AN EFFICIENT PEAK VALLEY DETECTION BASED VAD ALGORITHM FOR ROBUST DETECTION OF SPEECH AUDITORY BRAINSTEM RESPONSES

Ranganadh Narayanam

Department of Electronics & Communications Engineering
Bharat Institute of Engineering & Technology
Mangalpally (V), Ibrahimpatnam (M)
Hyderabad, A.P., India – 501 510
rnara100@uottawa.ca, rnara100@gmail.com

ABSTRACT

Voice Activity Detection (VAD) problem considers detecting the presence of speech in a noisy signal. The speech/non-speech classification task is not as trivial as it appears, and most of the VAD algorithms fail when the level of background noise increases. In this research we are presenting a new technique for Voice Activity Detection (VAD) in EEG collected brain stem speech evoked potentials data [7, 8, 9]. This one is spectral subtraction method in which we have developed our own mathematical formula for the peak valley detection (PVD) of the frequency spectra to detect the voice activity [1]. The purpose of this research is to compare the performance of this SNR based PVD (SNRPVD) method over Zero-Crossing rate detector [5] and statistical analysis based algorithms [10]. We have put into application of these three algorithms on these particular data sets of this experiment [7, 8, 9] and VAD is verified and compared the results of these three. MATLAB routines were developed on these particular methodologies. Finally we concluded that the method of SNRPVD surely performing better than the ZCR and statistical algorithms.

KEYWORDS

Zero Crossing Rate; Peak valley detection; Voice Activity Detection; stastical anlysis method.

1. INTRODUCTION

In this paper we are presenting three algorithms for the purpose of the Voice activity detection in EEG collected brain stem speech evoked potentials. In our experiment [7,8,9] EEG Eelectrode recordings are made from brain stem in response to complex, speech-like, sound stimuli. There is increasing interest in recording auditory brainstem responses to speech stimuli (speech ABR) as there is evidence that they are useful in the diagnosis of central auditory processing disorders [7]. However, the frequency content of natural speech is neither concentrated in frequency or in time,

Jan Zizka (Eds) : CCSIT, SIPP, AISC, PDCTA - 2013
pp. 315–321, 2013. © CS & IT-CSCP 2013 DOI : 10.5121/csit.2013.3636

the recording of speech ABR of sufficient quality may require tens of minutes [6]. Even with a synthetic consonant-vowel stimulus, a recording time of several minutes was required [8]. Speech ABR is believed to originate in neural activity that is phase-locked to the envelope or harmonics of the stimulus. As a result, the recorded responses are remarkably speech-like. In fact, speech ABR is quite intelligible if played back as a sound [9]. As a result, methods used for Voice Activity Detection (VAD) can be useful for the detection of speech ABR. Once the response is detected, then other noise suppression algorithms could in principle be applied to improve the signal-to-noise ratio (SNR).

We implemented three algorithms for the purpose of the VAD for our experiment of brain stem speech evoked potentials: (a) Linear-interpolation zero-crossing rate algorithm [5] explained in the section 2 A, specifically for our application. (b) A new proposed VAD algorithm that is based on a binary weighting of the spectral components of the signal under test [1]. This algorithm, explained in the section 2 B, is based on the property that vowels have distinctive spectral peaks. These are likely to remain higher than their surroundings even after severe corruption. Therefore, by developing a method of detecting the spectral peaks of vowel sounds in corrupted signal voice activity can be detected as well even in low signal-to-noise ratio (SNR) conditions. (c) Two more statistical algorithms [10] are also implemented, based on a statistical approach that has become the standard for detecting harmonic components in a related evoked response, the auditory steady-state response (ASSR).

We provided the results in the sections of 3 and 4. Finally we found the peak valley detection based SNRPVD algorithm perorming better than the remaining two.

