Effect of Wavelet Entropy on Speech and Non-speech Differentiation

[1]Win Mar Shwe, [2]Ma Khin Saw, [3]Kyaw Soe Lwin

***Abstract*— This paper presents the detection of the speech and non-speech by using wavelet entropy research methodology. Voice activity detection (VAD) is utilized a decision algorithm. The different noise levels of car, street, train, babble and white noise are produced and extracted speech signal according to the wave files from the first to the fifth level respectively. The implementation of the algorithm development for VAD and all the data analyses are utilized MATLAB .wav format with 8 kHz sampling frequency in mono channel. All the human speech signals are encoded with wave resolution of 16 bits. Comparisons have been done using various noise signals.**

***Index Terms*— Voice Activity Detection, wavelet, signal, Noise, MATLAB, algorithm.**

1) INTRODUCTION

This paper intends to detect between the speech and non-speech in the noisy environment. Speech signals take a very special place amongst all other audio signals. Speech information can be not only generated but also carried. Even in this modern age, much of the information can be received in the form of speech. And most other audio signals that do not be carry any information as such. Indeed, much would be classed as noise in everyday situations.

The term speech detection is often used interchangeably with the term Voice Activity Detection (VAD), even though voice activity may of course be a variety of things other than strictly speech. In any case, the majority of human voice activity could be called speech and the two terms are also used interchangeably in this paper. VAD is also used to refer to any algorithm or system that is designed to detect speech, and then stands for Voice Activity Detector. It can be viewed as a binary classifier that label search frame of a speech signal as speech or non-speech respectively. The structure of a VAD is often very complicated and tweaking of VAD-parameters might be required for optimal performance.

Voice activity detection (VAD) is important in many areas of voice signal processing such as voice coding, voice recognition, voice enhancement etc. An effective voice activity detection (VAD) algorithm is proposed for improving voice quality performance in noisy environments.

Manuscript received January, 2019.

***Win Mar Shwe**, Electronic Engineering Department, Technological University (Sittwe), Sittwe, Rakhine State, Myanmar, Mobile No.+959421711402.*

***Ma Khin Saw**, Electronic Engineering Department, Technological University (Sittwe), Sittwe, Rakhine State, Myanmar, Mobile No.+9594217113872.*

***Kyaw Soe Lwin**, Electronic Engineering Department, Technological University (Sittwe), Sittwe, Rakhine State, Myanmar, Mobile No.+9595062554.*

Speech and non-speech detection is an unsolved problem in speech processing and affects numerous applications including robust speech recognition, discontinuous transmission, and real-time speech transmission on the Internet or combined noise reduction and echo cancellation schemes in the context of telephony. The speech and 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. The approach taken in this work for the development of VAD systems is based on the discrete wavelet entropy calculating on critical bands. Such systems are able to give the presence of speech, and do not need to learn to do this through training on audio-signal examples.

This paper is based on the MATLAB software program to detect the different noise levels through Voice Activity Detection (VAD). The software development of speech and non-speech detection is based on the MATLAB. The signal to noise ratios is used to create on the variable environments. This paper is mainly focused on a speech database for testing the robustness issue on the detection of speech and non-speech. The implementation of the algorithm development for VAD is used MATLAB programming as well.

2) VOICE ACTIVITY DETECTION

Voice activity detection (VAD) in noisy environment is an important process in many speech signal processing algorithms.

Nitin N Lokhande,2011 Pravara Rural Engineering College, Loni Ahmednagar-Maharashtra said to determine the beginning and the termination of speech in the presence of background noise is a complicated problem and presented to concern with labeling sections of speech samples based on whether they are silence, voiced or unvoiced speech. The labeling is done using calculations over the speech samples; zero crossing and short-term energy functions. The short-term energy and zero crossing rate of speech have been extensively used to detect the endpoints of an utterance [1].

