

# Vocabulary Gaussian Clustering model using AELMS Filter

**Jong-Sub Lee**

Dept. of Liberal Education, Semyung University  
65 Semyung-Ro, Jecheon-Si, Chungbuk,  
390-711, South Korea  
[99jslee@semyung.ac.kr](mailto:99jslee@semyung.ac.kr)

**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](mailto:syoh1234@gmail.com)

**Abstract**— 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, and a gaussian model was clustered as a method to make noise more robust. By composing a gaussian clustering model, which is a robust speech recognition clustering model, in a noise environment, a recognition performance was evaluated. The study shows that SNR of speech, which was gained by canceling the environment noise which was kept changing, was enhanced by 2.7dB in an average and a recognition rate was improved by 3.1%.

**Keywords**— AELMS Filter, Gaussian model

## I. Introduction

Even though performance of a speech recognition system has been improved significantly, still noise is an attractive theme for the study of a speech recognition area due to its natural feature of numerous environmental changes. Recognition performance of a speech recognition system is varied over locations due to environmental noises. A speech recognition model is to calculate a probability of the concurrence of a recognition model used for the learning and an actual input speech. To enhance a recognition rate of a speech recognition rate, the LMS(Least Mean Square) Adaptive filter is used for noise cancelation. 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, and a gaussian state model was clustered as a method to make noise more robust . By composing a gaussian clustering model, which is a robust speech recognition clustering model, in a noise environment, a recognition performance was evaluated. The study shows that SNR of speech, which was gained by canceling the environment noise which was kept changing, was enhanced by 2.7dB in an average and a recognition rate was improved by 3.1%.

## 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].

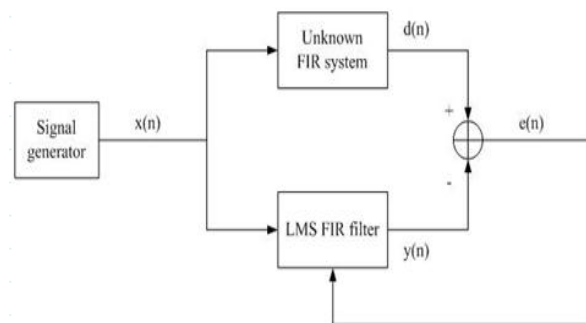


Fig. 1 Structure of LMS Filter

### 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 modeling of a state of HMM and it represents a speech model which allows state transition[3].

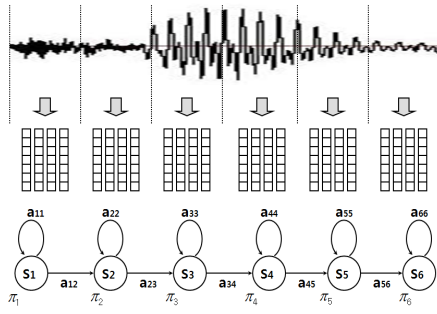


Fig. 2 HMM state modeling process

### III. Vocabulary Gaussian Clustering Model using AELMS FILTER

By using a clustering structure, it becomes possible to decrease the number of parameters which would increase in otherwise, so efficiency of modelling training can be expanded. With the decrease in the number of model parameters, it becomes possible to have less calculation to have the proper recognition, and clustering can be divided into model clustering, state clustering, distribution clustering, and feature clustering upon the degree of clustering. A state clustering sound modeling, based on a decision tree, composes a robust model at a noise including acoustic speech knowledges of recognition speech and established an efficient model along the maximum probability technique. The figure 3 represents the clustering of states.

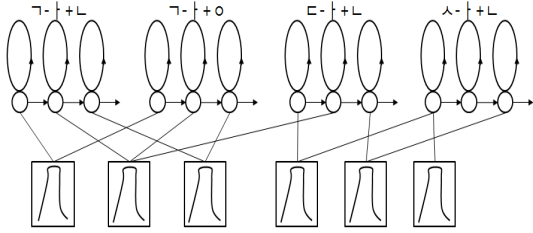


Fig. 3 gaussian state Clustering

Considering a random  $S$  is a group of HMM states and  $L(S)$  as log likelihood of the  $S$  and every state contained in the  $S$  is created from the  $F$ , a group of frames of training data under the assumption that all the states of the  $S$  are bound to,  $\mu(S)$ , common average, and  $\Sigma(S)$ , distribution, would be clustered. Also, if it is assumed that array of frames for each state of the bound states is not changed,  $L(S)$  can be represented mathematically as the following.

$$L(S) = \sum_{f \in F} \sum_{s \in S} \log \left( Pr(o_f | \mu(s)) \right)^G \quad (1)$$

$G$  is a posterior probability created by the state  $s$ , and  $o_f$  represents an observation frame.

Acoustic features of a tri-phone is reflected by collecting gaussian distributions of the HMM into a single large pool and

observation probabilities of each tri-phone uses gaussians contained in a common pool and varies weighted values. A method composing a common gaussian clusters speech feature vectors which has the same central phoneme. A 3-state single phoneme model initial group with a single gaussian output probability density function is created and trained.

### 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 + Gaussian Noise	20dB	62
	10dB	48
	5dB	21

### V. Conclusion

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, and a gaussian state model was clustered as a method to make noise more robust. For representing the intensity of noise in an environment, a signal to noise ratio for each environment was calculated and it was used for the performance evaluation. By composing a gaussian clustering model, which is a robust speech recognition clustering model, in a noise environment, a recognition performance was evaluated.

### 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.