hence the speech pause detection should mainly be made upon the information in the low-pass band. To account also for the case that it has be checked but now with reverse roles of the low-pass and the high–pass bands to determine whether target speech is absent (HP Speech Pause).

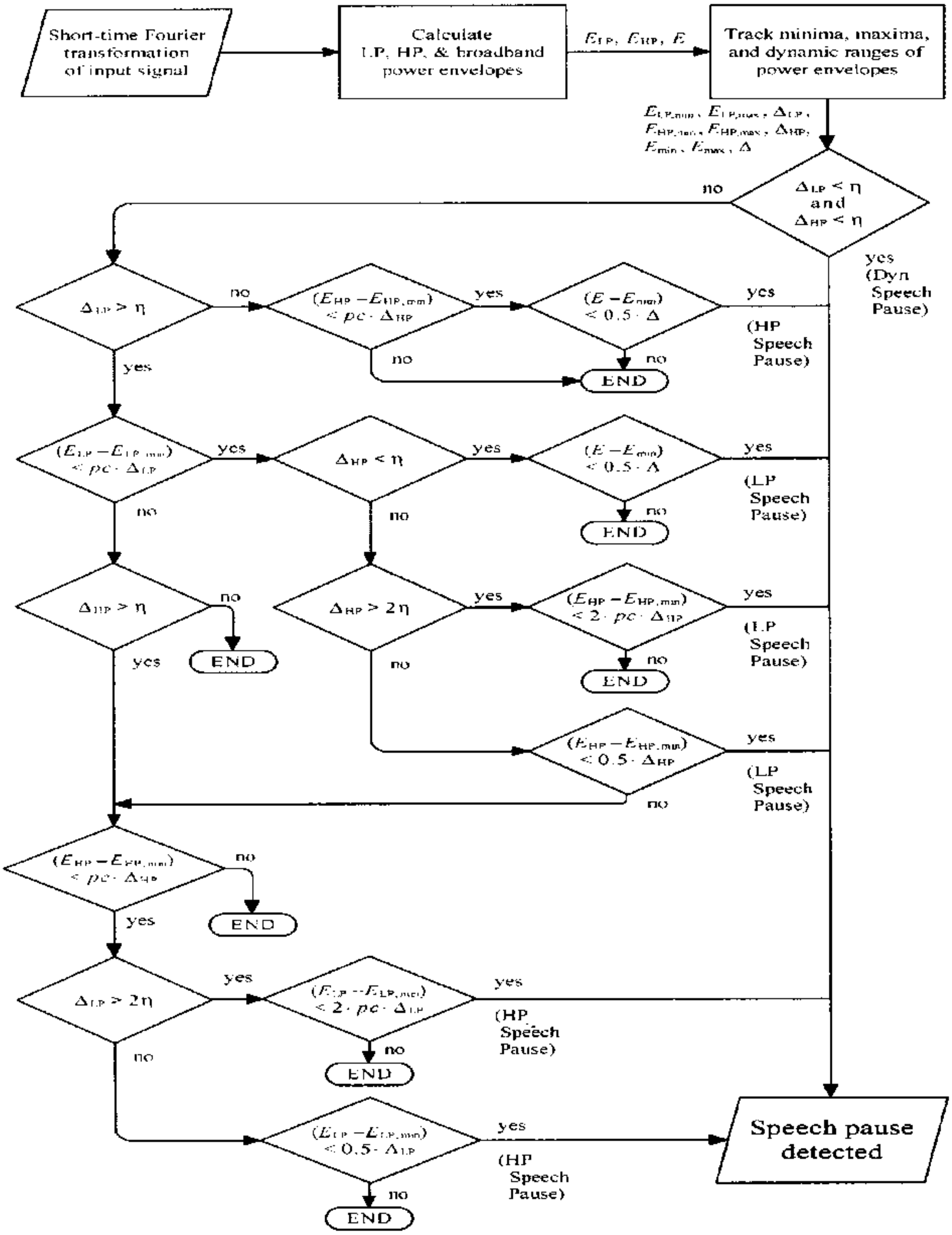
Short-time Fourier transformation of input signal
Calculate LP, HP, & broadband power envelopes
E_{LP}, E_{HP}, E
Track minima, maxima, and dynamic ranges of power envelopes
$E_{LP,min}, E_{LP,max}, \Delta_{LP}, E_{HP,min}, E_{HP,max}, \Delta_{HP}, E_{min}, E_{max}, \Delta$
$\Delta_{LP} < \eta$ and $\Delta_{HP} < \eta$
no
yes (Dyn Speech Pause)
$\Delta_{LP} > \eta$
no
$(E_{HP} - E_{HP,min}) < pc \cdot \Delta_{HP}$
yes
$(E - E_{min}) < 0.5 \cdot \Delta$
yes (HP Speech Pause)
no
no
END
yes
$(E_{LP} - E_{LP,min}) < pc \cdot \Delta_{LP}$
yes
$\Delta_{HP} < \eta$
yes
$(E - E_{min}) < 0.5 \cdot \Delta$
yes (LP Speech Pause)
no
no
no
END
$\Delta_{HP} > \eta$
no
END
yes
$\Delta_{HP} > 2\eta$
yes
$(E_{HP} - E_{HP,min}) < 2 \cdot pc \cdot \Delta_{HP}$
yes (LP Speech Pause)
no
END
no
$(E_{HP} - E_{HP,min}) < 0.5 \cdot \Delta_{HP}$
yes (LP Speech Pause)
no
$(E_{HP} - E_{HP,min}) < pc \cdot \Delta_{HP}$
no
END
yes
$\Delta_{LP} > 2\eta$
yes
$(E_{LP} - E_{LP,min}) < 2 \cdot pc \cdot \Delta_{LP}$
yes (HP Speech Pause)
no
END
no
$(E_{LP} - E_{LP,min}) < 0.5 \cdot \Delta_{LP}$
yes (HP Speech Pause)
no
END
Speech pause detected