Ji Wu and Xiao-Lei Zhang 2011 presented a new voice activity detection (VAD) algorithm that is based on statistical models and empirical rule-based energy detection algorithm. Specifically, two steps are needed to separate speech segments from background noise. For the first step, the VAD detects possible speech endpoints efficiently using the empirical rule based energy detection algorithm. However, the possible endpoints are not accurate enough when the signal-to-noise ratio is low. Therefore, for the second step, a new Gaussian mixture model-based is proposed to the multiple observation log likelihood ratio algorithm to align the endpoints to their optimal positions. Several experiments

All Rights Reserved © 2019 IJSETR

are conducted to evaluate the proposed VAD on both accuracy and efficiency [2].

3) Proposed System

Five main parts of the proposed system flow are preprocessing, framing and decomposition of the wavelet bands, spectral entropy calculation and VAD decision scheme in Figure1. This method deals with detection of the speech signals in noisy background. In this preprocessing method, the entire incoming signal should be MATLAB readable .wav format with 8 kHz sampling frequency in mono channel. All the human speech signals are encoded with wave resolution of 16 bits. When the incoming speech files have sampling frequency which is not 8 kHz, it will resample the original sampling rate into desired rate.

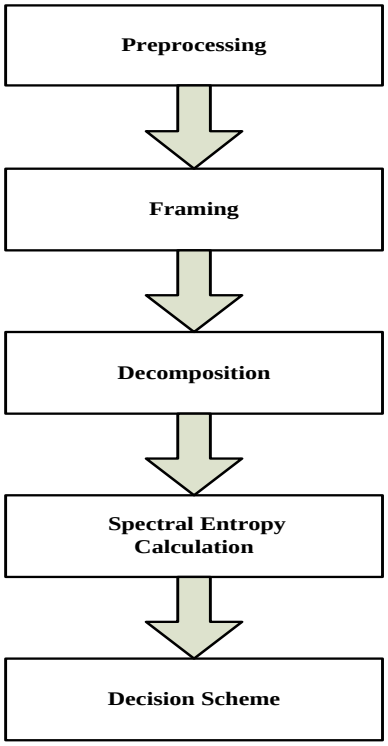

Figure1: Proposed System

This VAD method analyzes the signal characteristics with frame by frame basis. All VAD decision is made upon 20 ms blocks. The speech is partitioned into frames when the signal to be detected is fed into the Voice Activity Decision (VAD) algorithm.

It has been shown that frequency components of sound can be integrated into critical bands that refer to bandwidths at which subjective become significantly different. The frequency will be decomposed using the wavelet decomposition tree diagram.

Band	Low	High	CBW
1	0	250	250
2	250	500	250
3	500	750	250
4	750	1000	250
5	1000	1250	250
6	1250	1500	250
7	1500	1750	250
8	1750	2000	250
9	2000	2500	500
10	2500	3000	500
11	3000	3500	500
12	3500	4000	500
13	4000	4500	500
14	4500	5000	500
15	5000	6000	1000
16	6000	7000	1000
17	7000	8000	1000

Table 1: Decomposition of the Critical Band Frequency Scale

The speech file is designed to match the auditory critical bands as close as possible and it has been applied in various speech processing systems. Therefore, the speech signals are analyzed with critical bands with the frequency ranges as shown in Table 1.

All the frames have to be decomposed into 17 critical bands using discrete wavelet transform. The corresponding wavelet decomposition tree structure is shown in Figure2 which contains 16 basic wavelet decomposition cells and five decomposition levels. Each decomposition cell can be implemented via the filter bank approaches in MATLAB environment.

