Speech Recognition using MSHMM based on Fuzzy Concept

Tae Ock Ann*

※This Research was Supported by the ChonbukSanup University Research Grants in 1997.

ABSTRACT

This paper proposes a MSHMM(Multi-Section Hidden Markov Model) recognition method based on Fuzzy Concept, as a method on the speech recognition of speaker-independent.

In this recognition method, training data are divided into several section and multi-observation sequences given proper probabilities by fuzzy rule according to order of short distance from MSVQ codebook per each section are obtained. Thereafter, the HMM per each section using this multi-observation sequences is generated, and in case of recognition, a word that has the most highest probability is selected as a recognized word.

In this paper, other experiments to compare with the results of these experiments are implemented by the various conventional recognition methods(DP, MSVQ, DMS, general HMM) under the same data.

Through results of all-round experiment, it is proved that the proposed MSHMM based on fuzzy concept is superior to DP method, MSVQ method, DMS model and general HMM model in recognition rate and computational time, and does not decrease recognition rate as 92.81% in spite of increment of speaker number.

I. Introduction

In the speech recognition of word unit, there are DP method[1], VQ[2, 4] or MSVQ[6] method and HMM[7, 8, 9]. Generally DP method requires many memory for templates and has computational load. VQ or MSVQ take advantages of less memory for templates but its performance is not high.

Therefore, HMM in better recognition method than them is selected nowadays. But also in the case of general HMM, many speaker's data for training are needed, and the other speakers' speech(the speaker who did not participate in training models) has notably low recognition rate. So in this study, we will use multi-observation sequence brought in fuzzy concept. With this, in order to increase recognition rate and reduce computational time, we will get multi-observation sequence by MSVQ codebook, and train the MSHMM parameters using multi-observation sequence by multi-section. And in the recognitions, the MSHMM model compared with probabilities of candidate words.

General HMM selects the codeword with the most shortest distance between a vector of VQ codebook and vector by a certain frame of training data into VQ. On the contrary, This model get muti-observation sequence per section by fuzzy concept using distances among vectors of MSVQ codebook and a frame vector of training data. Therefore, the symbol that can't have a probability in training sets because it is not selected as the shortest distance in case of general HMM, can get proper probability, if such symbol belongs to multi-symbol by fuzzy concept, and because of them, we can be improved recognition rate.

And, as MSVQ which include time information against VQ reduces computation time and increases recognition rate, MSHMM against HMM reduces computation time and increases recognition rate by dividing section.

In this study, we do endpoint detection with ZCR and energy, and use double feature vectors which use LPC cepstrum as static feature vector and regression coefficient of LPC cepstrum as dynamic feature vector.

II. Proposed Speech Recognition System

In case of general discrete HMM, it is regarded that only a symbol is observed by each frame vector, but this study applies fuzzy rule which can make multi-symbol with appropriate weights for each frame vector. The symbol

*Department of Computer Engineering, Chonbuk Sanup University of Korea.

Manuscript Received: August 20, 1997.

which was not included in recognition experiment can't be recognized because the symbol which did not selected in HMM training can not have probability in training course. Therefore, setting up multi-symbol can make the possibility of recognition higher because it can have probability that belongs to some symbols that will observe in the same vector.

Moreover, in recognition by general HMM, VQ codebook is generated first, and observation sequence makes based on this codebook. So it has low recognition rate because it does not include information about variation of time. In this study, to overcome this we train MSHMM parameters after dividing training data into some sections and abtaining multi-observation sequence from MSVQ codebook, and also recognize after dividing input data into some sections in recognition.

Therefore, this model is called MSHMM, and when recognition is done by MSHMM, recognition time becomes faster and recognition rate becomes higher.

1. Basic Theory

In the model proposed, we make to change speech signal into multi-observation sequences through the courses of pre-processing, feature extraction, data compression and so on. After finishing these analysis, this study makes MSVQ codebook, and then we optimize MSHMM parameters using fuzzy concept and multi-section concept and recognize as the same method.

1.1 MSVQ Theory

According to Burton's[6] speech recognition which uses MSVQ codebook, we can know that recognition rate and recognition time can be higher and shorter than the ethod which uses VQ because of application of time information. For applying time reduction and time information to HMM model, this study carry out recognition experiment by MSHMM model using MSVQ codebook.