Fig 4: Flowchart of the proposed speech pause detection algorithm operating on a single time frame

Fig above gives a flowchart of the proposed power envelope detection algorithm. The flowchart is not fully symmetrical with respect to LP and HP speech pause detection since several redundant tests are omitted.

Due to its flexible design this novel approach for speech pause detection can easily be adjusted to obtain a rather low false- alarm rate by adapting the main parameters ή and *pc*. Generally, a Low-false-alarms rate is desirable to reduce speech distortions in the subsequent noise reduction process. However, this also results in a reduced hit rate.

During the development of the algorithm noisy signals generated from various different noise types and speech signals at several SNRs were used for performance verification. Finally, the following values were chosen for the free parameters: The input signal was digitized with a sampling Frequency of 8000 Hz and partitioned in Hamming-windowed segments of length 20 ms with 50% overlap. These segments were padded with zeros and a 512-point FFT was performed. This framework is compatible with most single-microphone noise reduction algorithms which can thus easily be integrated. Such short segments are motivated by the fact that then the same signal analysis and synthesis as necessary for a real-time noise reduction environment can be used. The cut-off frequency between low-pass and high-pass band was set to 2 kHz, motivated by the fact that excluding speech frequencies above 1.9 kHz has a roughly similar effect on speech intelligibility as excluding those below this value. With these settings a good approximation to the actual dynamic range of the signal and of its “placement” in the level area under a variety of conditions was achieved. However, systematic variations of these parameters were not investigated. The threshold was set to 5 dB and the fraction was set to 0.1.

III. SIMULATION RESULTS

In this section the procedure to carry out experiments is discussed and comparison between the proposed algorithm and conventional energy based algorithm is discussed

3.1 Procedure

The plain speech is recorded at a sampling rate of 8000Hz and is mixed with various noises like white noise, pink noise, brown noise, helicopter noise etc. with 0db, 3db, 5db, 10db, 15db SNR levels. The algorithms are implemented and simulated in MATLAB and simulation results are discussed below.