2. ALGORITHMS

A. Zero Crossing Rate usage for the purpose of Voice activity detection[5]

First, in this paper we have implemented the Voice activity detection for the collected EEG brain stem speech evoked potentials using the “linear interpolation Zero crossing rate algorithm”. In this algorithm the shape of the signal is very close to the straight line near the zero crossing. Near the zero crossing the samples are described as points in straight line defined by the angular parameter a, and a linear parameter b. The two consecutive samples on the X-axis can be expressed as

$$x_n = a \times n + b$$
$$x_{n+1} = a \times (n + 1) + b \quad (1)$$

The parameters a and b can be expressed in terms of the samples.

$$a = x_{n+1} - x_n$$
$$b = (n+1) \times x_n - n \times x_{n+1} \quad (2)$$

The raising zero crossing must be in between two consecutive samples which must satisfy the condition.

$$x_n \leq 0 < x_{n+1} \quad (3)$$

Sample displacement for interpolation parameter d, is defined when interpolated sample n+d is given by

$$x_{n+d} = a \times (n + d) + b \quad (4)$$

By making the interpolated value equal to zero we can calculate the desired instant n+d of the zero crossing.

$$n + d = (b/-a) = \frac{b}{-a} = [(n \times x_{n+1})-(n+1)x_n]/[(x_{n+1})-(x_n)] \quad (5)$$

The displacement d is given by

$$d = \frac{-x_n}{(x_{n+1})-(x_n)} \quad (6)$$

d is a fractional number as of the equation 3, which is the zero crossing instant in sample numbers n+d. The samples x_n and x_{n+1} are trailing and leading samples from the zero cross instant.

B. Peak valley detection algorithm for the Voice activity detection: Signal to Noise ratio Peak valley detection ratio[1]

This method uses spectral peaks of vowel sounds to detect Voice activity in this particular experiment of EEG collected brain stem speech evoked potentials. Using this method we reduce the problem of detecting the voice activity to the problem of detecting the presence of vowels. In this the assumption is that the vowel sounds are nearly unique to speech. Why is this vowel sound detection is the reason that consonant sounds are always accompanied by vowel sounds and their duration is short so presence of vowel sounds can be directly related to the presence of speech. Vowel sounds are having distinctive spectral peaks of energy at specific spectral bands. Even though there is severe noise corruption the peaks remain higher than their surroundings.
We assume that the positions of major spectral peaks are the most important factor in recognizing the vowel sounds rather than the relative sizes of peaks or the shapes in spectral valleys, which are vulnerable to noise. Using this concept we propose the Signal to noise Ratio peak valley difference (SNRPVD) which calculates the similarity between the peak signature vector S of a registered vowel sound and the spectrum X of the input signal. In this by using one conventional existing peak valley detection formula [1] and applying it on several of our data sets and for our application we have concluded and modified the formula to this following formula.

S Vector: The peak signature vector S contains the peak position information for a vowel sound. It is a binary vector designed by us for this type of data collection. We already know that the locations of the places where these vowel sounds peaks occur i.e.> for example at 400 Hz, 500 Hz, 700 Hz, 800 Hz, 1000 Hz. So after we locate which frequencies of vowel peaks we need to select then we have to use the following formula for the detection of the locations of the frequencies in the given vector size of 1024 and for the frequency sampling rate of 3202 Hz. Then put “1” in those calculated locations and put zeros in the remaining locations which gives the “S”

$$A = \sum_{k=0}^{n-1}(X[k] \times S[k]) \quad (7)$$

$$B = \sum_{k=0}^{n-1} S[k] \quad (8)$$

$$C = \sum_{k=0}^{n-1}(X[k] \times (1 - S[k])) \quad (9)$$

$$D = \sum_{k=0}^{n-1}(1 - S[k]) \quad (10)$$

$$SNRPVD(X, S) = (A/B)/(C/D) \quad (11)$$

vector for that particular data set. Then we can apply the above SNRPVD formula for the voice activity detection. There are several ways to design this S vector depending on the application and the data collection.

S vector frequency location calculation formula.

Frequency location = [(Sampling Frequency / Number of Samples) (frequency for which we need to find the location)] + 2. (12)

3. RESULTS

In this section the results of the three algorithms are presented in the form of bar graphs which are useful for the analysis of the performance of the three algorithms. In this we observed the SNRPVD [1] algorithm is far better than the ZCR [5, 9] and Statistical analysis algorithm [10]. The results are discussed in the discussion section, which are given in the tabular form Table 1. We have 10 different subjects for analysis but we have presented here graphical results of one subject for example purposes for all the three algorithms in Fig 1,2,3,4.