MSVQ codebook about training data shares those into some sections, and then VQ codebook of each section is made by using clustering algorithm. That is, 2 MSVQ codebook are made by dividing training data into 2 sections, and the codebook of each section is composed of 128 codeword, and the number of total codewords is 256. And, 4 MSVQ codebook are made by dividing training data into 4 sections, and the codebook of each section is composed of 64 codeword, and the number of total codewords is 256.

1.2 Multi-symbol using fuzzy

In the same words, their speeches are not of the same waves because of the difference of individual's utterence. Even same person's same words differ according to variation of time. This makes speech recognition more difficult.

In this paper, we will overcome these problems using fuzzy concept and multisection concepts. That is to say, the method to get the observation sequence in general HMM chooses one symbol of the shortest distance between the vector of each frame and codewords of VQ codebook. By the way, in this study, we suggest the way that chooses several symbols of similar characteristics per vector of each frame.

fuzzy rule which is used in computing multi-observation sequence is such as

$$w_s = \frac{2^{S-s}}{\sum_{i=1}^{S} 2^{i-1}} \tag{1}$$

Where, s means the order when it is marked by the order of small distance values between vectors of a frame and each codeword of codebook, and S means the number of multi-symbol which will be selected out of vector of each frame. And w_s is the weight of s^{th} symbol.

This study experiments on S, varying it from 2 to 8. According to the experiment, it is proved that selection of so many number of symbols is meaningless. we get the best recognition rate in case of 6, so fuzzy concept used in this study is adopted S = 6.

2. Speech recognition by proposed method

2.1 HMM Principle using fuzzy concept

The model notation used in MSHMM which applies fuzzy concept by two features vector is the same as below

N = number of states in the model
M = number of observation symbols
T = length of the multi-observation sequence
S = number of symbols observed of each vector
$Q = \{q_1, q_2, \ldots, q_N\}$, states
$V = \{v_1, v_2, \ldots, v_M\}$, discrete set of possible symbol observations
$A = \{a_{ij}\}$, $a_{ij} = Pr(q_j$ at $t+1 \mid q_i$ at $t)$, state transition probability distribution

When k^{th} vector has multi-symbol set $\{v_{k1}, v_{k2}, \ldots, v_{ks}, \ldots, v_{kS}\}$ under q_j state, if the weight set at that time is $\{w_{k1}, w_{k2}, \ldots, w_{ks}, \ldots, w_{kS}\}$, then

$$\sum_{s=1}^{S} w_s(k_s) = 1 \tag{2}$$

So probability that k^{th} vector selects multi-symbol $v_k = \{v_{ks}\}$ under q_j state is

$$B = \{b_j(k)\},\ b_j(k) = \sum_{s=1}^{S} w_s(k_s) b_{js}(k_s) \tag{3}$$

$(1 \le j \le N),\ (1 \le k \le M),\ (1 \le s \le S)$

$\pi = \{\pi_i\},\ \pi_i = pr(q_i \text{ at } t = 1)$, initial state distribution

$O = O_1, O_2, ..., O_t, ..., O_T$, observation symbol

$O_t = \{O_{t1}, O_{t2}, ..., O_{tS}\}$, composition of multi-symbol of t^{th} vector.

According to results of reference (8), we apply for doubly feature vector in this study. That is, symbol of static feature vector at time t, O_t^I is not correlation with symbol of dynamic feature vector, O_t^D, as follows

$$b_s(O_t) = Pr(O_t | S)$$
$$= b_s^I(O_t^I) b_s^D(O_s^D) \tag{4}$$

$b_s^I(O_t^I)$ is the probability of symbol by static vector in state S.

$b_s^D(O_t^D)$ is the probability of symbol by dynamic vector in state S.

Therefore,

$$\sum_{s=1}^{S} w_s(O_s) b_s(O_s) = \sum_{s=1}^{S} w_s(O_s) b_s^I(Ob_t^I) b_s^D(O_t^D) \tag{5}$$

In using doubly feature vector, each weight w is computed when multi-symbol is obtained by each vector. If we call weight by static feature vector $w_s^I(k_s^I)$ and weight by dynamic feature vector $w_s^D(k_s^D)$, fuzzy concept is adopted with 5 techniques.

① $w_s(k_s) = w_s^I(k_s^I)$ (6)

② $w_s(k_s) = w_s^I(k_s^I) \cdot w_s^D(k_s^D)$ (7)

③ $w_s(k_s) = w_s^I(k_s^I) + w_s^D(k_s^D)$ (8)

④ $w_s(k_s) = w_s^I(k_s^I) \vee w_s^D(k_s^D)$ (9)

⑤ $w_s(k_s) = w_s^I(k_s^I) \wedge w_s^D(k_s^D)$ (10)

This models can be marked $\lambda = (A, B, \pi)$ and when it is adopted practically we can consider it by sharing to algorithms for making model and for recognizing.

