A Korean Large Vocabulary Speech Recognition System for Automatic Telephone Number Query Service

자동 전화번호 안내를 위한 한국어 대용량 음성 인식 시스템

(J. M. Koo*, H. S. Kim**, C. K. Un***)

구 준 모*, 김 형 순**, 은 종 관***

ABSTRACT

In this paper, we introduce a Korean large vocabulary speech recognition system which can recognize sentence utterances with vocabulary size of 1160 words, and be used for automatic telephone number query service. This system consists of four sub-systems. The first one is an acoustic processor recognizing words in an input sentence by Hidden Markov Model(HMM) based speech recognition algorithm. The second one is a linguistic processor which estimates input sentence from the result of the acoustic processor and determines next words after a word by using the syntactic information. The third one is a time reduction processor reducing recognition time by limiting the number of candidate words to be computed in the acoustic processor. The time reduction processor uses linguistic information and acoustic information contained in an input sentence. The last one is a speaker adaptation processor which adapts parameters of the speech recognition system to new speakers as soon as possible. The last subsystem uses VQ adaptation and HMM parameter adaptation based on the spectral mapping.

We also present out recent works to improve the performance of the large vocabulary speech recognition system. These works are focused on the enhancement of the acoustic processor and the time reduction processor for speaker-independent speech recognition. New approach for the speaker adaptation is also described.

요 약

본 논문에서는 인식어휘수가 1160단어이며 자동 전화번호 안내에 사용될 수 있는 한국어 대용량 음성 인식 시스템에 관하여 소개하였다. 이 시스템은 네개의 부시스템으로 구성되어 있다. 첫번째는 HMM 방식으로 입력음성중의 단어를 인식하는 음향학적 처리부이다. 두번째는 문법정보를 이용하여 인식대상 어휘를 결정하는 언어학적 처리부이다. 세번째는 음향학적 처리부에서 인식할 어휘를 제한하므로써 인식시간을 감축시켜 주는 인식 시간 감축부이다. 이 부시스템은 언어학적 정보뿐만 아니라 음향학적 정보도 이용한다. 마지막은 음성인식 시스템의 파라미터를 새로운 화자의 음성에 신속하게 적응시켜 주는 화자적응부이다. 마지막 부시스템은 VQ 적응방식과 스펙트럼 mapping 방식에 근거한 HMM 파라미터 적응방식을 이용한다.

또한, 본 논문에서는 대용량 음성인식 시스템의 성능을 향상시키기 위한 최근의 연구결과들에 관하여 살펴보았다. 이 연구들은 화자 독립 음성인식을 위한 음향학적 처리부와 인식시간 감축부의 성능향상에 초점이 맞추어져 있다. 마지막으로 화자적응을 위한 새로우 연구결과라도 기술하였다.

*DigiCom Institute of Telematics, Seoul, Korea
**Department of Electronic Engineering
Pusan National University, Pusan, Korea
***Communications Research Laboratory
Korea Advanced Institute of Science and Technology, Taejon, Korea

I. Introduction

Research on speech recognition in Korea was started before 1980's. Since then, many universities, governmental laboratories and private companies have made their efforts to develope Korean speech recognition systems, thereby some commercial products are now available. As one of thoese research results, we present the works being done at Communications Research Laboratory of Korea Advanced Institutue of Science and Technology, with especial focus on a Korean large vocabulary speech recognition system implemented for automatic telephone number query service. In the early days of our reserach, the main research areas were the analysis, modeling and coding of speech signal. Based on these works, we started research on speech recognition for Korean language in the early 1980's. At first, we studied isolated word recognition algorithms based on the vector quantization(VQ) or the dynamic time warping(DTW) method[1-3]. In recent years, speech recognition algorithms using the hidden Markov model(HMM) were the main tool. As a result of these works, we implemented a 1000-word speech recognitions system. Since it consists of some subsystems which are concernd with important elements for Korean speech recognition, we decided to introduce the system in this paper.

