Scientific Research and Essays Vol. 6(1), pp. 18-22, 4 January, 2011
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE09.600
ISSN 1992-2248 ©2011 Academic Journals

Full Length Research Paper

Real time noise-speech discrimination in time domain for speech recognition application

N. Mokhtar*, H. Arof, F. R. Mahamd Adikan and M. Mubin

Department of Electrical Engineering, Faculty of Engineering, University of Malaya,
50603 Lembah Pantai, Kuala Lumpur, Malaysia.

Accepted 16 December, 2010

A simple noise-speech discrimination method in time domain is presented. The random signal noise characteristics were studied in time domain. Using the characteristics information, in real time processing, a simple algorithm to detect starting and ending point of speech samples in time domain is being demonstrated. Out of 100 attempts, about 94% of successful attempts of noise-speech discrimination have been obtained with white noise background by choosing the critical thresholds values.

Key words: Noise-speech discrimination, noisy signal, local mean.

INTRODUCTION

In successful automatic speech recognition systems, it is essential to detect the starting and ending point samples of test and reference patterns (Junqua et al., 1994, Rabiner and Juang, 1993). In order to perform the discrimination of noise-speech in real time processing, important criteria that need to be considered are simplicity and reliability. In this case, simplicity is to minimize unnecessary operation to enable the performance of real time analysis in time domain. Reliability is the ability to detect the beginning and ending point of speech samples in time domain with high accuracy.

Approach used by Junqua et al. (1994) and Rahmani et al. (2009), for speech-noise discrimination is done in the frequency domain by selecting the energy in the frequency band 250 – 3500 Hz. Cohen obtained the noise estimation by averaging past spectral power recursively using time-varying frequency-dependent to differentiate speech absence and speech presence (Israel, 2003). Voice activity detection by Tanyer combined two methods which also used energy threshold and overlapping frames of 16 ms as vital information to differentiate the presence of speech and noise (Gökhun and Őzer, 2000). Online noise estimation by Zhao has also been conducted in the frequency domain and using overlapping frames of 16 ms (Zhao et al., 2008).

In this work, real time noise-speech discrimination is successfully demonstrated in time domain without using any overlapping frames. The noise characteristics under noisy environment such as amplitude and mean of the amplitudes are studied for thresholds values determination. Based on the noise characteristics, an algorithm to perform noise-speech discrimination in starting and ending point of speech samples is proposed. Speech samples with background noise are demonstrated. The extracted speech samples are discussed in the experiments and results section.

Signal and noise models

Total input signal from unidirectional microphone can be defined as:

$$s(t) = y(t) + w(t) \qquad (1)$$

where *y(t)* is the clean speech samples and *w(t)* is the background noise. Background noise can consist of white, pink, blue, red and other types of noises. White noise is defined to be a stationary random process having a constant spectral density (Zhao et al., 2008). In this work, background noise *w(t)* is white noise which has uniform amplitudes over 10000 samples per second. Statistically, white noise has zero mean value over samples taken in a frame (Brown, 1983).

*Corresponding author. E-mail: norrimamokhtar@um.edu.my. Tel: +603-79676806. Fax: +603-79675316.

