A New Rejection Algorithm Using Word-Dependent Garbage Models

*Gang Sung Lee

※This work was sponsored by Institute of New Technology in Kwangwoon Univ. in 1997.

Abstract

This paper proposes a new rejection algorithm which distinguishes unregistered spoken words (or non-keywords) from registered vocabulary. Two kinds of garbage models are employed in this design: the original garbage model and a new word garbage model. The original garbage model collects *all* non-keyword patterns where the new word garbage model collects patterns classified by recognizing each non-keyword pattern with registered vocabulary. These two types of garbage models work together to make a robust reject decision. The first stage of processing is the classification of an input pattern through the original garbage model. In the event that the first stage of processing is ambiguous, the new word dependent garbage model is used to classify the input pattern as either a registered or non-registered word. This paper shows the efficiency of the new word dependent garbage model. A Dynamic Multisection method is used to test the performance of the algorithm. Results of this experiment show that the proposed algorithm performs at a higher level than that of the original garbage model.

Ⅰ. Introduction

The rejection technique is an important part of the speech recognition and word spotting systems. A word spotting system can be considered as an extension of the speech recognition system. It spans its ability to detect keywords from input speech. Both systems should determine at the end of their procedure whether or not the extracted or recognized words are in the registered vocabulary.

The use of garbage models (or filler models) has been found to be effective in detecting unregistered words from keywords[1]. A garbage model is made from the collection of non-keyword or noise patterns. If the garbage model's score of the test pattern is lower than that of recognized word, then it is classified as a non-keyword.

The technique which doesn't use garbage models uses the difference in log-likelihood of the two highest ranking keywords[2]. If the difference is lower than the threshold, the input pattern is rejected. Although this method reduces the keyword rejection error rate, false alarm rate increases to an amount which is unacceptable to the applications[3].

A two pass classifier for utterance rejection was proposed that utilizes both garbage models and the differences in likelihood scores[4]. It uses discriminant training methods to improve the classification performance. In this technique, parameters are adjusted or trained to maximize the discrimination between the registered and non-registered vocabulary.

A garbage model which is used in the methods described above is made from all non-keyword patterns, regardless of their pattern differences. So the garbage model is the average configuration of non-keywords and certain features of non-keywords may be found to be unacceptable.

This paper presents a new method for keyword/non-keyword classification using both the new word dependent and the original garbage models.

Ⅱ. DMS-Based Rejection System

The method used to evaluate the new rejection algorithm is a Dynamic Multisection Model(DMS)[5]. This model is a composite form of a Dynamic Programming method[6] and a Multisection Vector Quantization method[7]. A DMS model is composed of a number of sections with a codebook and time duration information. A codebook represents sub-patterns in one stable section, and time duration is the average length of a given section.

*Institute of Computer Science and Humanities, Kwangwoon University.

Manuscript Received: June 16, 1997.

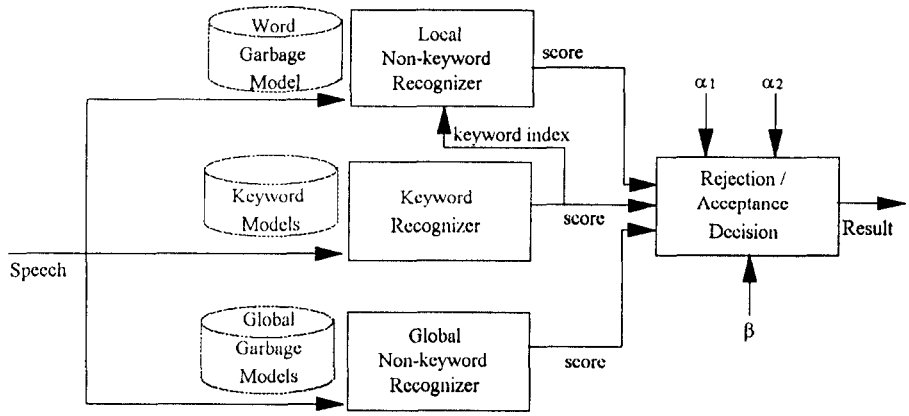

Figure 1. Non-keyword Rejection System Block Diagram

To make a model, all the training patterns are divided into a given number of sections so each section has a similar vector sequence. All vectors in the section are gathered and a codebook is constructed using a clustering technique. The average length of the section is written as time duration information. Models in this system are represented by a DMS method.