2.1.1 Model generation

When we wish to calculate the probability of the observation sequence O, given the model λ in front of reestimation for HMM parameters, forward-backward procedure is used as a efficient one. First, we can consider the forward variable, $\alpha_t(i)$, defined as:

$$\alpha_t(i) = Pr(O_1, O_2, ..., O_t, i_t = q_i | \lambda)$$
$$(O_t = \{O_{t1}, O_{t2}, ..., O_{tS}\})$$

this is the probability of partial multi-observation sequence and q_i at time t, given the model λ. We can solve for $\alpha_t(i)$ inductively, as follows.

step 1. Initialization

$$\alpha_1(i) = \pi_i b_i(O_1) = \sum_{s=1}^{S} \pi_i w_s(O_{1s}) b_{is}(O_{1s}) \tag{11}$$
$$, 1 \le i \le N$$

step 2. for $t = 1, 2, ..., T-1,\ 1 \le i \le N$

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i) a_{ij} \sum_{s=1}^{S} w_s(O_{(t+1)s}) b_{js}(O_{(t+1)s}) \tag{12}$$

Step 3. then,

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{13}$$

In a similar manner we can consider a backward variable, $\beta_t(i)$, defined as:

$$\beta_t(i) = Pr(O_{t+1}, O_{t+2}, ..., O_{T-1}, O_T | i_t = q_{ti}, \lambda)$$
$$(O_t = \{O_{t1}, O_{t2}, ..., O_{tS}\})$$

we can solve for $\beta_t(i)$ inductively, as follows:

step 1. Initialization

$$\beta_T(i) = 1,\ 1 \le i \le N \tag{14}$$

step 2. for $t = T-1, T-2, ..., 1,\ 1 \le i \le N$

$$\beta_t(i) = \sum_{i=1}^{N} \sum_{s=1}^{S} a_{ij} w_s(O_{(t+1)s}) b_{js}(O_{(t+1)s}) \beta_{t+1}(j) \tag{15}$$

Where, a problem is to adjust the model parameters A, B and π to maximize the probability of the multi-observation sequence given the model. Therefore an iterative procedure, such as the Baum-Welch method for optimization must be used.

Using the formulas (16), (17), and (18), we can use the Baum-Welch method to reestimate value of the HMM parameters using fuzzy concept.

$$\bar{\pi}_i = \gamma_1(i),\ 1 \le i \le N \tag{16}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1}\sum_{s=1}^{S} \alpha_t(i)\, a_{ij}\, w_s(O_{(t+1)})\, b_{js}(O_{(t+1)})\, \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \alpha_t(i)\beta_t(i)} \tag{17}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1,\, O_t=v_k}^{T} \alpha_t(j)\beta_t(j)}{\sum_{t=1}^{T} \alpha_t(j)\beta_t(j)} \tag{18}$$

2.1.2 Recognition Algorithm

Algorithms used in recognition is not only forward-backward algorithm but also Vitervi algorithm[7]. Forward algorithm according to result of reference (11) is used as recognition algorithm of this paper.

2.2. Structure of this recognition system

Generally, HMM gets observation sequence with VQ codebook, but MSHMM using fuzzy concept gets multi-observation sequence with MSVQ codebook, divides into sections of training data as many as sections of MSVQ, and trains HMM per each section. Figure 1 shows flowchart for it.