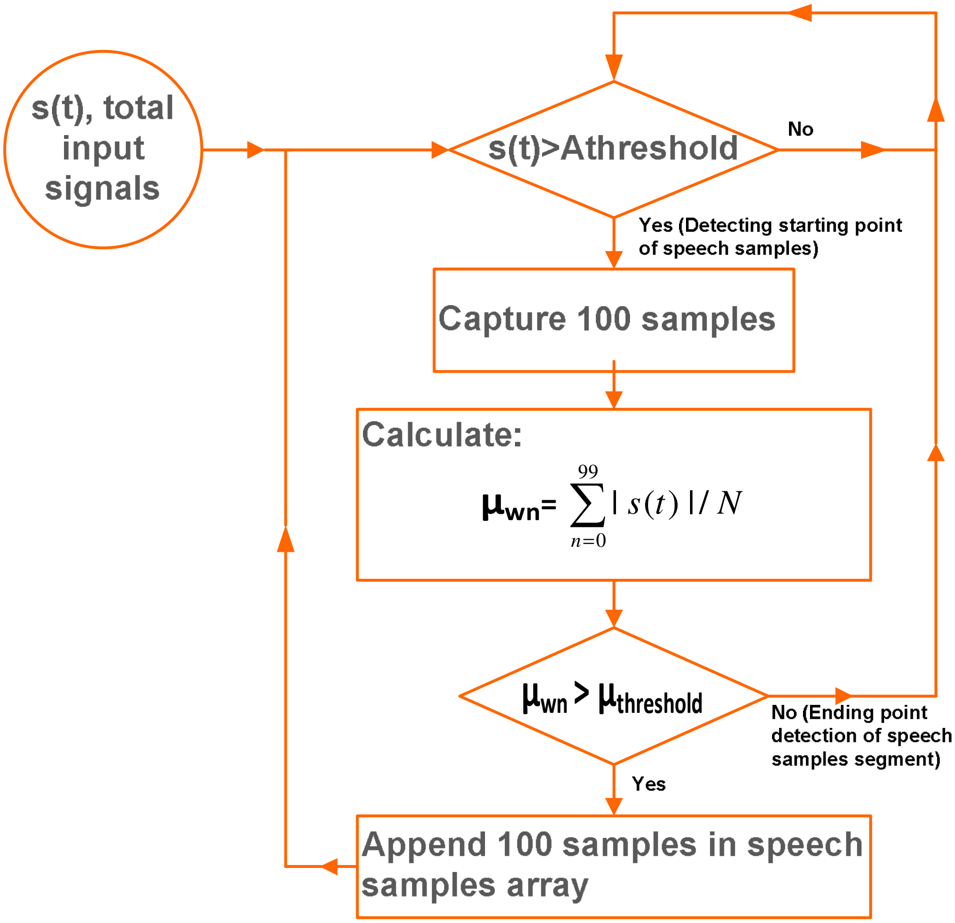

s(t), total input signals
s(t)>Athreshold
No
Yes (Detecting starting point of speech samples)
Capture 100 samples
Calculate:
$\mu_{wn} = \sum_{n=0}^{99} | s(t) | / N$
$\mu_{wn} > \mu_{threshold}$
No (Ending point detection of speech samples segment)
Yes
Append 100 samples in speech samples array

Figure 1. Proposed algorithm for noise-speech processing, and starting and ending point detection of speech samples.

$$\mu_{wn} = \sum_{n=0}^{99999} s(t) / N \ = 0 \qquad (2)$$

where μ_{wn} is the mean for white noise, N is number of samples per second, which was 10,000 samples per second used in this experiment.

The proposed method utilizes the white noise characteristics in order to differentiate between noise-speech signals and capture the speech samples segment. However, characterization of white noise does not provide sufficient information. Therefore, an algorithm and simple statistical method are proposed, which is shown in Figure 1.

METHODOLOGY

Basically, from Figure 2, white noise characteristics is clearly illustrated which has uniform amplitude across a 1 s frame. By using this characteristic, the threshold (amplitude threshold) is set to 0.1. If the amplitude is greater than the threshold value, the program will trigger the starting point of speech samples. The threshold value is obtained under noisy environment with air-condition and radio background.

Theoretically, the mean for all samples in a frame is zero for white noise. Equation 2 is modified to Equation 3. It was used as a trigger to detect ending point detection in speech samples.

$$\mu_{wn} = \sum_{n=0}^{99} | s(t) | / N \qquad (N=100) \qquad (3)$$

By taking the mean of every 100 samples locally, the ending point of speech samples can be determined correctly in time the domain. $\mu_{threshold}$ is set to 0.05, which was obtained under testing of various noise conditions such as air-condition and radio background.

After the noise-speech discrimination and starting point of speech samples are successfully done, the 100 samples that successfully pass the thresholds, will be saved in memory and the next 100 samples will be appended together with the previous 100 samples of speech signals until ending point of speech samples is detected. Flowchart of the algorithm is shown in Figure 1, which described the whole process involved.

EXPERIMENTS AND RESULTS

Experimental setup, test conditions and software information are shown in Table 1. Figure 2 demonstrates the background noise with two conditions. It was clearly noticed that background noises with air-condition and