When an input speech is transferred through the microphone, the keyword speech recognizer and global non-keyword recognizer check the similarity of the input speech and produce a score. Given the scores of both aforementioned recognizers, the rejection/acceptance block produces a result. If the result is ambiguous, a local non-keyword recognizer reads the keyword index to choose a appropriate word garbage model, and matches input pattern with it. The score of this pattern match becomes the final result of the rejection/acceptance decision block (figure 1).

Ⅲ. Rejection Algorithm

A keyword model is constructed to fit one pattern where a garbage model fit many patterns. The disadvantage of a garbage model is compensated by decreasing the pattern matching score between input pattern t_k and garbage model g_i. This is expressed in the following equation:

$$s_i^g = \alpha\, d(g_i, t_k) \tag{1}$$

$$s_j^k = d(r_j, t_k) \tag{2}$$

where s_i^g is the pattern matching score between garbage model g_i and input pattern t_k, $d(g_i, t_k)$ is the distance function between two patterns, and α is weight parameter. In case of keyword model r_j, the score s_j^k is the same as the distance between input pattern t_k and keyword model r_j.

The value of α is in the range of $0 \le \alpha \le 1.0$. If the value of α is close to 0, the possibility that the garbage model will be chosen increases. If α is small enough and the result is a registered word, we can trust that input pattern is a keyword. If α is large enough and the result is a garbage model, we can trust that the input pattern is a non-keyword. So if we get the reasonable value of α_1, α_2 ($\alpha_1 < \alpha_2$), then we classify the input pattern which gives the obvious result as follows:

$$\begin{aligned} &\textit{if } s^k_{\min} \le \alpha_1 s^g_{\min} \textit{ then } t_k \textit{ is a keyword} \\ &\textit{if } s^k_{\min} \ge \alpha_2 s^g_{\min} \textit{ then } t_k \textit{ is a non-keyword} \end{aligned} \tag{3}$$

where $s^k_{\min}$ is the minimum score of all keyword scores and $s^g_{\min}$ is the minimum score of all global garbage model scores.

If the input pattern cannot be classified with the standards of equation (3), a word dependent garbage model is used for classification. A word dependent garbage model is made from the patterns which are sorted and collected from pattern matches between non-keyword test patterns and keyword models. Let the *i-th* non-keyword test pattern be t_i, *j-th* keyword model be r_j, and word garbage model which depends on the keyword m be g^w_m. The word garbage model g^w_k is made with the patterns:

$$S_m = \{t_i \mid d(t_i, r_m) \le d(t_i, r_l),\ \text{for all } i \text{ and } l\} \tag{4}$$

i.e., S_m is a set of test patterns which are the closest to the keyword model m. The patterns in S_m are processed to get the model g^w_m with a procedure of making model of

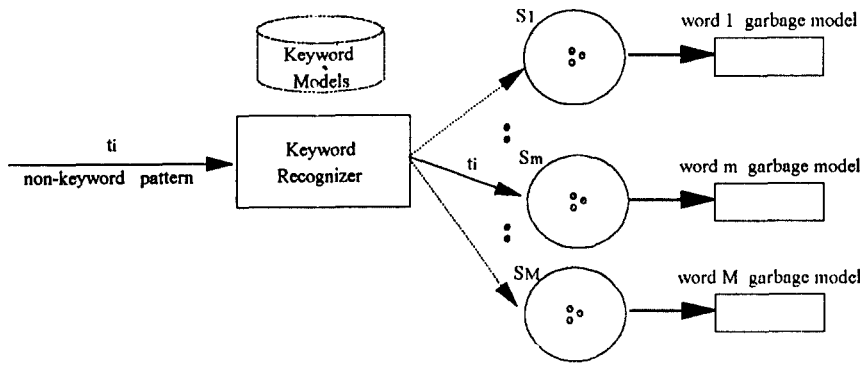

Figure 2. Making Word Dependent Garbage Model

DMS(Fig 2).