Following this introduction, we present the 1000-word Korean speech recognition system in section II. We describe the task and the structure of the system, and then each of the four sybsystem is described with experimental results. In section III, recent improvements of the large vocabulary speech recognition system and algorithms for real world application is given. And conclusion is made in the last section.

II. Large Vocabulary Speech Recognition System

A. Task

In Korea, people who want to know a telephone number of public offices, companies or any other institutes request information by dialing the telephone number 114. Thus, the service is called the 114 service, and handled by human operators.

The content of the 114 service can be divided into 13 categories including public office number service, changed number service, national code number service for international calls, area code number service for long-distance calls and so on. Since majority of the telephone number information requests are concentrated on some specific public offices, we decided to implement a prototype system for automatic telephone number query service, which can recognize the service requests by human voice and make a response to those inqueries. In order to make the system

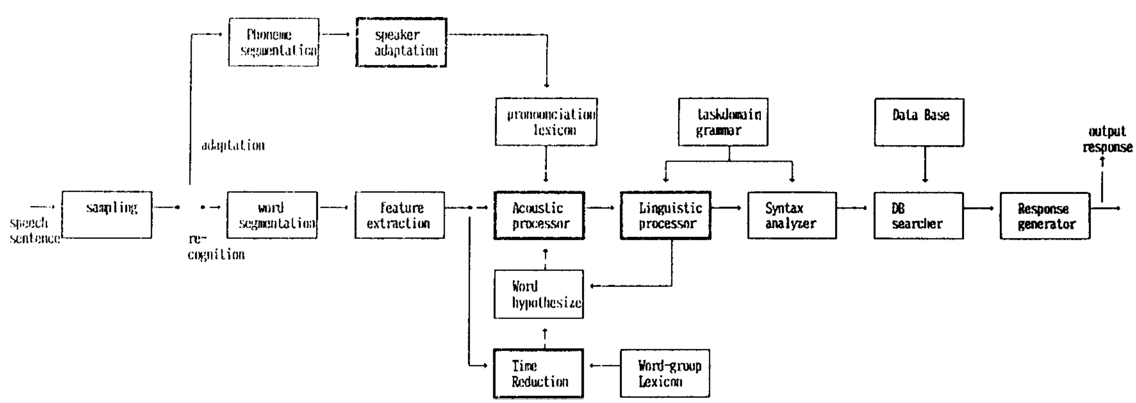

Fig. 1. Total configuration of a Korean Large vocabulary speech recognition system

user-friendly, we designed it to recognize sentence-type speech composed of 1160 words, which contains major public office names, geographical names, interrogative sentences and so on. The automated telephone number query service deals 11 categories which can cover a large part of the current 114 service.

B. Configuration

Our large vocabulary speech recognition system consists of four subsystems-acoustic processor (AP), linguistic processor(LP), time reduction processor(TRP), and speaker adaptation processor(SAP). The relationship of the four subsystems and the total configuration are depicted in Fig. 1.

As shown in Fig. 1, the recognition system has two operation modes, which are the recognition mode and the adaptation mode. In recognition mode, feature extraction procedure is performed first and then words are extracted form the input sentence. Since we assumed that the words are pronounced clearly and with pauses between words, they can be distinguished each other rather easily. For each word, likelihood from the word models are computed by the acoustic processor which uses HMM, where the word models used in this stage are predetermined by the word hypothesizer. The word hypothesizer employs information from the linguistic processor and the time reduction processor which determins candidates words by phoneme-class recognition and word-class selection. The linguistic processor receives likelihoods between input words and the candidate word models, and then corrects errors of the acoustic processor and constraints the number of candidate words by using the task domain syntax. The task domain syntax is also used to analyze the meaning of the recognized sentence to search data base for information request. Finally, the recognition system generates response to the input sentence.

In the adaptation mode, input speech is given as a form of isolated words. From the input words, phonemes are segmented automatically and they are used to modify the VQ codebook and HMM parameters in the pronunciation lexicon. By using the speaker adaptation subsystem, a robust recognition accuracy can be obtained for different speakers.

