Voice Activity Detection Using Feature Vectors

Jagadeesh Thati[1] and Fazal Noorbasha[2]

[1]Department of Electronics and communications, JNTU Hyderabad (AP.), India
[2]Department of Physics &Electronics, Dr. Harisingh Gour Central University, Sagar (MP.), India
e-mail: [1] jagadeeshthati@gmail.com , [2] skfazalahmed@radiffmail.com

Abstract

Effective speech communication can be achieved by taking the speech signal when microphone is active and suppressing the noise when it is passive. The model we proposed in this paper is to take Feature vectors of pre-defined speech and noise signal's, which already are stored for processing. Then centriods of the Feature Vectors were calculated using k-means algorithm. The Minimum distance between framed feature vectors of input signal and centriods of pre-defined signals are estimated using Euclidian distance. The ratio obtained between noise and speech minimum distance vectors will represent the voice activity at the microphone. The evaluation of the ratio indicates the significant performance of voice activity detection in noisy environment as well.

Keywords: Voice activity, Feature vectors, k-means clustering.

1. Introduction

Effective speech communication is increasingly becoming importance in the modern era of telecommunication system. This increase can be largely attributed to the desire to lower the average bit-rate of speech communication systems, whether this is for mobile telephony or VoIP communications [1]. Algorithm for effective speech communication can be developed from various methods specific to different applications. Voice activity detection is one of the methods which can be used to improve the quality of the speech signal, minimizing the channel capacity and suppressing the noise when there is no speech signal.

2. Review of the State of Art

Voice activity detection techniques have been formularized for speech enhancement, low bandwidth and minimum error rate while transmission of speech signals. Bayesian adaptive algorithm is used for the lost detection rate and the error rate of transmitted packets [2]. Some other methods were also proposed for the Voice activity detection using wavelet transform [3], in which energy of the sub bands and feature extraction is done. In this process of estimation, the main concentration was only on the parameters, feature extraction and then getting output after fixed threshold.
To enhance and suppress the noise signal in the microphone is the area of research, which has gained much importance in the field of speech processing's techniques. Some of them are included as the, Standard Voice activity detection algorithms, Hybrid algorithm and PHAT methods.

3. Research Question and Problem Statement

The way of extracting the feature vectors of pre-defined signals and input signals, how to determine the voice activity at microphone using these feature vectors. This could be done by extracting feature vectors using Levinson-Durbin algorithm, and then these feature vectors of pre-defined speech and noise signals will be compared to the input signal to detect the speech activity at microphone.

4. Problem Solution

Robust Voice Activity Detection is a field that is receiving considerable attention because of its relationship, for example, speech recognition [4].

To detect the voice activity at microphone, firstly we need to extract the feature vectors of pre-defined speech and noise signals. In our case we took 4 signals each for speech and noise.

In real time processing it is very hard to take the whole signal to estimate the voice activity at microphone, for this we have to process short frames of the signal to get the quick response. After framing the signals we get 160 samples per frame. Then extracting 10 feature vectors (Γn) from each frame of pre-defined signals using Levinson-Durbin algorithm [5]:

$$\Gamma_i = {}_{ri}\text{-}j=1_i\text{-}1\alpha_{ji}\text{-}1(i\text{-}j)E_i\text{-}1 \qquad (1)$$

Where a_i is:

$$a_i i=\Gamma_i \qquad (2)$$

$$\alpha_{ji}=\alpha_{ji}\text{-}1\text{-}k_i\alpha_i\text{-}j_i\text{-}1 \qquad (3)$$

$$\varepsilon_i=1\text{-}\Gamma^2{}_i\varepsilon_i\text{-}1' \qquad (4)$$

Γ_i is the feature vectors, a_i is the prediction filter and ε_i is prediction error, r_n is the autocorrelation of the signal. Length of the feature vectors is dependent upon the order of r_n. As 10 feature vectors will be calculated for each frame, so the order of r_n should be 10. ε_i is the prediction error which will be minimized during the training of feature vectors