In case of the classification using equation (3), minimum score s^k_{min} is compared with the score s^w_k of a word garbage model as follows:

if $s^k_{min} \le \beta s^w_k$ ***then*** *accept the word* $\arg(s^k_{min})$ *as a recognized word (keyword)*

else *reject the input pattern (non-keyword)* (5)

$\arg(s^k_{min})$ is a word index which has a minimum score s^k_{min}. The weight value satisfies $0 \le \beta \le 1$, but it is closer to 1.0.

Ⅳ. Experiments

A. Database

Input speech was sampled at the rate of 10kHz with 16 bits resolution. LPC cepstrum coefficients of order 10 were used as a feature parameter. A frame size is 128 samples and adjacent frames are separated by 128 samples.

Fifty isolated words were recorded three times from 28 male speakers. Thirty of those words were used as keywords and twenty as non-keywords (Table 1). Keyword templates were designed with 1440 utterances from 16 speakers. The remaining 960 (16 speakers × 20 words × 3 repeats) utterances of 16 speakers were trained to make garbage models. 960 utterances were clustered to construct five global garbage models using modified k-means clustering algorithm. A word-dependent garbage model was made for a keyword, if training patterns were classified to the keyword. 1800 (12 speakers × 50 words × 3 repeats) utterances of rest speakers were used for test patterns.

B. Determination of α_1 and α_2

To determine the value of α_1 and α_2 and to get the results of conventional method which uses global garbage models only, experiments were performed using the fol-

Table 1. Word List

No	Korean	English	No	Korean	English
Keyword list					
0	계산기	kyaesanki	1	프린트 매니저	print manager
2	달력	talyuk	3	휴지통	hujitong
4	시계	sikae	5	영한사전	yunghan sajeon
6	도스창	dos chang	7	도움말	toummal
8	윈도우 탐색기	window tamsaeki	9	우리집	urijip
10	워드	word	11	사무실	samusil
12	엑셀	accel	13	일어	yeoreo
14	파워포인트	power point	15	저장	jeojang
16	볼랜드씨	borland c	17	단아	tada
18	MSC	msc	19	다음창	taum chang
20	클립보드	clip board	21	다음그룹	taum group
22	메일박스	mail box	23	아이콘	icon
24	그림판	kurimpan	25	제일크게	cheil kugae
26	메모장	memo jang	27	원래크기	wonrae kugi
28	디스크검사	disk keomsa	29	시작	sijack
Non-keyword list					
30	페이지업	page up	31	백스페이스	back space
32	페이지다운	page down	33	엔터	enter
34	위	we	35	취소	chuiso
36	아래	arae	37	삽입	sabip
38	오른쪽	orunchok	39	삭제	sakjae
40	왼쪽	waenchok	41	한글	hangul
42	탭	tab	43	한자	hanja
44	홈	home	45	영문	yungmun
46	엔드	end	47	대문자	taemunja
48	스페이스바	space bar	49	소문자	somunja

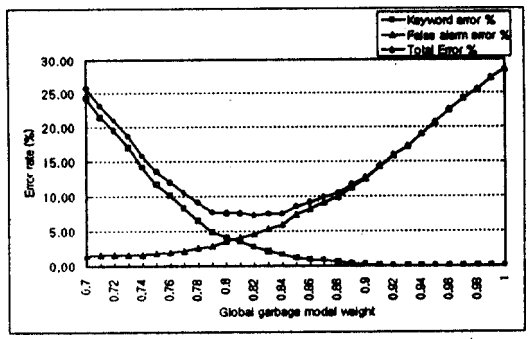

Figure 3. Classification Performance with Global Garbage Models

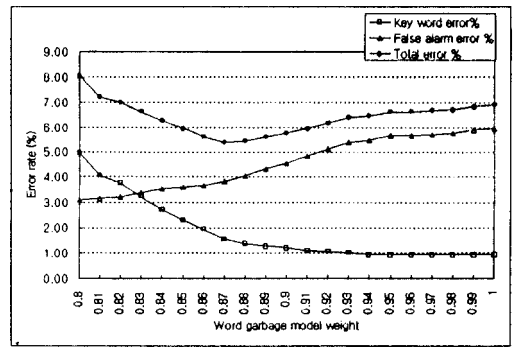

Figure 4. Classifier Permance with Global Garbage Models and Word Dependent Garbage

lowing method:

$$\textit{if } s^{k}_{min} \le \alpha s^{g}_{min} \textit{ then accept the word } \arg(s^{k}_{min}) \textit{ as a recognized word (keyword)}$$
$$\textit{else reject the input pattern (non-keyword)} \quad (5)$$