B.1 Acoustic Processor

Given an input utterance, the acoustic processor segments words from the input sentence by using the short-time energy and the zero crossing rate. And then, the acoustic processor extracts features for each word, where 17 non-uniform filter bank outputs based on the human peripheral auditory information are employed. The acoustic processor is utilizing the discrete hidden Mrakov models(DHHM) to compute the likelihood between an input word and word models. The word models are constructed by concatenating phoneme models. As shown in Table 1, Korean language has 43 phonemes which include 19 consonants, 10 vowels, 12 diphthongs, and 2 semi-vowels. Since some vowels are pronounced very similarly to other vowels in modern Korean and some consonants have very different acoustic characteristics according to their positions in a syllable, we reclassified the phonetic structure to make 47 phoneme-level HMM's including the silence pattern.

Table. 1. Korean Phonetic Inventory

Class	No	Phonetic Symbol
Consonant	19	p t c k s (lax)
		p´ t´ c´ k´ s´(tense)
		p^h t^h c^h k^h h (aspirated)
		m n ŋ (nasal)
		ℓ (liquid)
Semivowel	2	j w
Vowel	10	i y ï u(high)
		e ø ə o
		ε a (low)
Dipthong	12	je jə ju
		jε ja jo
		wi
		we wə
		wε wa
		ïj

For training the HMM's, we extracted training data from 75 phonetically balanced words where all Korean phonemes are distributed evenly. These training data are also used to design VQ codebook which is used for output symbol generation for HMM. Each phoneme-level HMM is a left-to-right model having four states and trained by the well-known Baum-Welch algorithm[4].

To compute the likelihood score between an input word and the word model, the Viterbi scoring algorithm[5] is used. The word models are constructed by referring the pronunciation dictionary which consists of multi-path word models to accomodate the phonological variations in pronunciation.

B.2 Linguistic Processor

Before describing the linguistic processor, we consider the task language model. The task language model is a finite regular language consisting of 1160 words. The syntax of the task language can be represented by a stste transition diagram with 148 states, 294 state transitions and 165 labels. Sentences of the task language have similar grammar structures and many word sets performing the same function. The maximum and the minimum number of words after a word is 667 and 9, respectively. A sentence can be composed of at least 4 words or at most 14 words. Since the number of words at each position in a sentence is much less than the total number of words, this information is useful in the linguistic processing. The operation of linguistic processor is depicted in Fig. 2 and can be described as follows.

Given an input sentence or a sequence of words, the linguistic processor starts its operation when receiving the number of words in the sentence. The linguistic processor sends the list of candidate words at each position in the sentence to the acoustic processor, and receives the likelihood scores between the input word and the candidate word models from the acoustic processor. Based on these scores, the linguistic processor searches the state transition diagrams to determine the next candidate words. In searching the state transition diagrams, we employed the dynamic programming techniques and limite the number of branches to be searched at the next position[6].

At the end of the given sentence, the linguistic processor performs back tracking of the search-path to find three most possible sentences ac-

