International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 07 | July 2018 www.irjet.net p-ISSN: 2395-0072

# Unwanted Voice Suppression by Using VAD Algorithm

**Divya T.R[1], R.M Nilajkar[2]**

[1]*M.Tech (D.E), Dept. of ECE, GMIT, Davangere, India*
[2]*Assistant Professor, Dept. of ECE, GMIT, Davangere, India*

---

**Abstract -** *As the interest for the limit of digital communications increases, the effective utilization of correspondence channel transmission capacity is the vital issues in digital communications increases. Transmission capacity (information) pressure in computerized coding of discourse can be accomplished utilizing two strategies. To start with strategy is effective source coding, and other is discourse addition. The discourse addition technique is used in two-way phone correspondence. A speech signal comprises of voiced, unvoiced and noiseless parts. The non-speech waveforms, the unimportant acoustic materials are approximately alluded to as silence. Likewise amid ordinary discussion through phone, every supporter represents just around 40% to half of the time (dynamic time). Aside from the dynamic time (sit channel) it comprises of tuning in, stops and intrusions amongst words and sentences. Information pressure can be accomplished by dispensing with these dead circumstances. Therefore the transmission connect activity limit can be expanded or multiplied by getting to different subscribers. The extent of this exposition work is to execute voice action recognition calculations utilizing the standard signal processing algorithm, for example, MATLAB with flag preparing tool stash. The exhibitions of a few such calculations should be dissected. The prerequisite is to recognize the fundamental highlights in these calculations which could be material for a present military task.*

***Key Words***: **Speech chain, VAD, zero crossing rate. Average energy etc….**

## 1. INTRODUCTION

The key motivation behind discourse is correspondence, i.e., the transmission of data. Demonstrates the total procedure of delivering and seeing discourse from the definition of a message in the cerebrum of a talker, to the making of the discourse flag, lastly to the comprehension of the message by an audience. In the exemplary prologue to discourse science, Danes and Pinson suitably alluded to this procedure as the "speech chain".

The wide varieties originating from the mouth of the speaker can be delegated voiced, unvoiced and silent portions. Additionally in an ordinary Telephone discussion, every supporter represents just around 40% to half of the time (dynamic station) the rest of the time (sit without moving station) comprises of tuning in, stops and intrusions amongst words and sentences. The interjection framework exploits this dead time and additionally the non-acoustic material (unvoiced and noiseless parts). Information pressure can be accomplished by wiping out these non-discourse materials from discourse. In this way the movement limit of the transmission connection can be expanded or multiplied by getting to alternate subscribers.

## 2. LITERATURE SURVEY

VAD calculations work by taking in a digitized audio signal, preparing this signal, extricating specific highlights from the handled signal, passing the removed highlights of the signal as parameters to a model depicts include in commotion and in speech, lastly yielding the choice in view of edges characterized in the model. There are a wide range of highlights of VAD calculations models are normally utilized and removed parameters like Fourier coefficients, periodicity and Zero intersection rates. So also there are different models that VAD calculations use to portray these highlights, some in light of heuristics while others in view of measurable models. Well known factual models incorporate Gaussian appropriations and Laplacian disseminations. In light of the choice administer characterized in the model, the VAD yields a banner to show the nearness or nonappearance of speech.

J. Sohn et al. (1999) The much referred to Sohn VAD in view of a factual model [2] has a fundamentally the same as way to deal with the Ying VAD [1], as far as its utilization of the vitality in a flag as its essential parameter for show correlation, and its utilization of Gaussian appropriations to demonstrate the dispersion of the discourse and non-discourse energies. The Sohn VAD however does not separate the flag into various recurrence groups, but rather considers the circulation of the whole range. The figuring uses the Decision Directed methodology to assess the from the prior SNR in the flag .A probability proportion is then registered utilizing the SNR in the present casing and the evaluated from the earlier SNR which is then contrasted with some limit dictated by the appropriation model to make the discourse/non-discourse choice. This calculation additionally executes a headache plan to keep the cut-out of feeble discourse tails, however instead of actualizing a straightforward postponement experiencing significant change from a discourse to a non-discourse pointer, the aftereffect conspire depends on a Hidden Markov Model

whereby the discourse choice of a present edge just relies upon the present edge and the past edge, making the connection between's sequential discourse outlines express.