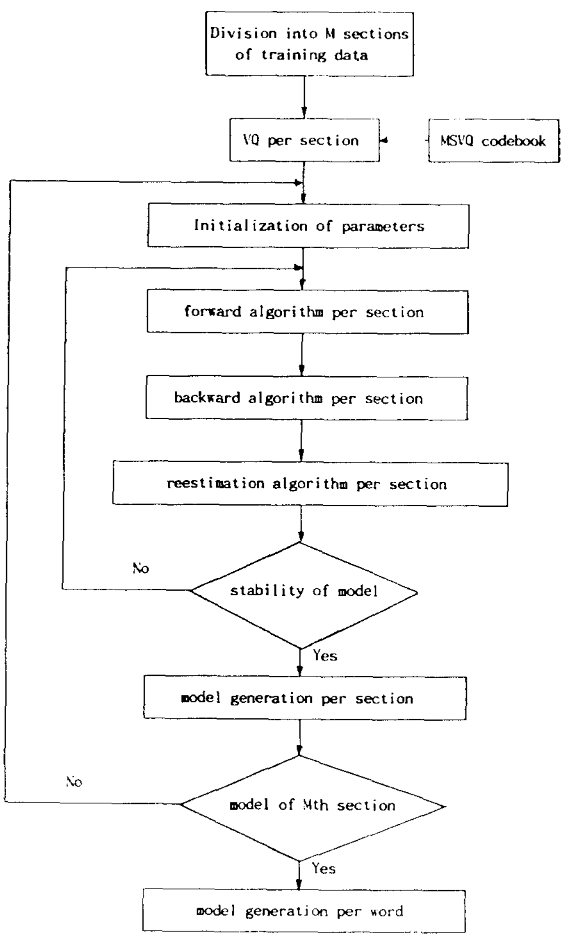

Fig. 1. The method of model generation

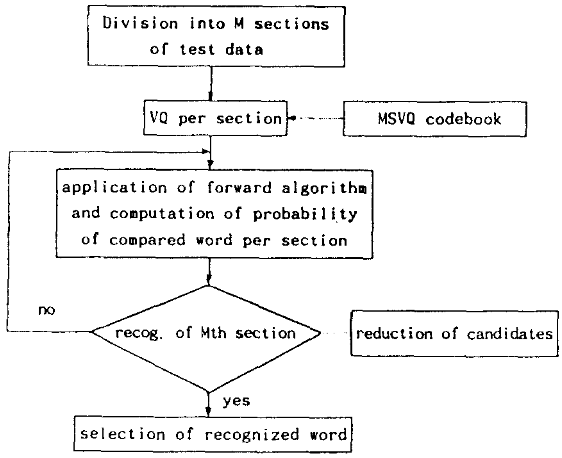

Fig. 2. The recognition method of this model

Also in recognition, after the sections of test data is divided into as those of training set, it is recognized on the basis of it. Because the probability can be computed per each section, the recognition experiment can be accomplished at the faster speed by decreasing candidates. the flowchart of it is shown in figure 2.

This study uses left-to-right model that allow only single transitions.

Ⅲ. Evaluation experiments and results

In speech recognition of male speaker-independent, we chose out Korean 146 DDD area names as the recognition vocabulary, model is made by words spoken two times by five men and we recognize with training data spoken two times by five men. And, in case of the other methods, we use the same data for comparison.

1. Construction for recognition system

Fig. 3 represents the speech recognition system according to proposed model, and we use sampling frequency as

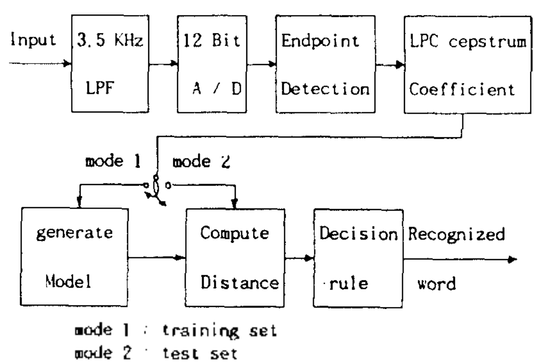

Fig. 3 Block diagram of system

8KHz, LPF as 3.5KHz, and feature parameter as 10^{th} LPC cepstrum coefficient.

2. Result by other classical Recognition Method

In this experiment, in the case of DP pattern matching, clustering algorithm is used for the method selecting reference pattern, and the method[1] of Sakoe and Shiba is used for gloval and local path constraint. In the case of speech recognition by MSVQ[3, 6], the number of section per words is experimented in 8, and the codebook size for each section is 4 and codeword of the total 32 is selected. Result for those is shown in table 1.

Table 1. Recognition rates by classical method

(unit : %)

method / speaker	DP method	MSVQ method
speaker A	81.85	75.34
speaker B	79.45	72.60
speaker C	81.51	77.05
speaker D	80.82	79.45
speaker E	79.45	73.97
Total	80.62	75.68

3. Result by general HMM Recognition Mothod

In general HMM, one method that LPC cepstrum coefficients is used as static feature parameters and the other method[8, 11] that LPC cepstrum coefficients and regression coefficients of its are used as double feature vector is experimented with. In this, 256 codewords are given for each feature vectors and number of state is 8. Result for those is shown in table 2.