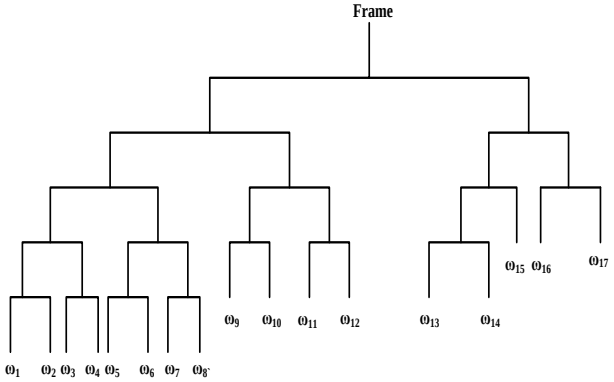

Figure 2: The Tree Structure of the Decomposition Level

Shannon entropy of each frame is computed over decomposed wavelet coefficients. Shannon spectral entropy is defined as:

$$H(m) = -\sum_{n=1}^{17} P(n)\log P(n)$$

Where:
m = frame index,
n = band number
P(n)= Probability density function of the wavelet coefficient

4) Simulation and Experimental Results

Based on the information of the female speech test wave file, sampling frequency is 8000 Hz, number of samples is 20839 samples, number of frame is 130 and duration is 2.6s.

It is common to test the effectiveness of voice activity detection with speech signals in noisy environments. The speech wave file is tested in five noisy backgrounds. The noisy are artificially added at specified SNR. Five considered noise types are car noise, street noise, train noise, babble noise and white noise.

All Rights Reserved © 2019 IJSETR

The clean speech signal in car noise at 20 dB SNR and 5 dB SNR displayed in Figure 3.

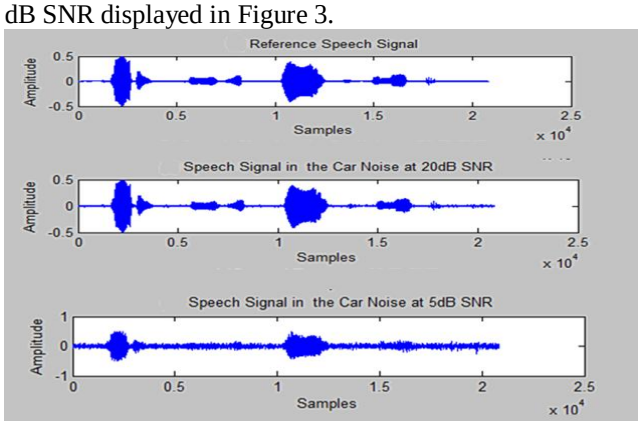

Figure 3. Comparison between the car noise for 20 dB SNR and 5 dB SNR

Similarly, the speech signal in street noise with the SNRs: 20 dB and 5 dB in Figure 3.

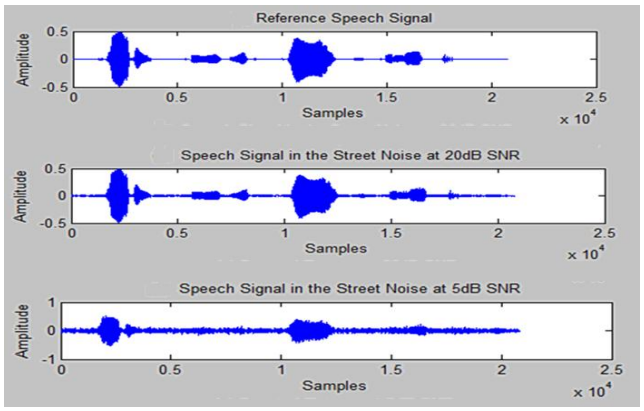

Figure 4. Comparison between the car noise for 20 dB SNR and 5 dB SNR

The wave file is coupled with the train noise at 20 dB SNR and 5 dB SNR.

