Robust Vocabulary Recognition Model using Average Estimator 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

Kyung-Yong Chung
School of Computer Information Engineering,
Sangji University, Usan-dong, Wonju-si, Gangwon-do,
220-702, South Korea
dragonhci@hanmail.net

***Abstract*— Noise estimation and detection algorithm should adopt to a changing environment in a fast manner so they use a LMS filter. However, there are some negative points as well. A LMS filter is very low and it consequently lowers a speech recognition rate. In order to overcome such weak point, I would like to propose a method for the establishment of a robust' speech recognition model in a noise environment. Since this proposed method allows the cancelation of noise with the AELMS filter in a noise environment, a robust speech recognition model can be established in a noise environment.**

Keywords— AELMS Filter, Vocabulary Recognition

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. In many cases, the LMS filter is designed to adopt to a changing environment in a fast manner to be applied for noise estimation and detection algorithm, but it can lower a speech recognition rate due to its lower convergence speed. To supplement this weakness, a method for the establishment of a robust speech recognition model in a noise environment is suggested. Since this proposed method allows the cancelation of noise with the AELMS filter in a noise environment, a robust speech recognition model can be established in a noise environment.

The study shows that SNR of speech, which was gained by canceling the environment noise which was kept changing, was enhanced by 2.8dB in an average and a recognition rate was improved by 4.1%.

II. Related Research

2.1 LMS(Least Mean Square) 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].

It starts from an input signal x(n)of an adaptation filter and y(n), representing an output, is shown after subtraction from a d(n). The difference between output y(n) and d(n) is expressed as e(n), an error signal, and if there is a difference between these two values and an error signal exist, it is processed as far-end and if there is no difference and these two values are the same, a signal is send to near-end to process it. The figure 1 represents a structure of the LMS filter[2].

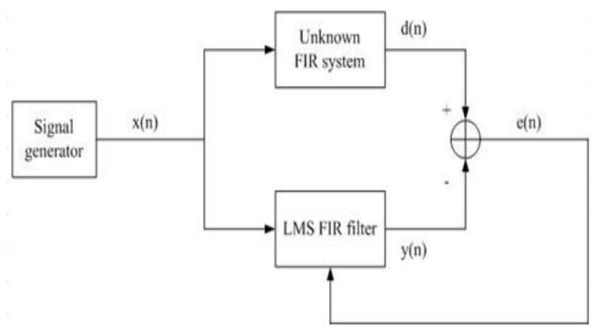

Fig. 1 Structure of LMS Filter

2.2 HMM Speech Recognition Model

HMM is composed of a process which cannot be observed and an observation process which connects an acoustical speech signal vector which is led from a speech signal to the state of not being able to observed. By estimating statistical features of speech, which cannot be observed at HMM, with an observable vector array, it could reflect statistical variability of speech. The figure 2 shows a process of the modelling of a state of HMM and it represents a speech model which allows state transition[3].

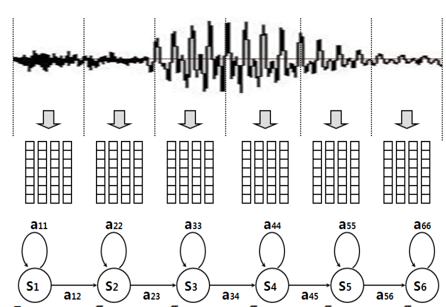

Fig. 2 HMM state modeling process

III. Robust Speech Recognition Model using AELMS FILTER

The AELMS filter is composed as a time-invariant system and it represents a filter for noise cancelation with the AEMLS filter in the figure 3[4].

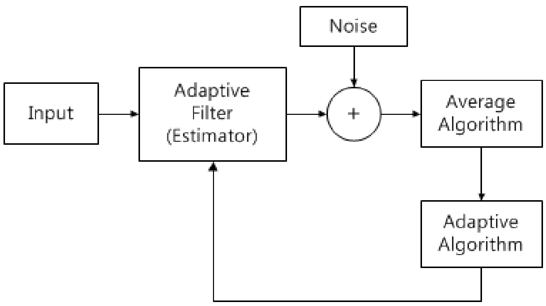

Fig. 3 AELMS Filter

Detection of parameters shows locations values of hv rather than m number of 0(zero), and the detection is made with a cost function and represented in a mathematical manner as the following.

$$J(N) = J_{LS}(N) + m\sigma_v^2 \log N \tag{1}$$

m represents the number of kinetic parameters which have not been indexed and σ_v^2 represents the deviation of $v(k)$ mathematically as the following.

$$J_{LS}(N) = \sum_{k=1}^{N}[v(k) - hvU(k)^T]^2 \tag{2}$$

$X_j(N)$ represents active measured values and J_{LS} minimizes and index of $k = j$. $T(N)$, an active threshold, can be represented mathematically as the following.

$$T(N) = \sigma_v^2 \log N \approx \frac{\log N}{N}\sum_{k=1}^{N} v^2(k) \tag{3}$$

By adjusting parameter thresholds, an estimated value of parameter is calculated and the measured parameter is indexed. With the renewal of $\hat{\theta}_j(k)$, a parameter vector upon its average value with k-time, it can be calculated, and it can be represented mathematically as the following.

$$\hat{\theta}_j(k+1) = \alpha^{1-g_j(k)}\hat{\theta}_j(k) + \pi + g_j(k)u(k-j) \tag{4}$$

$g_j(k)$ is represented as a parameter factor of the j –th $g(k)$, and this step gets repeated to extract a parameter factor and noise from the inputted signals gets canceled to make it converged into a signal to enhance a convergence speed.

IV. Result of experiment

The robust speech recognition model method which has been suggested by this paper has been tested in a noise environment with the AELMS filter. Speech data base which has been applied was down-sampled at 8kHz and about 100 words and noises in 50 speeches in Korean were tested. A signal to noise ratio for each environment was calculated to find the intensity of noise. For conducting the experiment in a noise environment(noise 60～80dB). a ratio of signal to noise was divided by 20dB, 10dB, and 5dB and a speech recognition rate for each was evaluated.

Table 1. Speech recognition rate(%)

Voice Recognition Data	S/R Ratio	recognition rate(%)
Clean Voice Data	20dB	76
	10dB	52
	5dB	34

V. Conclusion

Since noise could get canceled with the AELMS filter in a noise environment, a robust speech recognition model was established in a noise environment. With the AELMS filter, which can preserve sources features of speech and decrease the damage on speech information, noise of a contaminated speech signal got canceled. The study shows that SNR of speech, which was gained by canceling the environment noise which was kept changing, was enhanced by 2.8dB in an average and a recognition rate was improved by 4.1%.

References

[1] J. Homer & I. Mareels. (2004), LS detection guided NLMS estimation of sparse system. Proceedings of the IEEE 2004 International Conference on Acoustic. Speech, and Signal Processing(ICASSP). Montreal, Quebec, Canada.

[2] Li, Q., Zheng, J., Tsai, A., Zhou, Q., (2002). Robust endpoint detection and energy normalization for real-time speech and speaker recognition, IEEE Transactions on Speech and Audio Processing, vol.10, no.3, pp.146-157, March.

[3] Yamagishi, J., Kobayashi, T., Nakano, Y., Ogata, K. & Isogai, J. (2009). Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm. IEEE Trans. Audio, Speech, Language Process., 17(1), 66-83.

[4] Zoltan Tuske, Peter Mihajlik, Zoltan Tobler and Tibor Fegyo, (2005). Robust Voice Activity Detection Based on the Entropy of Noise Suppressed Spectrum, Interspeech 2005, pp. 245-248, Lisbon Portugal., september.