International Journal of Fuzzy Logic and Intelligent Systems, vol. 4, no. 2, September 2004 pp. 254-258

The Performance Improvement of Speech Recognition System based on Stochastic Distance Measure

B.S. Jeon*, D.J. Lee*, C.K. Song*, S.H Lee**, J.W. Ryu*

*School of Electrical & Computer Engineering, Chungbuk National University
**School of Electrical and Computer Engineering, Pusan National University

Abstract

In this paper, we propose a robust speech recognition system under noisy environments. Since the presence of noise severely degrades the performance of speech recognition system, it is important to design the robust speech recognition method against noise. The proposed method adopts a new distance measure technique based on stochastic probability instead of conventional method using minimum error. For evaluating the performance of the proposed method, we compared it with conventional distance measure for the 10-isolated Korean digits with car noise. Here, the proposed method showed better recognition rate than conventional distance measure for the various car noisy environments.

Key Words : speech recognition, distance measure, probability, Euclidian distance

1. Introduction

Although a computer occupies a part of our daily life with internet's development, there are many inconvenient elements for the inexpert in communication or interface with computer. So, several kinds of input medias have developed with a computer. For example, computer can be easily operated by touch screen or speech without a keyboard and a mouse. Expecially, speech is the most efficient and natural means of mutual communication. Also, speech can be utilized for naturally communicating between a human and a machine [1].

During recent years, the speech recognition system has been gaining more interests and various approaches have been studied in many universities and institutes. According to increasing safety and convenience, speech recognition system is already applied to operate equipments in car. With development of car industry, the number of handling button is relatively increasing. To prevent an accident from making a cellular phone call in driving a car, a hand free unit is used but it has a problem to handle manually when calling. Therefore, it is able to improve convenience and safety in driving to operate the car's device by speech instead of manual button.

However, it is difficult to design the excellent speech recognition system because the performance of speech recognition system severely degrades under noise. Also, the same speech words show different characteristic by intonation and an accent of pronunciation even if identical speaker [2]. Especially, there are many kind of noises occurred in driving or the one of surroundings [3].

Up to now, various methods have been studied for the robust speech recognition system under the noise. As the methods removing a noise at the preprocessing step of speech recognition, there are various approaches based on extracting robust features for robust feature extraction such as SMC(Short-Time Modified Coherence) [4], RASTA(RelAtive SpecTrAl) processing [5], the dynamic characteristic parameter [6], the cepstrum mapping measure [7], the spectral subtraction method [8,9]. and the method using the characteristic of an auditory system [10], and etc. These methods have the advantage to be able to process independently with a speech recognizer and reduce calculation time. However, these methods can't process properly the noise changing variously with time.

Among the these methods, the spectral subtraction is usually used to suppress noise. In this method, noise is estimated in advance before speech signal is produced. Then, the effect of noise can be easily reduced by subtracting the estimated noise from inputted speech signal. However, it has a problem to can't effectively remove the noise when the estimated noise is different from noise mixed in speech signal.

As the other methods, the wavelet technique has been used to calculate robust feature vectors[12,13] and carry out a comprehensive multiresolution decomposition[14]. The first method calculate the robust feature vectors by adding subband energies obtained by the wavelet transform in the each frame of speech signal. As the other method, mel cepstrum is used to extract feature vectors for static information as well as wavelet transform for dynamic information. However, two methods described above handled about the ideal signal without external noise, and there is no reference to the robustness on external noise. The research of [14] showed just 1-2% improvement of the recognition rate than the conventional method. Besides, the research of based on the wavelet filter banks used the Euclidian distance measure technique to compare a distance of a reference speech with a test speech. And then, the final speech recognition is

Manuscript received May. 25, 2004; revised Aug. 04, 2004.

performed by selecting model with minimum error regardless of distribution of error calculated by each model.

In this paper, we propose a robust speech recognition system under noisy environments. The proposed method adopts a new distance measure technique based on stochastic probability instead of conventional method using minimum error. The stochastic model for each speech is constructed by Bootstrap technique.

This paper is organized as follows. Section 2 describes the conventional VQ-based speech recognition and explain the proposed method based on stochastic probability. Section 3 presents simulation result obtained by using the proposed method. Finally, conclusion remarks are given in section 4.

2. SRS by Distance Measure based on Stochastic Probability

General speech recognition system(SRS) using the vector quantization consists of speech detection part, speech analysis part, training part, and speech recognition part as shown in Fig. 1. Let's briefly explain each part of SRS. In speech detection part, endpoint of speech signal is detected. In speech analysis part, we extract feature vectors of speech signal by using some methods such as the linear prediction coefficient, mel-frequency cepstrum coefficient, and so on.