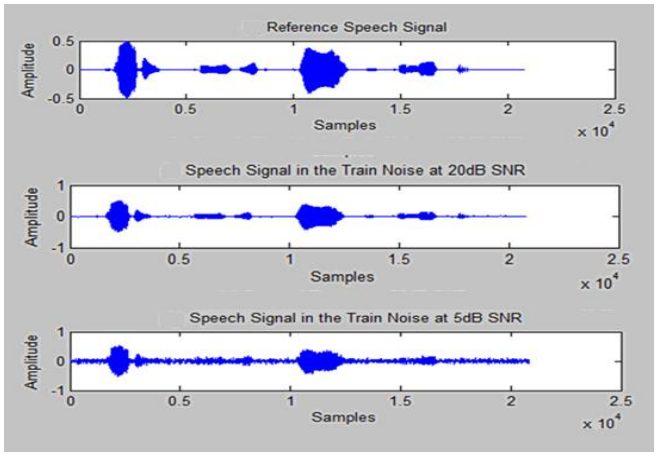

Figure 5. Comparison between the train noise for 20 dB SNR and 5 dB SNR

The babble noise is tested at 20 dB SNR and 5 dB SNR. In Figure 2 to Figure 6, at 20 dB SNR, the speech signal changes a little in five noisy environments when compared to the reference signal. And then, at 5 dB SNR, the speech signal alters to the noisy speech signal. The spectral entropy of the speech signal at 20 dB and 5 dB SNRs in the five noises in Figure 8 to Figure 12.

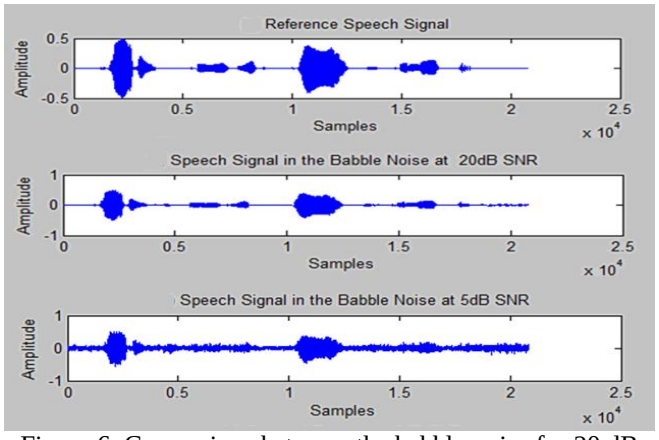

Figure 6. Comparison between the babble noise for 20 dB SNR and 5 dB SNR

The speech signal 20 dB and 5 dB SNRs are in the last noise, white noise.