A noteworthy recognizing variable of the Sohn VAD is the semi-regulated preparing of its Gaussian model. The commotion measurements are evaluated by accepting an underlying non-discourse district in a flag to prepare a loud model, which adds up to regulated learning, before consequent casings are then used to refresh the model in an unsupervised way. C. ITU G.729B The ITU G.729B VAD is a calculation utilized generally in correlations of various VAD calculations [3]. It utilizes four highlights as its parameters, the f,uguull and low-band outline energies, the arrangement of line otherworldly frequencies (which are Linear Predictive Coding (LPC) coefficients), and zero intersection rates.

D. ETSI AMR VAD There are two usage of the ETSI AMR VAD, alternative 1 (AMR 1) and choice 2 (AMR 2) [4]. The AMR 1 calculation works by first isolating the sound flag into various recurrence groups and after that distinguishing pitch and tone (varieties in pitch) nearness in the sub-groups as signs of discourse. An aftereffect conspire is added to represent low power endings of discourse blasts.

## 3. METHODOLOGY

This calculation is created to erase non-speech waveforms, the unimportant acoustic material inexactly alluded to as hush. Which utilizes two versatile abundancy limits and zero intersection rate to erase non-discourse material from speech.

Speech information for our work is digitized at a testing rate of 8000 Hz utilizing 8-bit PCM. The examples are fragmented utilizing rectangular window of size 100 examples/portion. For each fragment brief time normal energy and zero intersection rate (ZCR) are discovered utilizing the accompanying equations,

Average energy = ( 1/100 ) $\sum_{i=1}^{100} square[s(i)]$

Where s(i) is the i th speech sample

Zero crossing rate (ZCR) is given by,

ZCR = $\sum_{i=1}^{100} 1$ — [ s(i) . s(i+1) ] / [ s(i) . s(i+1) ]

Where s(i) is present sample

S(i+1) is the next sample

Energy level is turned out to be extremely valuable peacefully/discourse separation, especially when the limit to identify quiet to speech changes surpass the edge for discourse to hush advances and when the edges adjusted to neighborhood foundation commotion level advances. However vitality alone won't generally recognize discourse from quiet, we utilized moderately high ZCR estimations of powerless consonants to separate these from foundation clamor.

### 3.1 Algorithm Steps

Start with providing the mono channel 8 kHz .wav speech file as input to the algorithm. Then .wav file is read which provides the amplitude of samples. Average energy and average ZCR of the input file are determined. Total number of samples is divided in to several windows of 100 samples. For each window energy and zero crossing rate is determined. Then compare the energy and ZCR of the window with average energy and average ZCR. If energy of the window is greater than average energy and ZCR of the window is smaller than average ZCR, then the windowed signal is treated as voice or else treated as unvoiced. Perform this operation to every window of 100 samples until the total number of samples is over. The voiced and unvoiced part of the input file gets separated. Hence the compression achieves.