In training part, various methods are applied according to algorithms. In case of a vector quantization algorithm, the representative codebook is formed by applying feature vectors calculated in speech analysis part to conventional K-means clustering technique. In recognition part, the final speech recognition is performed by selecting the model with the minimum error after comparing an inputted data with a codebook representing each speech model[1].

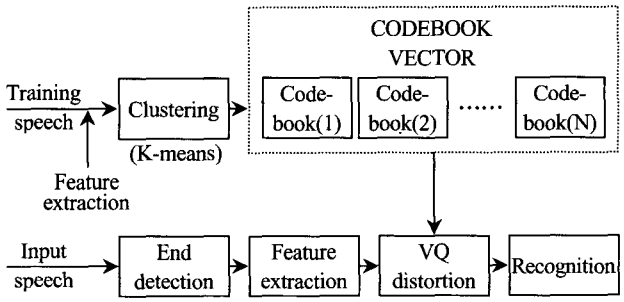

Fig. 1. SRS based on vector quantization

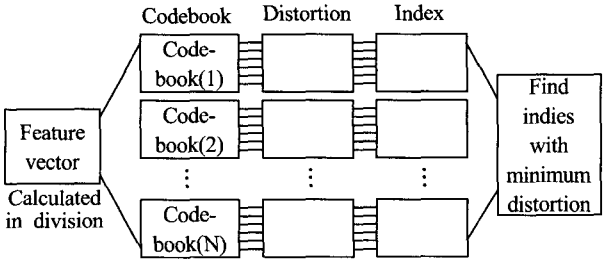

Fig. 2. Producing process of indies

A detailed analysis about an Euclidian distance measure technique is the same with fig. 2. Let's explain this method step by step as follows.

[Step 1] Calculate feature vectors after dividing input speech to each frame, where, if the number of frame is N and the order of feature vector is M, the size of feature vector becomes $N \times M$.

[Step 2] Find the error between the inputted feature vector calculated in the first frame and the feature vectors in codebook.

$$d(x, z) = \sqrt{(x-z)^T (x-z)} \tag{1}$$

where, x represents input vector calculated for each frame and z denotes feature vector in the codebook.

[Step 3] Find the indies of the smallest Euclidian distance calculated by [Step 2] and calculate the score for each speech model.

$$score(speech_i) = \begin{cases} 1 & \text{if } index \in speech_i \\ 0 & otherwise \end{cases} \tag{2}$$

where, $speech_i$ denotes the ith speech model. Here, we consider just isolated word such as /0/, /1/, ···, /9/, So, the number of speech models are 10.

[Step 4] Repeat [Step 2] until the last frame.

[Step 5] Select the speech model with maximum value for the score calculated by Eq.(2).

Unlike above method, the discrimination technique by the distance measure based on stochastic probability is shown in Fig. 3. The procedure of the proposed method is as follows.

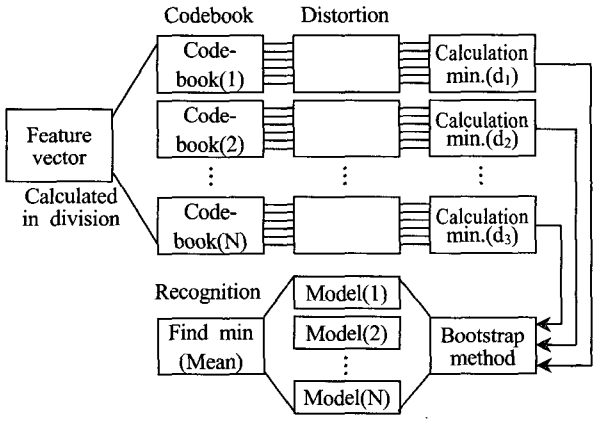

Fig. 3. Calculating process of the distance measure based on stochastic probability

[Step 1] Perform the process of [Step 1] of Euclidian distance measure technique.

[Step 2] Calculate the individual errors for each codebook model by Eq. (1) and calculate the smallest error in each model as

$$D_i(k) = \min(d_i(k)) \quad (3)$$

where, i is the ith speech model, k is the kth frame, and $d_i(k)$ denotes the error calculated in the kth frame for the ith model. $D_i(k)$ denotes the minimum error calculated in the kth frame for the ith model.

[Step 3] Repeat [Step 2] until the last frame.

[Step 4] Find the mean value from stochastic distribution of each speech model obtained by Bootstrap technique using errors calculated in Eq. (3).

[Step 5] Recognize the speech model with of minimum mean value.