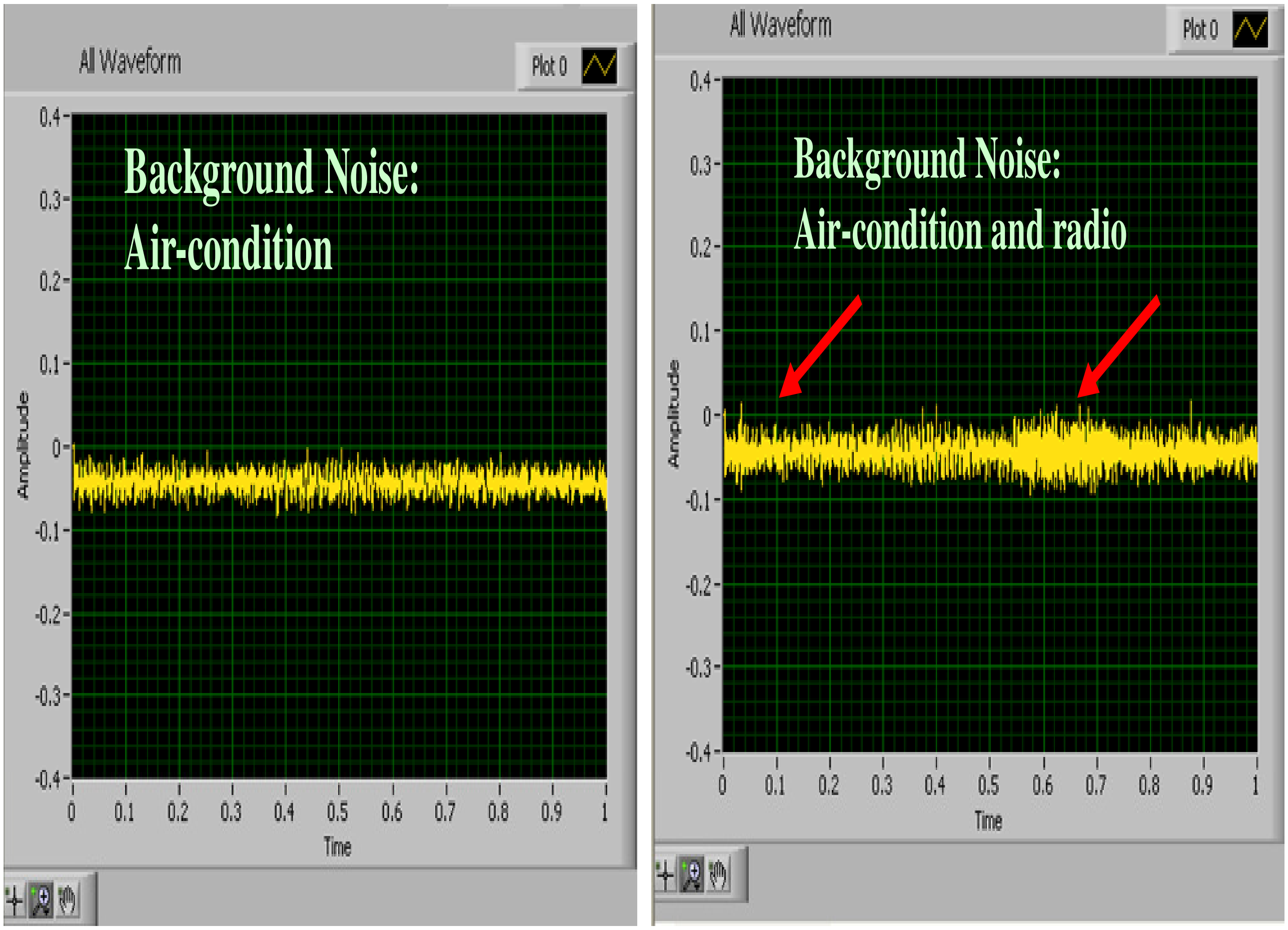


Figure 2. Observation of white noise with air-condition noise and radio background.

Table 1. Input setup, test conditions and software information.