3.2 Simulation results

3.2.1 Conventional energy based algorithm results

The following are the results of conventional energy based algorithm for voice operated switch

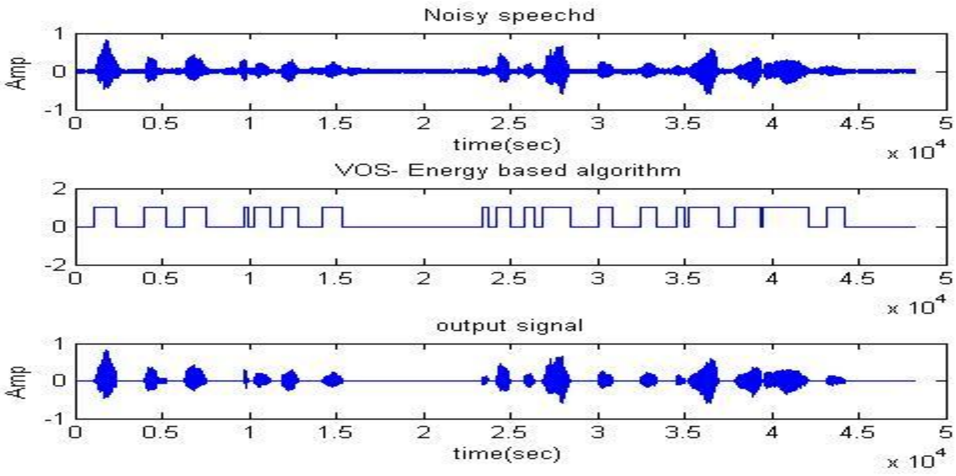

Fig 5: Plot of (a) 10dB white noise mixed speech signal (b) Switch on/off for energy based algorithm(c) output

Fig 5 shows the results of voice operated switch when plain speech is mixed with 10 dB white noise and the output of conventional energy based algorithm.

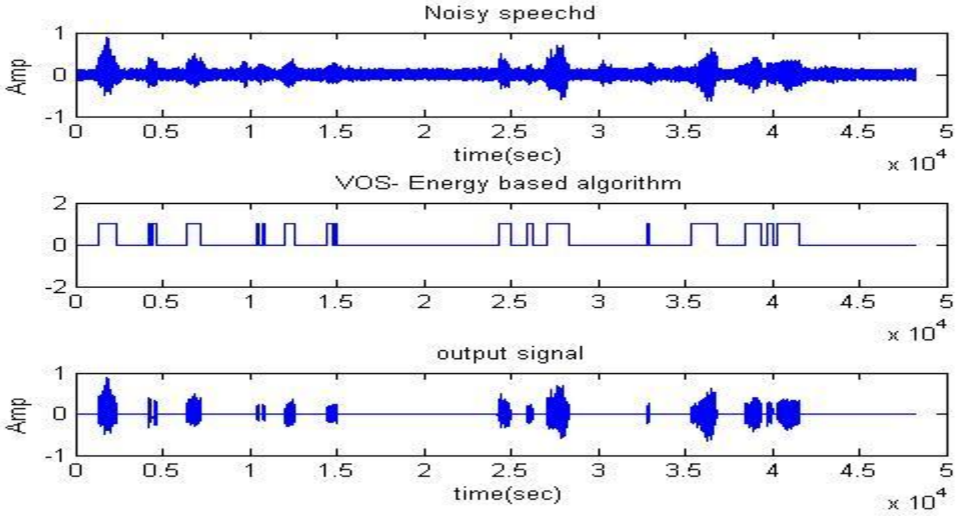

Fig 6: Plot of (a) 5dB white noise mixed speech signal (b) Switch on/off for energy based algorithm(c) output

Fig 6 shows the results of voice operated switch when plain speech is mixed with 5 dB white noise and the output of conventional energy based algorithm.

Fig 7 shows the results of voice operated switch when plain speech is mixed with 10 dB pink noise and the output of conventional energy based algorithm.

Fig 8 shows the results of voice operated switch when plain speech is mixed with 5 dB helicopter noise and the output of conventional energy based algorithm.