3. Experiment and Result

In this paper, SRS by distance measure based on stochastic probability is proposed. In order to improve the performance of recognition system against noise with high frequency components, a speech signal is decomposed by wavelet filterbanks. Speech data is composed of 10-isolated Koran digits as from 0 to 9. These speech data is recorded under environment without any noise. We used total 1000 of speech data containing of 10 individuals. In each speech model, we used 300 speech data as training. And 700 speech data are used for test. The sampling frequency is 11.025kHz and the size of a codebook of a reference pattern is 128. The feature parameter of speech signal is calculated per frame with 20ms Hamming window. Also, let window be in superposition with speech signal by each 10ms so that moved them to compensate the signal information of both ends. So, after dividing an original signal by frame using Hamming window, Mel-cepstrum coefficient of the 10th order is calculated from the speech signal included in each frame. It is very important that detect a start-point and an endpoint exactly in the process of vector quantization. Here, we detect these points by artificial method.

After detecting a start-point and a endpoint of speech, speech signal is transformed by high frequency filter as shown in Eq. (4) to emphasize the elements of high frequency. And then, Mel-cepstrum feature vector was calculated from speech signal.

$$H(z) = 1 - 0.95z^{-1} \quad (4)$$

In order to evaluate the availability of the proposed algorithm, the car's noise was recorded and analyzed at the various sides. Considering all possible cases, a record in a car was executed over the six cases as shown in table 1. The simulation was performed as the method like changing the volume of car's noises as shown in table 1. That is, after adding the noise changed each SNRs to the isolated words (0~9) recorded in the non-noisy condition, the recognition rate was investigated by the simulation.

Table 1. Kind of car's noise

Exp.	Kind of car's noise
I	Car's noise occurred in idling
II	Car's noise + Air-conditioner's noise
III	Car's noise in speed of 30km/h
IV	Car's noise in speed of 60km/h
V	Speed of 60km/h + Air-conditioner's noise
VI	Speed of 60km/h + Air-conditioner's + Radio's noise

A recognition experiment was performed in comparison with the case only using the general vector quantization technique of Euclidian base and it doing the proposed vector quantization technique by distance measure based on stochastic probability. A recognition result about the car's noisy environment I and II is shown in table 2. And a recognition result about the car's noisy environment III to VI is shown in table 3 and 4, respectively. As shown in table 2 to 4, the proposed method illustrated the improvement of recognition performance of 1.57% than the conventional method under the environment that noise doesn't exist. Also, the proposed method showed better recognition performance than the conventional method under the noisy environment. As a result,

Table 2. Experimental result [Noisy environment(I),(II)]

SNR	Noisy environment (I)		Noisy environment (II)	
	Proposed	conventional	Proposed	conventional
NO	97.86	96.29	97.86	96.29
25	95.29	92.71	96.75	94.29
20	93.0	89.57	96.75	93.43
15	87.0	81.43	94.57	91.14
10	67.4	57.9	85.43	78.43
5	31.29	29.71	74.86	68.00

Table 3. Experimental result [Noisy environment(III),(IV)]

SNR	Noisy environment (III)		Noisy environment (IV)	
	Proposed	conventional	Proposed	conventional
NO	97.86	96.29	97.86	96.29
25	98.00	95.43	97.86	95.86
20	97.29	95.14	97.86	95.86
15	96.57	94.00	97.29	95.14
10	96.86	91.71	95.0	93.86
5	86.29	85.29	93.14	91.86

Table 4. Experimental result [Noisy environment(V),(VI)]

SNR	Noisy environment (V)		Noisy environment (VI)	
	Proposed	conventional	Proposed	conventional
NO	97.86	96.29	97.86	96.29
25	97.86	95.71	97.57	95.0
20	97.43	95.57	96.86	94.86
15	96.29	94.71	96.14	93.43
10	95.14	93.29	95.14	93.43
5	91.86	90.74	90.89	89.29

the proposed method showed the improved recognition performance under the noisy environment as well as the environment that noise don't exist.

4. Conclusions

In this paper, SRS that recognize the isolated words by a new distance measure technique based on probability was proposed. In order to evaluate the propriety of a proposed method, we simulated about the Korean single numeral speech of ten words, /0/ to /9/. Then, the proposed method showed the improvement of recognition performance of 1.57% than the conventional method under the environment that noise doesn't exist. Also, the proposed method showed better recognition performance than the conventional method under the noisy environment. As a result, we knew that the proposed method shows the improved recognition performance under the noisy environment as well as the environment that noise don't exist. In the future, we wish to apply and utilize a proposed technique to a SRS for a large-scale continuous sound.