### 3.2 Flow Chart

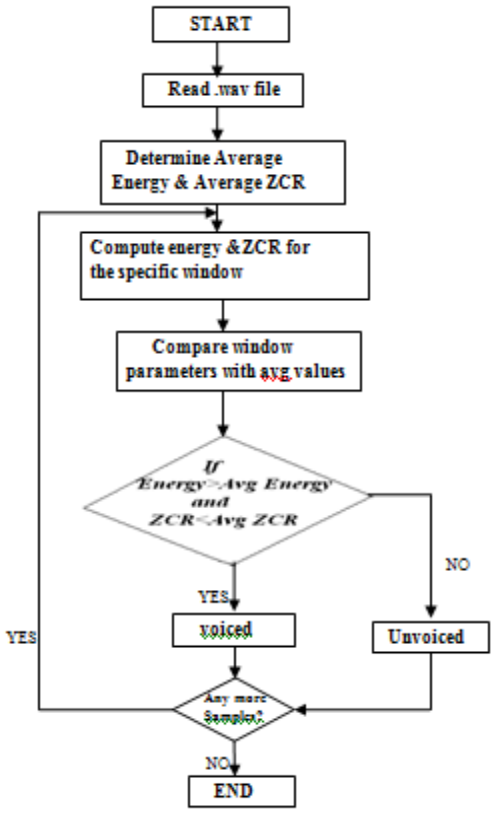


**Fig-1**.Flow chart of the algorithm

## 4. RESUT

A message '**this is the test message'** of 3 sec (24Kb) is taken for processing. This incorporates just information part, the header part is isolated. Subsequent to preparing the messages the information measured is packed and the portion in the header which speaks to 'information estimate' is altered and replayed.

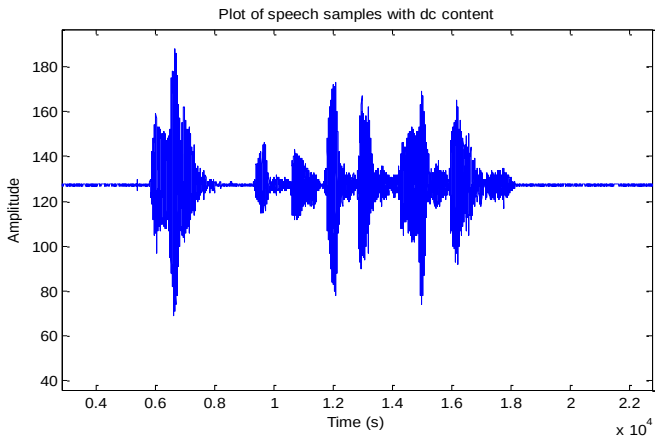


**Fig-2.** Speech signal with DC content

Results obtained when messages are processed by voice activity detection algorithm**.**