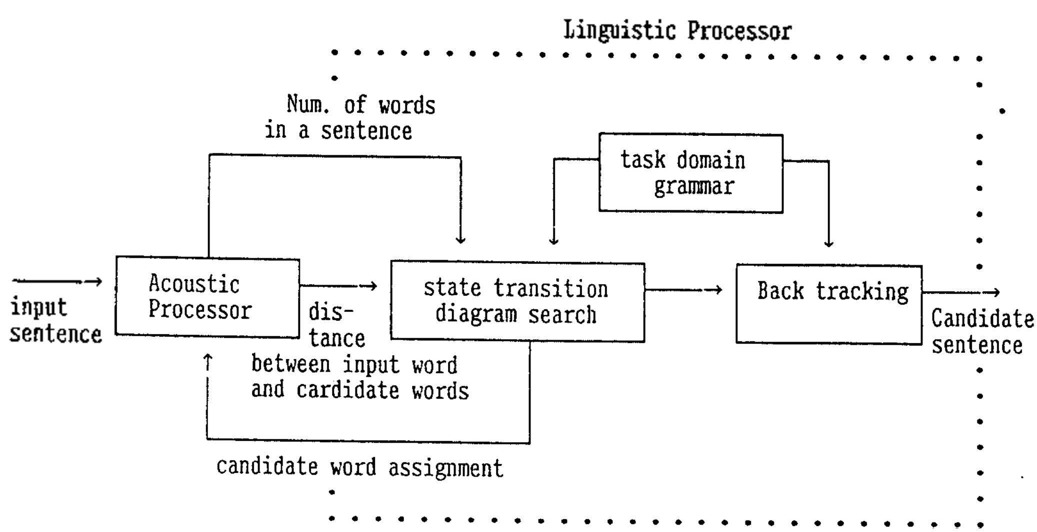

Fig. 2. Operation of Linguistic Processor

cording to the order of the likelihood scores, which are asked to the input speaker in order to be confirmed. If one of the three candidate sentences is correct, the recognition system makes a response to the speaker. In determining the second and the third candidate sentences, we used two methods. One is assuming that only one word in the first candidate sentence is incorrect, and the other is assuming that only one of keywords is incorrect. From our experiments, we found that the latter method yields a little higher semantic recognition rate and a littele lower sentence recognition rate than the former method.

B.3 Time Reduction Processor

One of the problems to be solved in large vocabulary speech recognition systems is how to reduce the recognition time or the computational complexity. Although the recognition time is reduced by using the linguistic information to choose the candidate word as described in section B.2, we devised another algorithm using the acoustic properties of input words to reduce the recognition time further[7]. The block diagram of the proposed algoritm is shown in Fig. 3.

In this algorithm, each phoneme is classified into four phonetic calsses according to the various acoustic features. Given an input word, phoneme segmentation is performed first, and then each phoneme segment is classified. The acoustic features used in this process are the zero crossing rate, the short-time energy, and the short-time energies passed through three different band-pass filters. We defined the four phonetic classes as stop, sonorant, fricative, and vowels. The segmentation and classification procedures are com-