References

[1] L. R. Rabiner and B. H. Juang, *Fundamentals of speech recognition*, Prentice-Hall, 1993.

[2] B. H. Juang, "Speech recognition in adverse environments", Computer Speech and Language, Vol. 5, pp. 275-294, 1991.

[3] H. W. Ruehl, S. Dobler, J. Weith and P. Meyer, "Speech Recognition in the Noisy Car Environment", Speech Commun., Vol. 10, No. 1, pp. 11-22, Feb, 1991.

[4] D. Mansour and B. J. Juang, "The Short-time modified coherence representation and its application for noisy speech recognition", IEEE Trans. ASSP., Vol. 37, No. 6, pp. 795-804, 1989.

[5] H. Hermansky, N. Morgan and H. G. Hirsh, "Recognition of speech in additive and convolutional noise based on RASTA spectral processing", Proc. ICASSP, pp. 83-86, 1993.

[6] T. H. Applebaun and B. A. Hanson, "Regression feature for recognition of speech on quiet and in noise", Proc., ICASSP, pp. 985-988, 1991.

[7] D. Mansour and B. J. Juang, "A family of distortion measures based upon projection operation for robust speech recognition", IEEE, Trans., ASSP, Vol. 37, No. 11, pp. 1659-1671, 1989.

[8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE, Trans., ASSP, Vol. 37, No. 2, pp. 113-120, 1979.

[9] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtracter (nss), hidden markov models and the projection, for robust speech recognition in cars", Speech Commn., Vol. 11, pp. 215-228, 1992.

[10] O. Ghitza, "Auditory nerve representation as a front-end for speech recognition in a noisy environment", Computer, Speech and Language, Vol. 1, pp. 109-130, 1986.

[11] Stephane Mallat, *A wavelet tour of signal processing*, Academic press, 1999.

[12] Gowdy, J. N. and Tufekcim, Z. "Mel-scaled discrete wavelet coefficients for speech recognition", Proc., ICASSP, Vol. 3, pp. 1351-1354, 2000.

[13] Yu Hao and Xiaoyan Zhu, "A new feature in speech recognition based on wavelet transform", Proc., ICASSP, Vol. 3, pp. 1526-1529, 2000.

[14] Kidae Kim, Dae Hee Youn, Chul hee Lee, "Evaluation of wavelet filters for speech recognition", Systems, Mans, and Cybernetics, Vol. 4, pp. 2891-2894, 2000.

Byeong-Seok Jeon

He received the B.S. and M.S. degree in Dept. of Electrical Engineering from Chungbuk National University, Cheongju, Korea, in 1995 and 1998, respectively. He is currently in the doctor's course at the same graduate school. His research interests include robust control theory, intelligence system and fuzzy theory.

Dae-Jong Lee

He received B.S., M.S. and Ph.D. degree in Dept. of Electrical Engineering from Chungbuk National University, Cheongju, Korea, in 1994, 1997, and 2002, respectively. He had worked Electrical Safety Test Research Institute from 2000 to 2003. Since 2003, he is working as a postdoctoral in University of Alberta. His research interests include speech recognition, intelligence system and fuzzy theory.

Chang-Kyu Song

He received the B.S. and M.S. degree in Dept. of Electrical Engineering from Chungbuk National University, Cheongju, Korea, in 1995 and 1997, respectively. He is currently in the doctor's course at the same graduate school. His research interests include image processing, image compression, fuzzy theory, and intelligence system.

Sang-Hyuk Lee
He received the B.S. degree in electrical engineering from Chungbuk National University, Cheongju, Korea, in 1988, the M.S. and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1991 and 1998, respectively. He also received M.S. degree in mathematics from Chungnam National University, Daejeon, Korea, in 2003. He served as a Research Fellow from 1996 to 1999 in the HOW Co. Ltd. Since 2000, he has been with Pusan National University. Currently he is a Chaired Professor in Specialized Group in Industrial Automation, Information, and Communication. His research interests include robust control theory, game theory, and fuzzy theory.

Jung-Woong Ryu
He received the B.S. degree in Dept. of Electrical Engineering from Hanyang University, Seoul, Korea, in 1965. He received the M.S. degree in Dept. of Electrical Engineering from Dankook University, Seoul, Korea, in 1976. He received the Ph.D. degree in Dept. of Electronics Engineering from Dankook University, Seoul, Korea, in 1987. He had worked Daejeon Industrial College from 1969 to 1979, where he had been an associate professor. Since 1979, he is a professor in School of Electrical & Computer Engineering, Chungbuk National University, Cheongju, Korea. His research interests include robust control theory, PID, QFT, intelligence system and fuzzy theory.