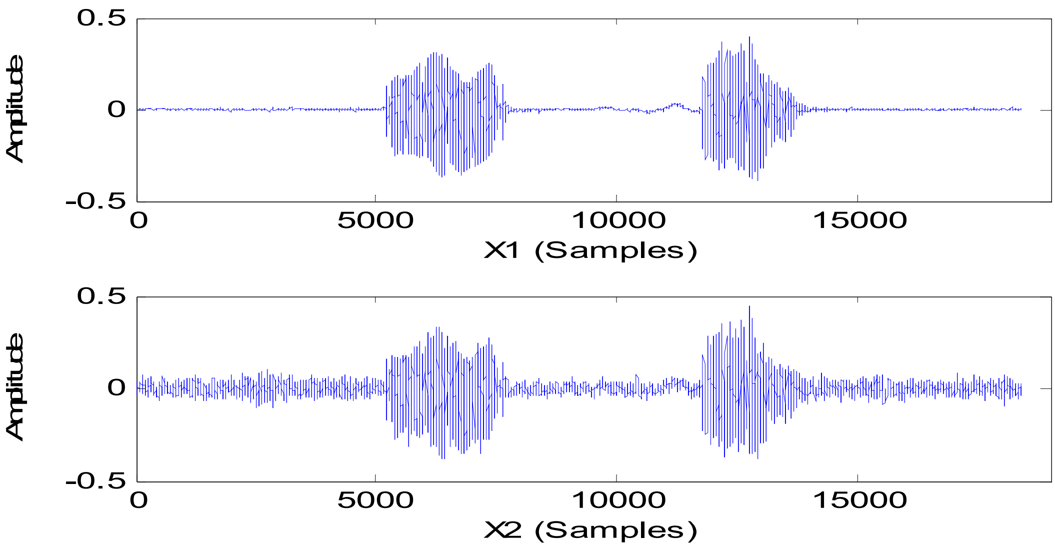

Fig. 1. X1 is the pure Input Signal from Microphone and X2 is the Distorted Signal.

These feature vectors now can be used to determine whether the signal coming from a microphone is a speech signal or a noise. For this we have to frame each incoming signal and extracting its feature vector and finding its minimum distance with the centriods of pre-defined speech and noise feature vectors. That minimum distance and centriods of pre-defined feature vectors can be determined using Euclidean distance and K-Means Algorithm [6], respectively.

$$E=argkmini=1K_j=1nx_{ij}-z_i^2 \qquad (5)$$

E is the minimum distance between the input signal's feature vector and centriods of pre-defined feature vector. x_{ij} is the feature vectors of input signal and zi is the centroid of the pre-defined feature vectors. We will calculate k=10 centroids each for both pre-defined speech and noise feature vectors. Minimum distance from the input signal's feature vector with centriods of pre-defined speech is ds and minimum distance from the input signal's feature vector with centriods of pre-defined noise signals will be dn.