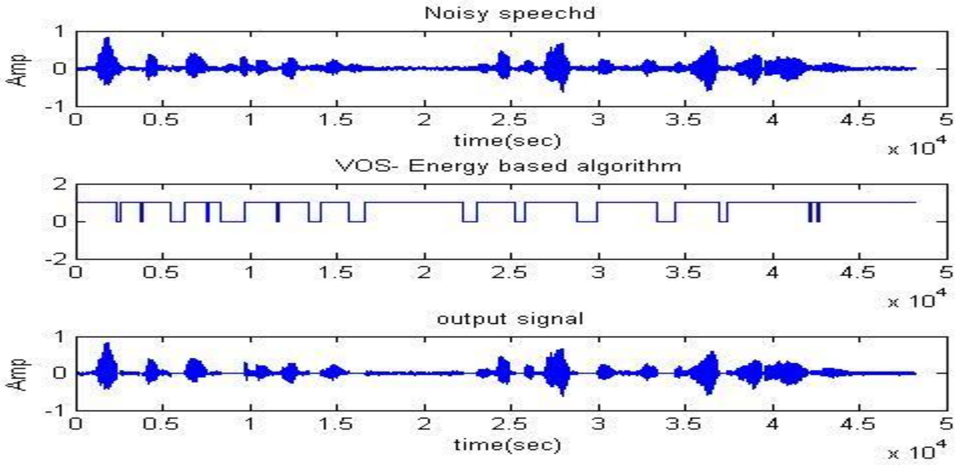

Fig 7: Plot of (a) 10dB pink noise mixed speech signal (b) Switch on/off for energy based algorithm(c) output

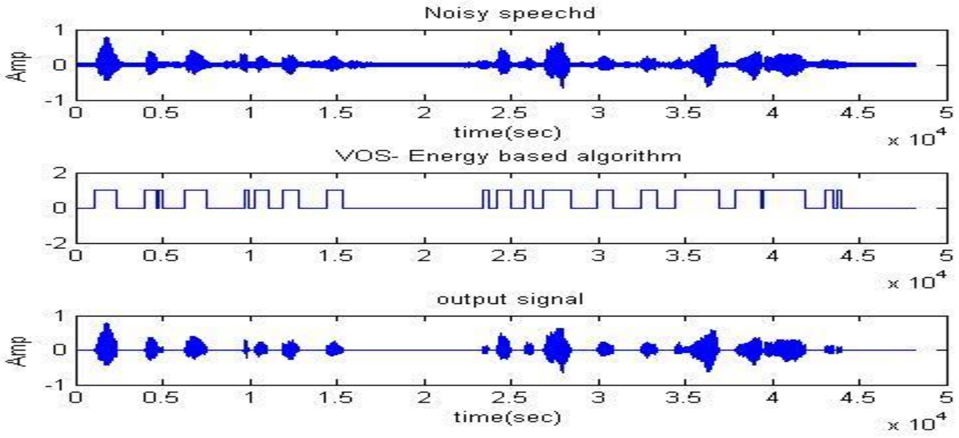

Fig 8: Plot of (a) 5dB helicopter noise mixed speech signal (b) Switch on/off for energy based algorithm(c) output