Subject 2

Figure 1. ZCR algorithm.

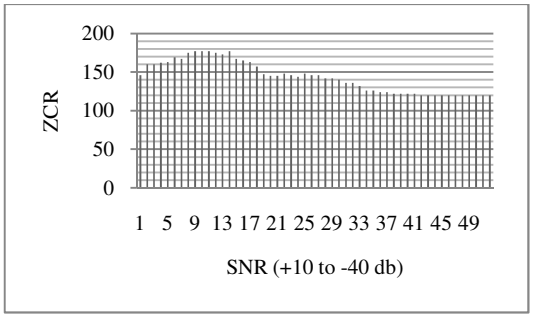


Figure 2. SNRPVD algorithm.

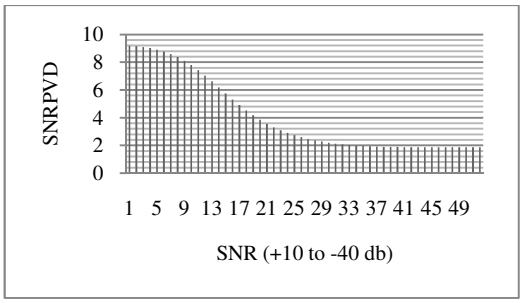


Figure 3. Statistical algorithm: for fundamental frequency of 100 Hz.

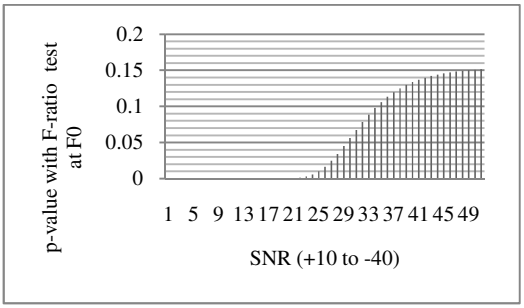
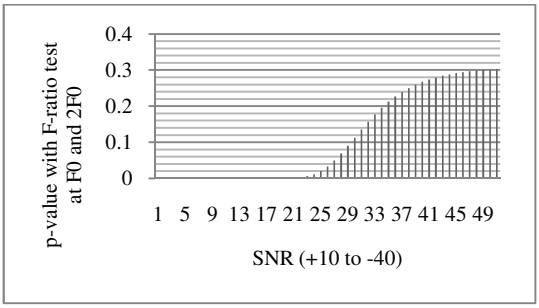


Figure 4. Statistical algorithm: For the frequency tone 100 Hz + 200 Hz.



For figures (1), (2), (3) & (4) on the X-axis SNR values in db from 10 dB to -40 dB in the steps of 1, and on the Y-axis ZCR, SNRPVD, Statistical algorithm: For fundamental frequency of 100 Hz, Statistical algorithm: For the frequency tone 100 Hz + 200 Hz respectively.

4. DISCUSSIONS

The results are given in the Tabular form Table 1 for all the three algorithms under evaluation. In this we have taken into consideration the EEG pure noise data which we have collected during the data collection. The ZCR [5, 8] and SNRPVD [1] values as reference for evaluation which are 120 and 1.8596 respectively for this EEG noise data. On these for precise evaluation purpose we have taken some additional percentage of 5% more on these values which are 126 and 1.95258 respectively to form as a threshold. So this is the 100% surety reference threshold from the ZCR

[8] and SNRPVD [1]. So for the statistical analysis purposes we will take as low p-value as possible for the reference threshold value to make it as close as possible for the 100% surety. In our case we have taken 0.01 is the threshold p-value we have taken. Form this we have 1% error i.e. 99% surety. For ZCR [8] and SNRPVD [1] values of all 10 subjects if the values are more than the taken threshold then it is considered as voice detection. For statistical analysis algorithm if it is less than the given threshold then it is considered as the voice detection. So in this case we started the experiment as adding the noise from 10 db to -40 db and also 1000db to the EEG pure noise signal and then adding this noised signal to the original data signal and then we did put into application this ZCR and SNRPVD and also statistical algorithmic procedures.

Table 1. The results for the three algorithms for the subjects 1-10.