Table 2. Recognition rate of general HMM

(unit : %)

feature vector / speaker	HMM using static feature vector	HMM using doubly feature vector
speaker A	85.96	89.38
speaker B	80.82	84.59
speaker C	83.90	88.01
speaker D	82.88	85.62
speaker E	81.51	84.93
Total	83.01	86.51

4. Result by the supposed Recognition System

First, the experiment is done by multi-observation sequence using fuzzy concept. On the basis of this, MSHMM by concept of multisection is experimented. This is experimented in two cases. The first method experiments in state 8 and the codebook size 256 by only LPC cepstrum coefficient. After surveying the result of the first experiment, using LPC cepstrum coefficients as static feature vector and regression coefficients of its as dynamic feature vector, the second method experiments in state 4 and the codebook size 256.

4.1 The case of using only static feature vector

4.1.1 Experiment result using fuzzy concept

Multi-symbol is obtained by expression (1). Number of multi-symbol S is from 2 to 8. The case of modeling and recognition by multi-symbol, and the other case of modeling by multi-symbol and recognition by only single symbol are shown in table 3. In these experiments, the case of recognition by only single symbol is better than the other. number of symbol 6 shows the best recognition rate.

Table 3. Result of HMM using fuzzy rule.

(unit : %)

number of symbol / speaker	2	3	4	5	6	7	8
multi-symbol	87.05	88.49	88.84	89.25	89.59	89.32	89.04
single symbol	85.00	87.60	88.36	89.38	90.14	89.59	88.77

4.1.2 Experiment result by MSHMM using fuzzy concept

In HMM recognition using fuzzy concept, the best recognition rate is shown when number of multi-symbol is 6. Therefore, this factor are used in MSHMM. This experiment is done in section 2 and 4. When the section is 2, number of symbol of each section is 128. When the section is 4, number of symbol of each section is 64. In the case of 2 section, after the first section is recognized, 20 candidates of 146 words are selected, and a word of those is recognized in the second section. In the case of 4 section, after the first section is recognized, 50 candidates of 146 words are selected, in the second section, 20 candidates, in the third section, only 5 candidates. Therefore, recognition time can be reduced. The recognition rate according to number of section is shown in the table 4.

The recognition rate of MSHMM using fuzzy concept, when number of section is 2, number of multi-symbol is

Table 4. Result of MSHMM by number of section

(unit : %)

number of section / speaker	1	2	4
speaker A	92.81	93.49	91.10
speaker B	88.01	89.04	84.93
speaker C	89.38	90.41	89.38
speaker D	90.07	91.44	89.04
speaker E	90.41	92.12	88.70
total	90.14	91.30	88.64

6, is the best in recognition by only static feature vector.

4.2 The case using doubly feature vector

4.2.1 Result by HMM experiment using fuzzy concept

Recognition experiments selected only static feature vector are shown best recognition rate, when nymber of multi-symbol is 6. Therefore, we apply this factor in the case of experiments using doubly feature vector. Expression (6), (7), (8), (9) and (10) is applied in HMM. The result of this experiment is shown in the table 5.

Table 5. Results of HMM's by 5 methods

(unit : %)

technique / speaker	①	②	③	④	⑤
speaker A	92.81	94.18	93.15	92.81	92.81
speaker B	90.75	90.75	90.75	90.75	90.75
speaker C	92.47	92.47	92.81	92.47	92.47
speaker D	91.44	91.79	91.10	91.44	91.10
speaker E	91.10	92.47	91.79	91.79	92.12
total	91.71	92.33	91.92	91.85	91.85

4.2.2 Result from MSHMM experiment using fuzzy concept

In HMM recognition using fuzzy concept, the best recognition rate is shown when expression (7) is applied. Therefore, this expression is used in MSHMM.

This experiment is the same as the case of MSHMM using only static feature vector. The experiment result by this system is shown in the table 6.

In this experiment, when number of section is 2, number of multi-symbol is 6, the recognition rate of MSHMM using fuzzy concept, is the best.

Table 6. Result of MSHMM by number of section

(unit : %)

number of section / speaker	1	2	4
speaker A	94.18	95.21	92.47
speaker B	80.75	90.41	88.70
speaker C	92.47	93.15	89.73
speaker D	91.79	92.47	89.38
speaker E	91.47	92.81	89.38
total	92.33	92.81	89.93

5. All-round experiment result

This study compares method by MSHMM model using fuzzy concept. In this comparision, memorys, computational times and recognition rates is shown in table 7. Rabiner's expression[10] is used in computation for these. In this, average frame number is 40.