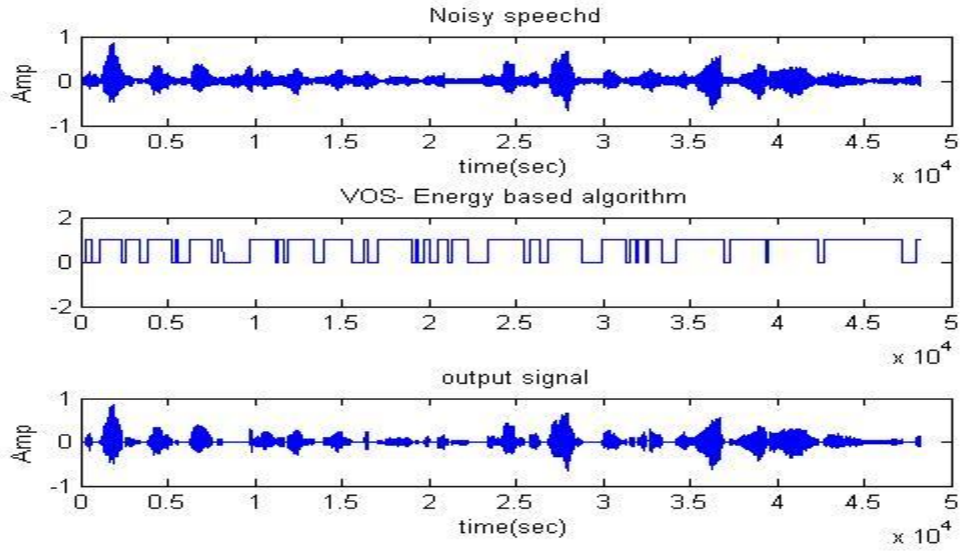

Fig 9: Plot of (a) 10dB babble noise mixed speech signal (b) Switch on/off for energy based algorithm(c) output

Fig 9 shows the results of voice operated switch when plain speech is mixed with 10 dB babble noise and the output of conventional energy based algorithm.

3.2.2 Power envelope based algorithm results

The following are the results of power envelope based algorithm for voice operated switch

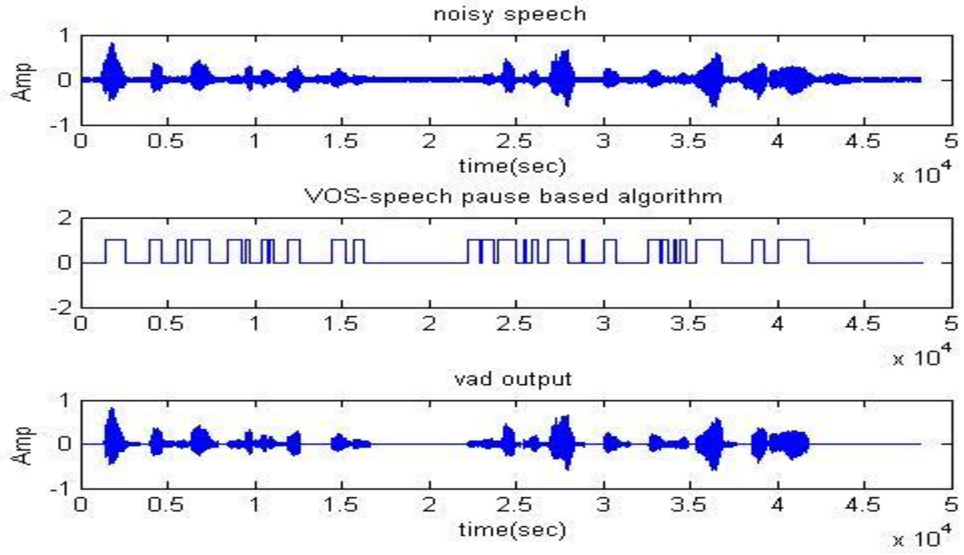

Fig 10: Plot of (a) 10dB white noise mixed speech signal (b) Switch on/off for power envelope based algorithm(c) output

Fig 10 shows the results of voice operated switch when plain speech is mixed with 10 dB white noise and the output of power envelope based algorithm