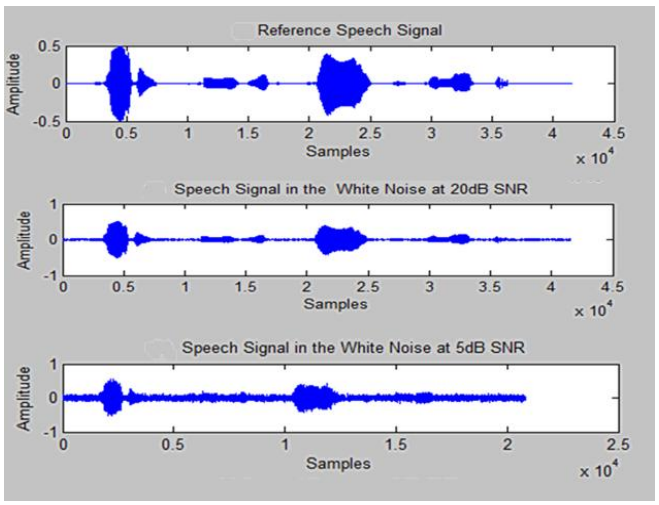

Figure 7. Comparison between the white noise for 20 dB SNR and 5 dB SNR

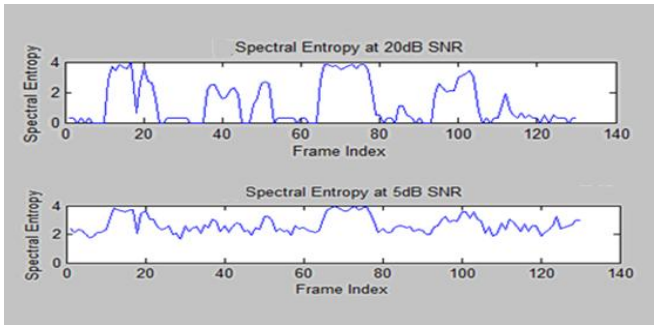

Figure 8. Spectral Entropy Comparison in car noise at 20 dB and 5 dB SNRs

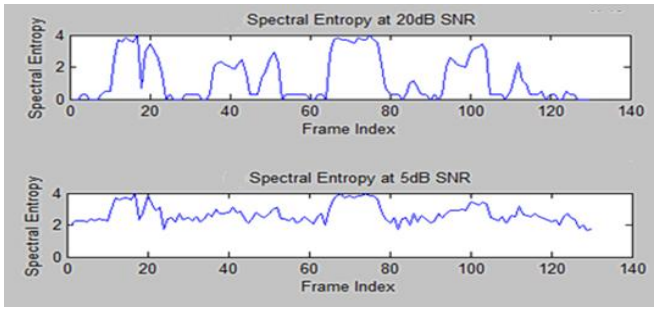

Figure 9. Spectral Entropy Comparison in street noise at 20 dB and 5 dB SNRs

All Rights Reserved © 2019 IJSETR

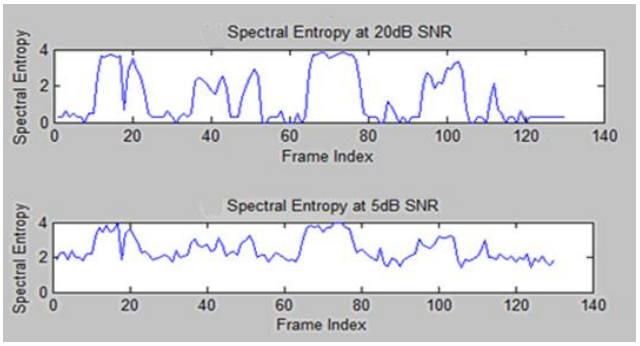

Figure 10. Spectral Entropy Comparison in train noise at 20 dB and 5 dB SNRs

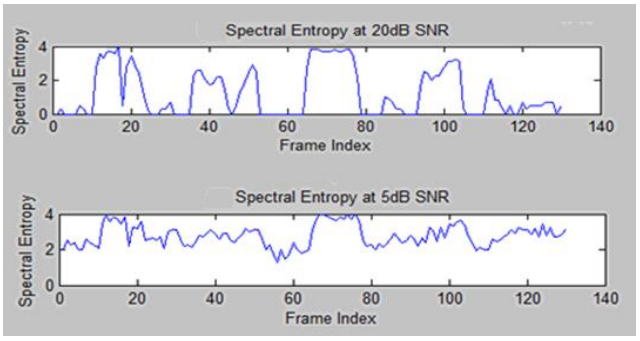

Figure 11. Spectral Entropy Comparison in babble noise at 20 dB and 5 dB SNRs

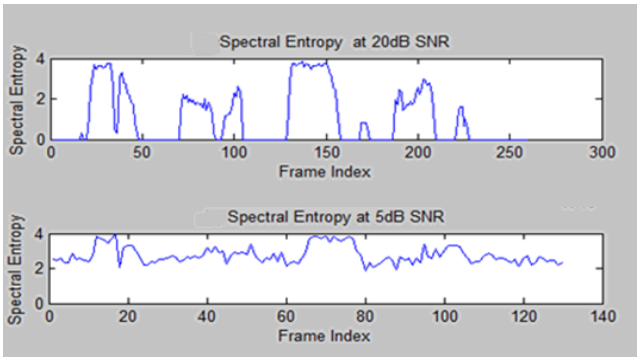

Figure 12. Spectral Entropy Comparison in white noise at 20 dB and 5 dB SNRs

The experimental results are shown in Figure 3 to 7 inferred that the VAD algorithm produces result for certain noises; car noise, street noise, train noise, babble noise and white noise. The spectral entropy results are shown in Figure 8 to 12 .The spectral entropy of the speech signal at 20 dB SNR is the same original speech signal. The speech signal looks like the original speech signal when the signal to noise ratio is greater than 20 dB SNR. The more small the signal to noise ratios, the noisier the speech signal. Thus speech and non-speech detection depend on the 20 dB SNR.