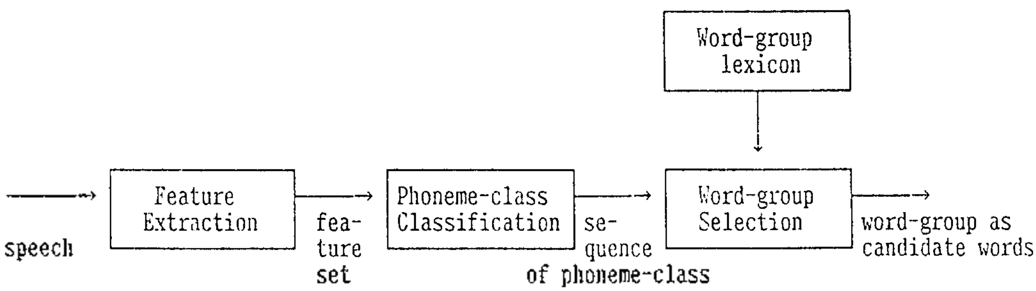

Fig. 3. Block diagram of time reduction processor

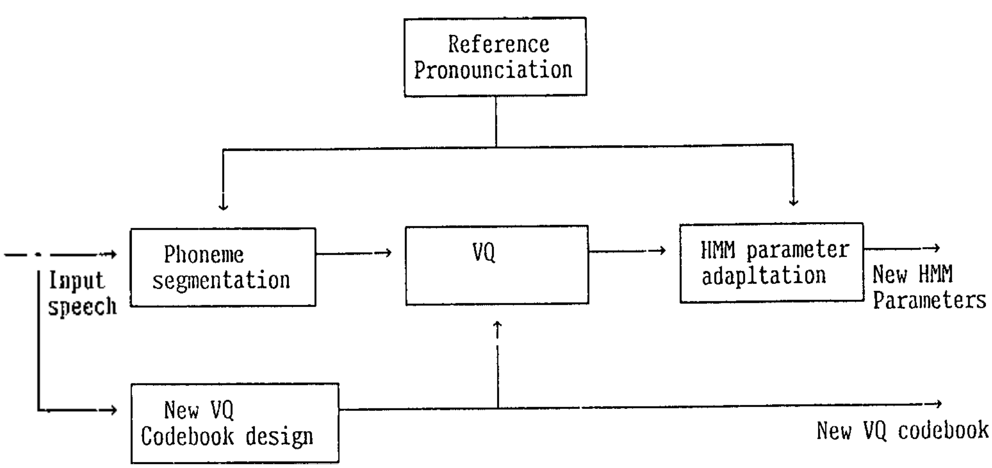

Fig. 4. Block diagram of speaker adaptation processor

plicated and implemented as a kind of rule-based system. Finally, and input word is represented as a sequence of phonetic classes, from which candidate words are determined by referring a word group lexicon. The word group lexicon lists words to the same group only if they are represented by the same sequence of phonetic classes.

By incorporating this time reduction procedure with the information from the linguistic processor, we implemented a word hypothesizer. Since the time reduction processor and linguistic processor are based on the mutually independent features, the number of candidate word can be reduced effectively.

B.4 Speaker Adaptation Processor

Since our large vocabulary speech recognition system is originally designed as a speaker-dependent system, we tried a speaker adaptation mechanism to achieve the speaker independency. The algorithm used in this work is a kind of the static supervised adaptation algorithm. The speaker adaption processor as shown in Fig. 4 consists of two parts-one for segmentation of speech to the adaptation units and the other for adaptation of those units. As the units for the speaker adaptation, we used phoneme-level units that were used in the acoustic processor, and we select 100 phonetically-balanced words to extract those units.

Given an input word from the phonetically-balanced word list, the phoneme-level units are extraced automatically by the dynamic time warping algorithm. That is, the utterance of a new speaker is time-aligned with that of the reference speaker by the dynamic time warping algorithm and segmentation information is obtained by matching phoneme-level unit boundaries of reference word to the new input word.

After all phoneme-level units are obtained through segmentation, they are vector-quantized by a new VQ codebook designed from the utterances of new speaker. If the spectral characteristics of the utterances spoken by new speaker are far different from those spoken by the reference speaker, they can not be represented by the reference VQ codebook effectively, and this is why we designed a new VQ codebook. After the vector quantization procedure, the spectral relation-ship between the new speaker and the reference speaker for each phoneme-level unit is estimated by computing the corresponding probability which shows the probability that an output symbol of the reference speaker is mapped into another output symbol of the new spaeker[8]. Under the assumption that the temporal characteristics of each phoneme-level unit is not much variant, we adapted only the output symbol observation probability matrix to the new speaker.

C. Performance Evaluation

Performance of our large vocabulary speech recognition system was evaluated by computer simulation. For training, 75 Korean phonetically-balanced words and 410 words from the task voacbulary with 1160 words were uttered by one speaker in order to design VQ codebooks and estimate HMM parameters. For the time reduction processor based on the rule-based algorithm, all of 1160 words were used for training. Speaker-dependent recognition test was performed using 100 sentences uttered by the same speaker, which were randomly selected by computer. Totally 568 words were included in the 100 sentences.