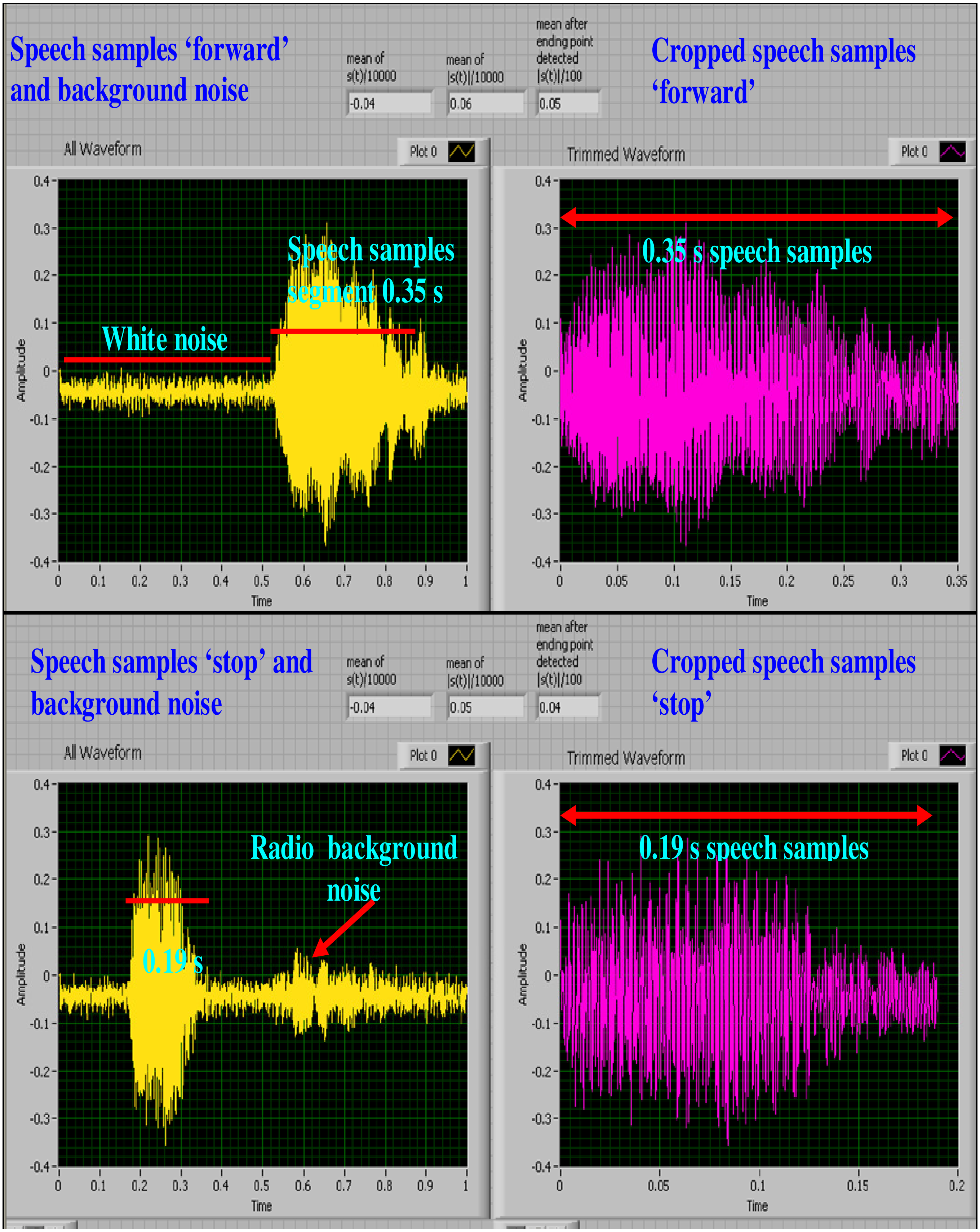
Input setup	Test conditions	Software
Unidirectional microphone	Silence with air-condition noise	Labview version 8.2 by National Instruments
One channel, resolution: 16 bits Sampling rate: 10KHz	Silence with air-condition noise and slow radio background	

radio background have about 15 - 20% higher amplitude as compared to background noise with only air-condition contribution.

From Figure 3, examples of speech samples and background noises were illustrated. Two types of speech samples were tested which are ‘forward’ and ‘stop’. ‘Forward’ speech samples have duration of 0.35 s which was extracted correctly in the trimmed graph from the all waveform graph. ‘Stop’ speech samples have duration of 0.19 s which was also extracted correctly in the trimmed graph. Mean for all samples was -0.04. By taking the mean of modulus *s(t)* locally for every 100 samples, it was demonstrated that the last μ_{wn} after the ending point of speech samples detected was 0.05 for ‘forward’ samples and 0.04 for ‘stop’ samples. These values are the triggered value set by the algorithm to detect the ending point of the speech samples.

Conclusion

Most speech processing techniques involving noise-speech discrimination were done in frequency domain and they used overlapping frames. In this work, although the algorithm was simple, it has been successfully demonstrated that this process can be done in time domain without using overlapping frames in order to save processing time and enable real time speech processing.



mean of s(t)/10000 -0.04

mean of |s(t)|/10000 0.06

mean after ending point detected |s(t)|/100 0.05

mean of s(t)/10000 -0.04

mean of |s(t)|/10000 0.05

mean after ending point detected |s(t)|/100 0.04

Figure 3. Examples of speech samples with air-condition and radio background.

Auto-threshold and speech recognition are future project of this work.

REFERENCES

Junqua J-C, Mak B, Reaves B (1994). A Robust Algorithm for Word Boundary Detection in the Presence of Noise. IEEE Trans. Speech Audio Processing, 2(3): 406-412.

Rahmani M, Yousefian N, Akbari A (2009). Energy-based speech enhancement technique for hands-free commun. Electron. Lett., 45(1): 1-2.

Israel C (2003). Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging. IEEE Trans. Speech Audio Processing, 11(5): 466-475.

Gökhun S, Tanyer HŐ (2000). Voice Activity Detection in Nonstationary Noise. IEEE Trans. Speech Audio Processing, 8(4): 478-482.

Zhao DY, Kleijn WB, Ypma A, De Vries B (2008). Online Noise Estimation Using Stochastic-Gain HMM for Speech Enhancement. IEEE Trans. on Speech Audio Processing, 16(4): 835-846.

Brown RG (1983). Introduction to Random Signal Analysis and Kalman Filtering. John Wiley & Sons.

Rabiner L, Juang B-H (1993). Fundamentals of Speech Recognition. Prentice-Hall International.