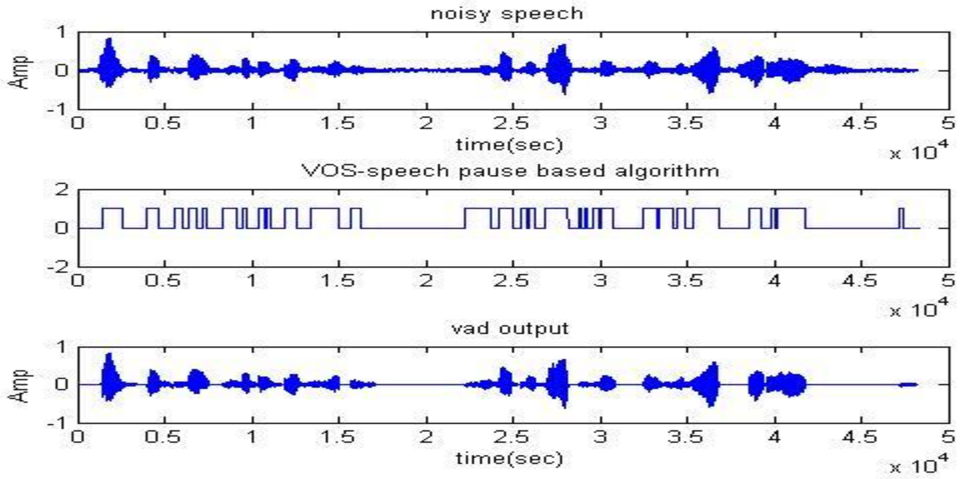

Fig 11: Plot of (a) 10dB pink noise mixed speech signal (b) Switch on/off for power envelope based algorithm(c) output

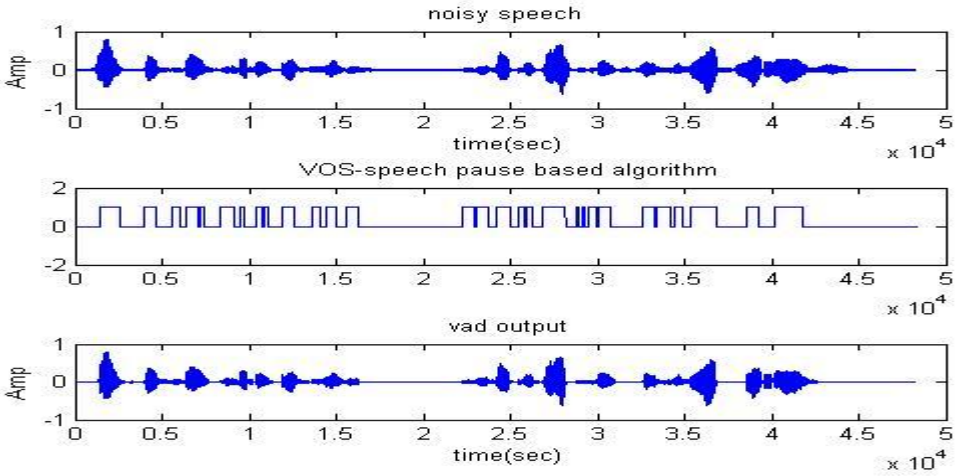

Fig 12: Plot of (a) 10dB helicopter noise mixed speech signal (b) Switch on/off for power envelope based algorithm(c) output

Fig 11 shows the results of voice operated switch when plain speech is mixed with 10 dB pink noise and the output of power envelope based algorithm

Fig 12 shows the results of voice operated switch when plain speech is mixed with 10 dB helicopter noise and the output of power envelope based algorithm

Fig 13 shows the results of voice operated switch when plain speech is mixed with 5 dB helicopter noise and the output of power envelope based algorithm