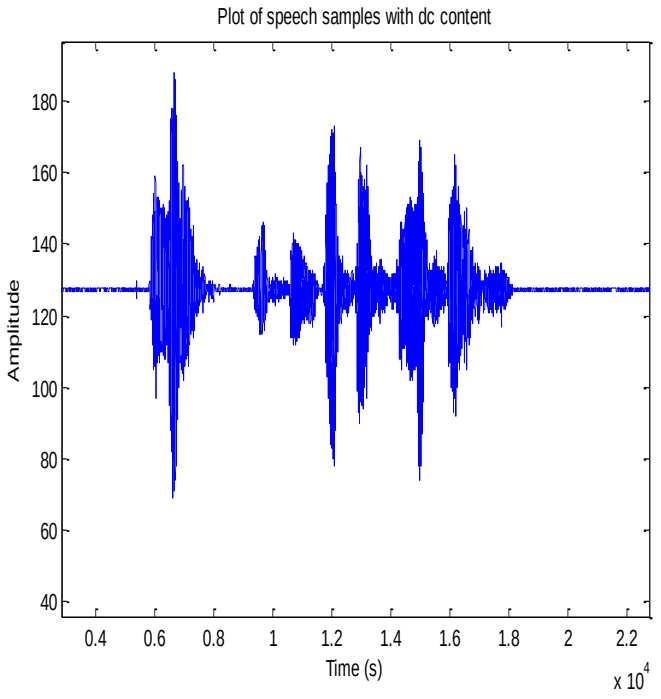


**Fig-3.** Speech signal with DC content

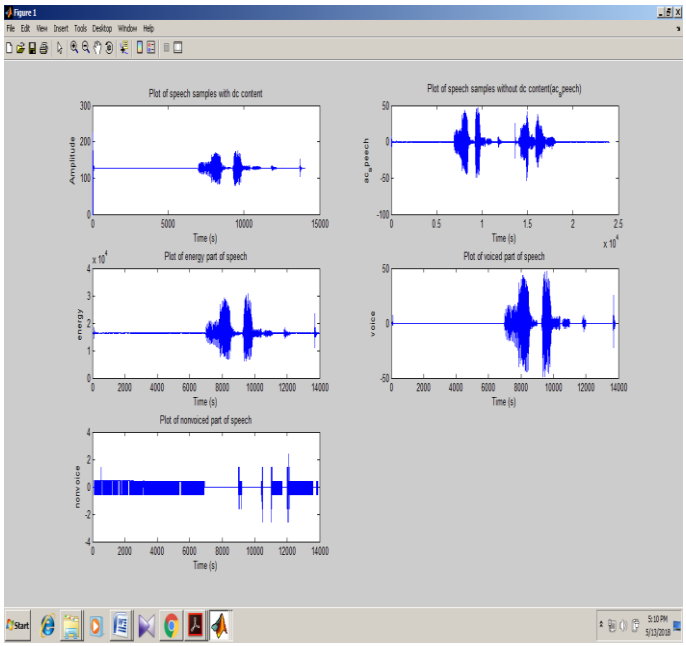


**Fig-4.** Extracted possible parameters

### 4.1 Result Analysis

| |
|---|
| Plot of speech samples with dc component……signal +noise =180(higher) |
| Comments: without processing of the samples… noise will be more (voiced part+ non voiced part+ silence part) |
| Plot of speech samples without dc component (ac content)…….only energy=48 |
| Comments: with processing of the sample noise will be less (removed non voiced part+ silence part) |
| Plot of energy part of speech……..28000 (higher energy value having and good frequency) |
| Plot of voiced part of speech = 48 |
| Plot of non voiced part of speech = 2 |
| Completely we cannot remove non voiced part 99% removed (remaining 1% negligible). |

## 5. CONCUSIONS

Discovery of voiced, unvoiced and hush segments of the discourse test is an essential part of programmed discourse acknowledgment .For the situation of ceaseless discourse acknowledgment it empowers long delays to be distinguished and erased preceding disentangling.

From the above procedure when the recorded speech is handled by calculation notwithstanding a great part of the quiet bit a few consonants are dispensed with. Along these lines it ends up hard to distinguish a few words which

contains more consonants extraordinarily words which start with consonants. Subsequently it influences the nature of speech. Be that as it may, the issue has not yet been enough unraveled, albeit numerous helpful outcomes and perceptions have been made.

The nature of the discourse can be enhanced by sending the data, for example, fragment numbers which are recognized as non-speech alongside the packed information and including the non-speech material at the best possible portion at the collector end.

In addition this calculation is prepared with recorded discourse information not at all like constant information.

## ACKNOWLEDGEMENT

I deeply thankful to Mr. R. M Nilajkar, Asst. professor, Department of Electronics and Communication Engineering, GM institute of technology, Davangere, for his boundless cooperation, constant support and valuable suggestions..

## REFERENCES

[1] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Transactions on acoustics, speech, and signal processing, Vol. assp-27, no. 2, April, 1979.

[2] ] U. Mahbub, T. Rahman and A. B. M. H. Rashid, "FPGA Implementation of Real Time Acoustic Noise Suppression by Spectral Subtraction using Dynamic Moving Average Method", IEEE Symposium on Industrial Electronics and Applications, Kuala Lumpur, Malaysia, October 4- 6, 2009.

[3] ] J. Whittington, K. Deo, T. Kleinschmidt and M. Mason, "FPGA Implementation of Spectral Subtraction for Automotive Speech Recognition", IEEE Transaction, 2009.

[4] F. Beritelli, "A Robust Voice Activity Detector for Wireless Communications Using Soft Computing", IEEE Transaction, 1998.

## BIOGRAPHIES

**Divya T.R** has completed her B.E.in E&CE discipline from Visvesvaraya Technological University at GMIT. She is currently pursuing M.Tech in Digital Electronics from Visvesvaraya Technological University at GMIT.

**Mr. R.M Nilajkar** has competed his B.E in ECE discipline at BVBCET Hubli and M.S in Electronics and Control at Pilani. He has published 2 journals and attended 1 international conference. He is currently working as a Assistant professor in Dept. of ECE at GMIT. Davangere, Karnataka.