Robust Voice Activity Detection at Noisy Environment using Average Estimate Least Mean Square Filter

Sang-Yeob Oh
Dept. of Interactive media, Gachon University
1342 Seong-Nam-Dae-Ro, Su-Jeong-Gu, Seong-Nam-Si
Kyung-Gi-Do 461-702, South Korea
syoh1234@gmail.com

Chan-Shik Ahn
Seoul Metro Rapid Transit Media Co.
85-2 kt solution support centers 4th floor, yeomridong-dong, Mapo-gu, Seoul, South Korea
coolsahn@gmail.com

Abstract— In this paper, noise gets reduced with an average estimate LMS filter in a car noise environment so robust voice activities are being detected in a car noise environment. For the noise reduction, an average estimate LMS filter, which is a method for maintaining a source feature of speech and decreasing damages on speech information, is used in a speech signal detection process to reduce noise from a polluted speech signal. An average estimator was calculated and a step size of a LMS filter was controlled with a frame measure to improve an adaptation speed. Moreover, the starting point of result speech, comparing to the previous method of using frame energy was found to be improved by 1.7% and 3.7% of an error rate and an end point error rate, respectively.

Keywords— AELMS Filter, Clustering Model

I. Introduction

Strengths of a LMS filter are that it can be easily applied and have a proper performance on the detection of noise from input signals. Performance of the LMS filter is varied upon performance of a FIR filter and designed upon general performance of a FIR filter as well. Therefore, the LMS adaptation filter is designed to adapt to a changing environment in a fast manner and used for noise estimation and detection algorithm. However, an adaptation speed of the LMS adaptation filter is not that fast and only after a certain period, it adapts and estimates a signal. In order to resolve such weakness, a method for the detection of robust voice activities in a car noise environment. In this paper, noise gets reduced with an average estimate LMS filter in a car noise environment so robust voice activities are being detected in a car noise environment. For the noise reduction, an average estimate LMS filter, which is a method for maintaining a source feature of speech and decreasing damages on speech information, is used in a speech signal detection process to reduce noise from a polluted speech signal. An average estimator amount was calculated and a step size of a LMS filter was controlled with a frame measure to improve an adaptation speed. It was confirmed that convergent performance of the result SNR, which conducted a noise reduction test with an average estimator LMS filter in a noise improvement, was improved. It was revealed that a convergent speed was varied over a step size and its noise reduction rate was more superior than of the existing method. For the speech detection of signals synthesized with car low-speed driving noise and high-speed driving noise, an energy spectrum entropy method was used. Moreover, the starting point of result speech, comparing to the previous method of using frame energy was found to be improved by 1.7% and 3.7% of an error rate and an end point error rate, respectively.

II. Related Research

2.1 LMS(Least Mean Square) adaptation filter

Noise can go through various changes due to environmental factors and shapes of noise even in one specific area are various. An impulse response filter, which has been modeled to adapt to such various changes in noise, is also changed over time. Since noise actually existed in an environment is being calculated in a digital domain, modeling with a linear time varying filter which has impulse responses is possible[1].

2.2 energy and zero crossing rate

A zero crossing rate, which is widely used for searching a frequency where energy is concentrated in a spectrum, refers to the number of crossing of a signal wave form and a zero axis in an interval analysis frame and it is usually shown when continuous sampling values at a discrete signal are very different signs [2]. Therefore, it can be very useful for the segmentation, analysis and recognition of speech. A zero crossing rate is less dependent on the voice of volume of a speaker and when speech is generated, a speech spectrum is reduced by a vocal cord in occurrence of speech so sonant concentrated on low-frequency and a zero crossing rate becomes low consequently. By using this feature, it is possible to detect between sonant and surd [3].

III. Robust Voice Activity Detection at Noisy Environment using Average Estimate Least Mean Square Filter

LMS adaptation filter is interpreted statistically, and when it is composed as a pattern model of speech with high correlation diagram of a probability value where an input signal and estimate are possible, due to a coefficient of a
pattern model which has high correlation coefficients, it experiences a phenomenon of the decrease in a convergent speed. Also, it lowers a noise reduction effect along with the decrease in speech detection rate performance since a speech detection performance gets poor as well. In order to improve such weaknesses, in this paper, an energy spectrum entropy method for noise reduction and robust voice activity detection with a LMS detection algorithm with an average estimator is recommended.

Upon the detection of speech in a very low SNR environment, entropy has the maximum spectral peak energy and the minimum spectral peak energy, which are not that different. So, as for properly detecting speech, harmonic ingredients included in entropy and speech for spectral energy of speech should be detected [4,5]. It could be assumed that a speech domain at a low SNR energy spectrum has a relatively high energy spectrum comparing to a non-speech domain. Such energy spectrum can be expressed similarly as information entropy explained by Shannon [6].

IV. Result of experiment

In order to test a robust voice activity detection method with an average estimator LMS filter, which has been proposed in this paper, a car noise environment speech database (Car3), distributed by SITISC(speech information technology industry support center) is used. As for speech database, speech data collected at a center microphone of a sun-visor in a low speed driving environment (40–60 km/h) and a high speed driving environment (70–90 km/h) of a car.

The comparison of the proposed technique and the previous method of utilizing a frame energy shows that an starting point error rate and an end point error rate of speech were improved by 1.7% and 3.7%, respectively.

V. Conclusion

In this paper, noise gets reduced with an average estimate LMS filter in a car noise environment so robust voice activities are being detected in a car noise environment. For the noise reduction, an average estimate LMS filter, which is a method for maintaining a source feature of speech and decreasing damages on speech information, is used in a speech signal detection process to reduce noise from a polluted speech signal. An average estimator was calculated and an adaptation speed was improved by controlling of a step size of the LMS filter. The test was conducted with 100 words with data collected in a noise environment with the low speed driving and high speed driving of a car. A signal to noise ratio for each environment was calculated to find the intensity of noise. By canceling energy of the noise estimated to energy of a speech data, energy with clean signals was gained and a signal to noise ratio was calculated with a ratio of a clean signal and estimated energy.

References