Subject number	SNRPVD SNR cutoff (db)	ZCR SNR cutoff (db)	P-values Column 2 SNR cutoff (db)	P-values column 4 SNR cutoff (db)
1	-18	-2	nothing	-6
2	-26	-20	-14	-13
3	-32	1000	-19	-16
4	-23	-21	-12	-11
5	-24	-11	-15	-12
6	-26	-21	-12	-11
7	-23	-2	-16	-15
8	-36	-15	-13	-16
9	-30	-18	2	-15
10	-32	-11	-14	-12

(Note: column 2 is for: For fundamental frequency of 100 Hz. Column 4 is for: For the frequency tone 100 Hz + 200 Hz.)

5. CONCLUSION AND FURTHER RESEARCH

After As of the observed results from the table I it is clear that we can do better in the case of Signal to Noise Ratio Peak Valley Detection (SNRPVD) algorithm than Zero Cross Rating and Statistical analysis algorithms. As a further research part we are implementing few more VAD algorithms existing and about to compare them with our SNRPVD technique for further verification of its better performance over them, and would like to standardize it & then to apply in real time applications.

REFERENCES

[1] In-Chul Yoo and Dongsuk Yook, “Robust voice activity detection using the spectral spectral peaks of vowel sounds”. ETRI Journal,Volume 31,Number 4,August 2009.

[2] R. Nicole, J. Sohn, N.S. Kim, and W. Sung, “A statistical Model Based Voice Activity Detection,” IEEE Signal Process. Lett., vol. 6, 1999, pp. 1-3.

[3] Javier Ramirez,Jos C segura, Carmen Benitez,Angel de la torre, Antonio Rubio, “Efficient voice activity detection algorithms using long-term speech information”, J. Ram_ırez et al. / Speech Communication 42 (2004) 271–287.

[4] I. Krekule, “zero crosssing detection of the presence of evoked responses”, Electroencephalography and clinical neurophysiology, Elsevier publishing company, Amsterdam – Printed in the netherlands.

[5] GBrson Eduardo Mog, Eduardo Parente kbeiro, “Zero Crossing determination by linear interpolation of sampled sinusoidal signal.

[6] Dajani, R.H., Purcell, D., Wong, W., Kunov, H., Picton, T.W. 2005. Recording Human Evoked Potentials That Follow the Pitch Contour of a Natural Vowel. IEEE Transactions on Biomedical Engineering 52, 1614-1618.
[7] Johnson, K.L., Nicol, G.T., Kraus, N. 2005. Brain Stem Response to Speech: A Biological Marker of Auditory Processing. Ear & Hearing 26, 424-434.
[8] Russo, N., Nicol, T., Musacchia, G., Kraus, N. 2004. Brainstem responses to speech syllables. Clinical Neurophysiology 115, 2021-2030.
[9] Galbraith GC, Arbagey PW, Branski R. Intelligible speech encoded in the human brain stem frequency-following response. NeuroReport 1995; 6: 2363-2367.
[10] M.S. John, T.W. Picton, MASTER: a Windows program for recording multipleauditory steady-state response. Computer Methods and Programs in Biomedicine 61 (2000) 125–150, Elsevier.

Author's Biography

Mr. Ranganadh Narayanam is an Assistant professor in the department of Electronics & Communications Engineering in Bharat Institute of Engineering & Technology(BIET).This research is continuation of the research done in univeristy of ottawa under the guidance of Dr. Hilmi Dajani. This current research was partly funded by BIET. Mr.Narayanam, a research student in the area of "Brain Stem Speech Evoked Potentials" under the guidance of Dr. Hilmi Dajnai of University of Ottawa,Canada. He was also a research student in The University of Texas at San Antonio under Dr. Parimal A Patel,Dr. Artyom M. Grigoryan, Dr Sos Again, Dr. CJ Qian,. in the areas of signal processing and digital systems, control systems. He worked in the area of Brian Imaging in University of California Berkeley. Mr. Narayanam is having around 5+2 years of full time teaching & research experiences respectively, and more than 5 years of entry level research experience and more than 10 publications. Mr. Narayanam's research interests include neurological Signal & Image processing, DSP software & Hardware design and implementations, neurotechnologies.