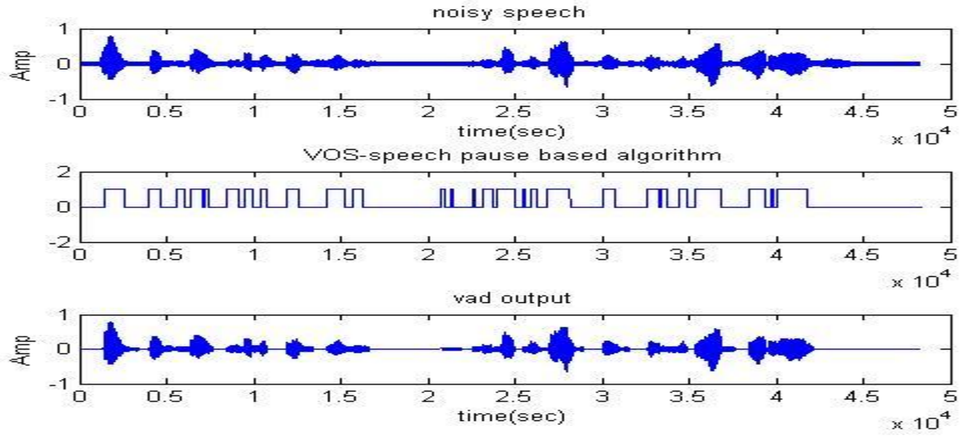

Fig 13: Plot of (a) 5dB helicopter noise mixed speech signal (b) Switch on/off for power envelope based algorithm(c) output

There were other simulation results considered and only few of them. In energy based algorithm the switch activates at few places even when only noise is present. The major disadvantage because of this algorithm is it doesn't perform well due to noise interference or heavy breathing or a side conversation etc.

IV. CONCLUSION

The present VOS technology works based on the conventional energy based algorithm, which works well for white noise at 10dB, 15dB etc., but results are not promising for helicopter, pink and babble noise. The pilot or co-pilot has to adjust the threshold (SNR) level which is inside the audio control system (as shown in the figure1) to hear a clear speech signal. Using the proposed algorithm, the pilot doesn't need to change the threshold control every time (unlike conventional energy based algorithm) to hear the speech signal clearly i.e. the speech is transmitted when speech is detected (VOS=1), it is muted when noise is present (VOS=0).

From the simulation results and observation it is clearly seen that the power envelope detection based algorithm gives better and efficient performance when compared with the conventional energy based algorithm. Many research are however to be analyzed on basis to modify the algorithm to improve the performance on different noise levels. The proposed power envelope detection algorithm maintains a low and approximately constant false-alarm rate over a wide range of SNRs. The hit rate decreases only slightly at poorer SNRs.

The efficient voice operated switches can be father used for commercial and defense purpose which leads to save battery life and helpful to save storage spaces on recording devices.

V. ACKNOWLEDGEMENT

I am grateful to my guides Mr. Ram Singh, Scientist 'E', and Mr. Satyanarayana, Assistant professor for constantly encouraging me, for their inspiration and innovative ideas, valuable suggestions and guidance without which I would not have achieved our goal.

REFERNCES

[1] Benyassine, E. Shlomot, H.-Y. Su, and E. Yuen, "A robust low complexity voice activity detection algorithm for speech communication systems," in *Speech Coding For Telecommunications Proceeding, 1997, 1997 IEEE Workshop on, p. 97, 1997.*

[2] ETSI, "Universal mobile telecommunication systems mandatory speech codec speech processing functions, AMR speech codec; voice activity detector *(3GPP TS 26.094 version 4.0.0 release 4)," 2001.*

[3] Sangwan, Chiranth M. C, R. Shah, V. Gaurav, R. Venkatesha Prasad, "Voice Activity Detection for VoIPTime and Frequency domain Solutions", *TENTH ANNUAL IEEE symposium on multimedia communications and signal processing, bangalore, nov 2001, pp 20-24.*

[4] R. Martin, "An efficient algorithm to estimate the instantaneous SNR of speech signals," *in Proc. EUROSPEECH'93, vol. 1, 1993.*

[5] H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," *in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing 1995, vol. 1, 1995, pp. 153–156.*

[6] M. Dendrinos and S. Bakamidis, "Voice activity detection in colorednoise environment through singular value decomposition," *in Proc. 5th Int. Conf. Signal Processing Applications and Technology.Waltham, MA: DSP Associates, 1994, vol. 1, pp. 137–141.*

[7] Weiwu Jiang, Wai Kit Lo and Helen Meng," A New Voice Activity Detection Method Using Maximized Sub-band SNR",*978-1-4244-5857-8/10/$26.00©2010, ieee,icalip2010*

[8] "Digital Processing of Speech Signals" by L. R. Rabiner and R. W. Schafer, *Prentice Hall Publication, 1978*

[9] "Discrete-time speech signal processing" by Thomas F. Quatieri