Table 7. All-round experimental result

class / method	method	computation	rate(%)
DP	64,240	1,284,800 multiply	80.62
MSVQ	51,392	515,088 multiply	75.68
HMM by single vector	311,168	206,080 multiply 93,440 log	83.01
HMM by doubly vector	306,976	272,000 multiply 46,720 log	86.51
HMM of fuzzy concept by single vector	311,168	206,080 multiply 93,440 log	90.14
MSHMM of fuzzy concept by single vector	311,168	106,240 multiply 49,920 log	91.30
HMM of fuzzy concept of doubly vector	306,976	272,000 multiply 46,720 log	92.13
MSHMM of fuzzy concept of doubly vector	306,976	137,600 multiply 24,960 log	92.81

Ⅳ. Conclusions

In speech recognition using HMM model, it is important to learn parameters of model. This study proposes MSHMM model using MSVQ codebook and fuzzy, and isolated word recognition experiment of speaker-independent is accomplished by this model. Besides the speech recognition experiments by the proposed model, for comparison with it, we perform the experiments by DP, MSVQ and general HMM under same condition and data.

In the case of recognition by DP pattern matching method, storage is too large, and time is too long. In MSVQ, storage is small, but recognition rate is low. In the case of HMM, it takes too long time to make a model, it demands many data, and when it is used for speech recognition of speaker-independent, it demands many training data of many speakers.

In this system, we can be reduced error rate by MSHMM using fuzzy concept, and can be much shorter recognition time and can much higher recognition rate by reduction of candidates because model is learned and recognized in each section.

In the proposed method, computational time and recognition rate of MSHMM that the state number is 4 and LPC cepstrum coefficient and regression coefficient are used as feature vectors are better than those of MSHMM that the state number is 8 and LPC cepstrum coefficient is used as feature parameter.

On the whole, it is proved that MSHMM model using MSVQ codebook and fuzzy concept proposed in this paper is superior to DP method, MSVQ and general HMM model.

References

1. Hiroaki Sakoe and Seibi 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.
2. R. M. Gray, "Vector Quantization", IEEE ASSP Magazine, Vol. 1, pp. 4-29, Apr. 1984.
3. Tae Ock Ann and Sun hyub Kim, "An automatic Speech Recognition of Computer Using Time Sequential Vector Quantization", The Korea Institute of Telematics and Electronics, Vol. 27, No. 7, July. 1990.
4. F. K. Soong, A. E. Rosenberg, L. R. Rabiner and B. H. Juang, "A Vector Quantization Approach to Speaker Recognition", IEEE Trans. of Acoustics, Speech, Signal Processing, Vol. ASSP-33, No. 4, Oct. 1985.
5. D. K. Burton and J. E. Shore, "Speaker-Dependent Isolated Word Recognition using Speaker-Independent Vector Quantization Codebooks Augumented with Speaker-Specific Data", IEEE Trans. of Acoustics, Speech, and Signal Processing. ASSP-33, No. 2. pp 440-443. Apr. 1985.
6. D. K. Burton, J. E. Shore and J. T. Buck, "Isolated-Word Speech Recognition using Multisection Vector Qantization Codebooks", IEEE Trans. of Acoustics, Speech, Signal Processing, Vol. ASSP-33, No. 4, Aug. 1985.
7. L. R. Rabiner and B. H. Juang, "An Intorduction to Hidden Markov Models", IEEE ASSP Magazine, JAN. 1986.
8. M. Niishmura, "HMM-based Speech Recognition using Dynamic Spectral Feature", Proc. ICASSP 89 pp 298-301, May 1989.
9. Kai-Fu Lee, "Automatic Speech Recognition, The Development of the SPHINX System", Kluwer Academic Publishers, 1989.
10. L. R. Rabiner, S. E. Levinson and M. M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Model to Speaker-independent, Isolated Word Recognition", Bell System Technical Journal, Vol. 62, No. 4, Apr. 1983.
11. T. O. Ann, Y. G. Byun and S. H. Kim, "Korean Speech Recognition using DHMM", Acoustical Society of Korea, Vol. 10, No. 1, pp 52-61, Feb. 1991.

▲Tae Ock Ann

Tae Ock Ann received the B.E degree in material engineering from University of Wool-San, Korea, in 1981, and the M.E., Ph.D. degree in computer engineering from University of Kwangwoon, Seoul, Korea, in 1987 and 1991. He is currently a professor in the Chonbuk Sanup University. His research field is speech recognition, computer vision, AI, neural network, and robotics.