5) Conclusion

Based on the spectral entropy VAD algorithm, a relatively accurate speech and non-speech classification can be obtained at low computational cost. Speech and non-speech detection is produced correctly and the running time is very fast. Thus this paper tends to the VAD algorithm works well and can extract speech signal at different noisy environments.

Acknowledgment

First of all, the authors would like to express their special thanks to Dr. Myo Thein Gyi, Union Minister of Education, the Republic of the Union of Myanmar, for his encouragement to do research works for regional development and applied science. The authors' grateful thanks go to Dr. Kyaw Hlaing Oo, Acting Pro-Rector and Principal of Technological University (Sittwe) for his kind guidance, suggestion and directions throughout the preparation of research work, and valued motivations and encouragement as well. Finally, the authors are also sincerely thankful to our colleagues from the Department of Electronic Engineering, Technological University (Sittwe) offered strong moral and physical support.

REFERENCES

[1] Nitin N Lokhande, Navnth S Nehe and Pratap S Vikhe , "International Conference in Computational Intelligence (ICICA)" (2011)
[2]Ji Wu* and Xiao Lu Zhang , "Wu and Zhang EURASIP Journal on advances in Signal Processing 2011, 2018,8 http:// asp.eurasip journals.com / 2011/1/18
[3]Prasad, R.V., Sangwan , A., Janmadagni, H. S. and Chiranth , M. C.: "Comparison of voice activity detection algorithms for VOIP, in Proc. IEEE Symposium on Computer and Communication", vol.5, 2002
[4]Gold, B. and Morgan, N.: "Speech and Audio Signal Processing ", New York: John Wiley and Sons, (2000)
[5]Marzinzik, M. and Kollmeier, B.: "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. IEEE Transactions on Speech and Audio Processing", 10 (2): ,(2002)

Win Mar Shwe has received her B.E. degree in Electronic Engineering (in 2005) from Technological University (Sittwe), M.E. degree in Electronic Engineering (in 2010) from West Yangon Technological University. She is dedicated to teaching field from the last 13 years. She has supervised four under graduate students. Her research areas are digital signal processing, electronic circuit design and Communication Circuit Design. At present she is working as lecturer of Electronic Engineering Dept. at Technological University (Sittwe), Rakhine State, Myanmar.

Ma Khin Saw has received her B.E. degree in Electronic Engineering (in 2005) from Technological University (Sittwe), M.E. degree in Electronic Engineering (in 2010) from West Yangon Technological University. She is dedicated to teaching field from the last 13 years. She has supervised four under graduate students. Her research areas are electronic circuit design and Communication Circuit Design. At present she is working as lecturer of Electronic Engineering Dept. at Technological University (Sittwe), Rakhine State, Myanmar.

Kyaw Soe Lwin has received his B.E. degree in Electronic Engineering (in 2002), M.E. degree in Electronic Engineering (in 2004) from Yangon Technological University and Ph.D. degree in Electronic Engineering (in 2009) from Mandalay Technological University. He is dedicated to teaching field from the last 16 years. He has supervised 25 M.E students, guided 5 Ph.D. students and 50 under graduate students. His research areas are power electronics, electronic circuit design (including VHDL) and RF and Microwave Circuit Design. At present he is working as Professor and Head of Electronic Engineering Dept. at Technological University (Sittwe), Rakhine State, Myanmar.

All Rights Reserved © 2019 IJSETR