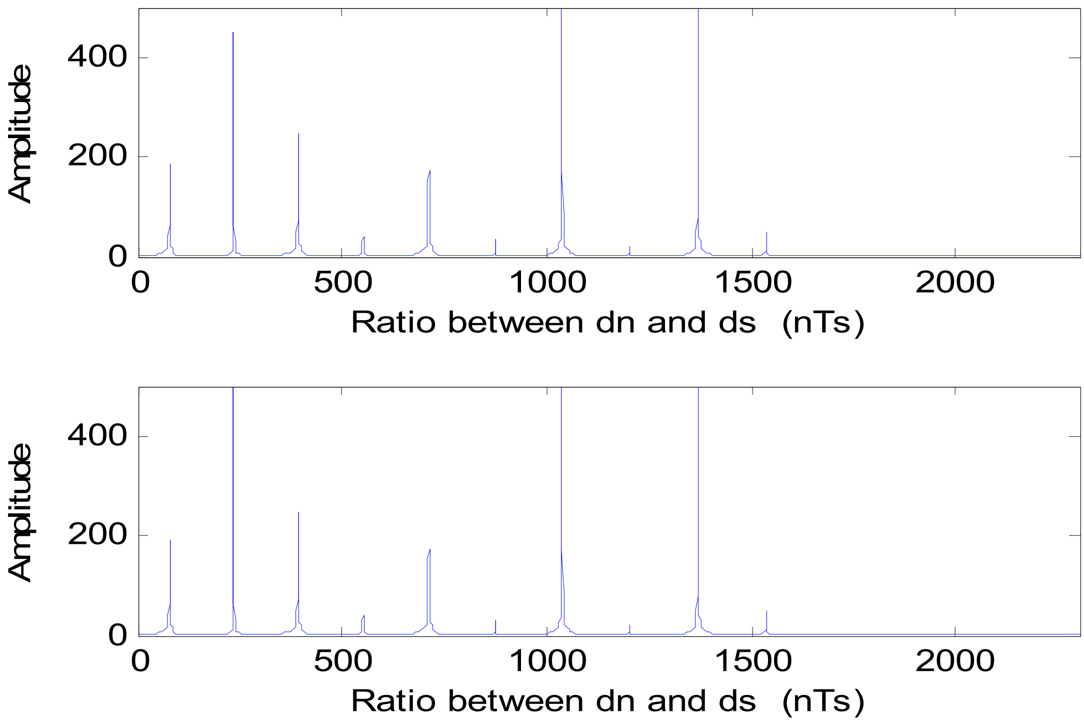

Fig. 2. The ratio corresponds to X1 & X2 input signals.

Ratio calculated from dn and ds is:

$$r=dnds \qquad (6)$$

This shows the peaks when there is a speech signal and valleys when there is noise. Fig2 shows the same ratio values for pure speech signal as well as the distorted speech signal. This shows that the algorithm is robust in noisy environment. So we can use this ratio to attenuate the noise in the incoming signal from the microphone.

5. Conclusions

This paper presented the voice activity detection using feature vectors. In the proposed model, we first framed the signal into blocks, and then extracted the feature vectors of pre-defined speech and noise signal's, which were already stored for processing. We used k-means algorithm to calculate the centroids of the pre-defined framed feature vectors. Levinson-Durbin algorithm is used to extract the feature vectors. The minimum distance between framed feature vectors of input signal and centroids of pre-defined signals are estimated using Euclidian distance. The ratio obtained from the Euclidean distances of noise and speech, represents the voice activity at the microphone. The evaluation of the ratio indicates the significance performance of voice activity detection in ideal and noisy environment.

This project paves the way to advancement in the field of Acoustic localization and Voice activity detection techniques, such as effective channel utilization. Moreover this technique has the capacity of working on the bit rate, which can be increased, thus making the channel more efficient.

References

[1] R. V. Prasad, A. Sangwan, H. S. Jamadagni, M. C. Chiranth, R. Sah, and V. Gaurav, "Comparison of voice activity detection algorithms for VoIP," in Proc. Int. Symp. On Computers and Communications, pp. 530–535, 2002.

[2] Jianping Zhang, Wayne Ward and Bryan Pellom, "Phone based voice activity detection using online Bayesian Adaptation with conjugate normal distributions, Centre for Spoken Language Research,2002,pp.0-7803-7402-9

[3] Kh.Aghajani, M.T.Manzuri, M.Karami, H.Tayebi, "A robust voice activity detection based on Wavelet Transform", In Second International Conference on Electrical Engineering, 25-26 March 2008, pp. 978-1-4244-2293-7

[4] A. Alvarez, P. Gomez, V. Nieto, R. Martinez, and V. Rodellar,"Application of a First-Order Differential Microphone for Efficient Voice Activity Detection in a Car Platform", Proc. Interspeech, pp. 2669-2672, 2005.

[5] R.Yu,X.Lin,C.C.Ko, "A Multi-Stage Levinson-Durbin Algorithm", Laboratories for Information technology,2002, pp.0-7803-7576-9

[6] Los Alamos Science, Vance Faber, "Clustering and the Continuous K-means Algorithm", Los Alamos Science, Number 22, 1994.