In this experiment, word dependent garbage models were not used. The error rate was calculated with a value of α from 0.7 to 1.0. The result is shown in figure 3. The keyword error rate is the estimated number of keyword to non-keyword misclassifications. The false alarm error rate functions in the opposite way, non-keyword to keyword misclassifications. Total error rate is the sum of both the keyword error rate and the false alarm error rate. When α has a value of 0.82, the minimum error rate of 7.28% was obtained. This value will be compared with the error rate in a new method.

As the value of α increases, keyword error decreases. In the range of $\alpha \ge 0.86$, keyword error rate is less than 1.00%. In contrast, as the value of α decreases, the false alarm error rate decreases. At the value of 0.74, the false alarm rate reaches to minimum value of 1.56%. So α_1 is set at 0.74 and α_2 is set at 0.86.

C. Determination of β

We get the score s^{k}_{min} by pattern-matching a test pattern with keyword templates, and s^{g}_{min} by pattern-matching with global garbage models. Test patterns are classified with the equation (3), the value of $\alpha_1 = 0.74$ and $\alpha_2 = 0.86$. If it is not classified, equation (5) is used for final decision. To get the optimal value of β, the error rate was calculated with a value from 0.80 to 1.00. The result is shown in figure 4. When β is 0.87, error rate is minimal with 5.39%.

V. Conclusions

A new rejection algorithm for keyword/non-keyword classifier was presented. The new algorithm uses a garbage model which depends on the keyword. A input pattern is classified using the global garbage models. In the event that the result is unclear, a word garbage model is used to give a more accurate result. With this method, the total error rate decreased up to 1.89% giving an error rate 5.38%. This result shows that using the word dependent garbage model is useful in classifying keywords and non-keywords.

References

1. J. G. Wilpon, L. R. Rabiner, C. H. Lee and E. R. Goldman, "Automatic Recognition of Keywords in Unconstrained Speech using Hidden Markov Models," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 38, No. 11, pp. 1870-1878, Nov. 1990.
2. P. Moreno, D. Roe, P. Ramesh, "Rejection Techniques in Continuous Speech Recognition using Hidden Markov Models," *Proc. Signal Processing Conference*, Barcelona, Spain, pp. 1383-1386 Sep. 1990.
3. Luis Villarrubia and Alejandro Acero, "Rejection Techniques for Digit Recognition in Telecommunication Applications," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, Minnesota, Vol. 2, pp. 455-458, Apr. 1993.
4. Rafid A. Sukkar and Jay G. Wilpon, "A Two Pass Classifier for Utterance Rejection in Keyword Spotting," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, Minnesota, Vol. 2, pp. 451-454, Apr. 1993.
5. Lee, G. S. "Speaker Independent Isolated-Word Speech Recognition Using DMS," *Journal of the Institute of New Technology in Kwangwoon Univ.*, Vol. 26, 1997.

6. H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-26, No. 1, pp 43-49, Feb. 1978.
7. D. K. Burton, J. E. Shore and J. T. Buck, "Isolated-Word Speech Recognition using Multisection Vector Quantization Codebooks," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-33, No. 4, Aug. 1985.

▲Gang Sung Lee

Gang Sung Lee was born on January 15, 1964 in Seoul Korea. He received the B.S. degree in computer engineering, and the M.E. and Dr. Eng. degrees in the same field from Kwangwoon University, Seoul, Korea, in 1986, 1988, and 1993, respectively.

In 1990, he joined the faculty of the institute of Computer Science and Humanities at Kwangwoon University, and he has been an Assistant Professor since 1994. His research interests are speech recognition, speaker recognition, word spotting, speech synthesis, and digital signal processing.