At first, the performance of the acoustic processor and the linguistic processor was evaluated and the result is shown in Table 2. In this table,

Table. 2. Speaker Dependent Recognition Accuracy (%) with the AP and the LP

Performance Accuracy	AP Word	AP+LP Word	AP+LP Sentence	AP+LP Semantics	Reduction Rate of LP
Top1	71.5	85.4	43	66	19.1
Top2	85.0		59	80	
Top3	89.8		65	86	

the word recognition accuracy of the AP means that of the AP without the LP. And the word recognition accuracy of AP+LP means that of the AP when the candidate words are given by the LP. The sentence accuracy is very low since

Table. 3. Speaker Dependent Recognition Accuracy (%) with AP, LP and the TRP

Performance Accuracy	AP Word	AP+LP Word	Sentence	Semantics	Reduction rate	Error rate of TRP
Top1	78.90	88.91	60	74	3.42	5.99
Top2	86.80		68	78		
Top3	89.44		72	83		

the sentence is counted as incorrectly recognized one even if only one word is not correctly recognized. But, if the meaning of the recognized sentence is correspondent with the intention of speaker, the sentence is counted as correct one in computing accuracy of semantic recognition. This recognition accuracy is 86% when including upto third candidate sentences. Also, the recognition vocabulary size reduction rate of linguistic processor is evaluated and showed about 20%, that is only one of five words are tested.

A recognition experiment to evaluate the performance of time reduction processor was performed and the results is shown in Table 3. By combining TRP, the number of candidate words are reduced further than the system with only AP and LP from 19.14% to 3.42%. Also, the recognition accuracy for the first candidate words are improved, while slightly degraded for top three case. This fact can be explained considering the error rate of TRP which shows the probability that a correct word for an input speech is not contained in the candidate word lists. Since words recognized incorrectly in top case are usually out of candidate word list, the recognition accuracies including top two or top three candidate word are degraded.

We also performed speaker independent and speaker adaptive recognition experiments, the first half of 100 sentences are used. For comparison, we first evaluated the performance of speaker dependent recognition system for 50 sentences and the results are given in Table 4. And then, speaker independent experiments are performed and very disappointing results are shown in Table 5. Comparing this result with the speaker dependent case, word recognition accuracy of AP only is degraded by 50.16% and the semantic recognition accuracy is degraded as much as 72%. These results imply that the speaker adaptation processor is indispensible. The performance of SAP combined with the speaker independent system is shown in Table 6. For speaker adaptation, we used 100 phonetically-balanced words. By speaker adaptation, the performance of AP improved by 48% and the word, sentence and semantic recognition accuracy are improved by 43%, 30% and 60% respectively. From this experiment, it is known that the speaker dependent system can be used effectively through speaker adaptation procedure.

Table. 4. Speaker Dependent Recognition Accuracy (%) for Comparison

Performance Accuracy	AP Word	AP+LP Word	Sentence	Semantics
Top1	72.35	88.42	48	74
Top2	86.82		66	82
Top3	91.64		72	88

Table. 5. Speaker Independent Recognition Accuracy (%)

Performance Accuracy	AP Word	AP+LP Word	Sentence	Semantics
Top1	22.19	40.84	0	2
Top2	32.80		0	8
Top3	40.82		2	14

Table. 6. Speaker Independent Recognition Accuracy (%) with AP+LP+SAP

Performance Accuracy	AP Word	AP+LP+SAP Word	Sentence	Semantics
Top1	70.74	84.24	30	62
Top2	85.53		44	78
Top3	89.07		56	84

Thus far, the susbystems of the large vocabulary recognition system and its performance have been described. In the next section, recent researches improving the large vocabulary recognition system will be introduced. Also, researches for the real-world application of the large vocabulary recognition system will be reviewed briefly.

III. Recent Studies For Improvements

Although the large vocabulary recognition system was implemented by integrating subsystems considered as important elements for Korean speech recognition, it still has some limitations to be used for the 114 service task. One of the most serious constraints is the speaker dependency. As the results shown in Table 5, speaker independent recognition accuracy is degraded severely. In order to alleviate this problem, some algorithms have studied and they showed some promising results.

A. Improvement of Acoustic Processor

New algorithms improving the speaker independent recognition performance of AP have been studied. They are using the fuzzy mapping concepts[9] for HMM parameter estimation or post-processing of results for competing candidates. The fuzzy mapping concepts showed their efficiency in speaker-independent speech recognition, because the lack of generality comming from the small amount of training data can be alleviated by the estimation of training data distribution based on the fuzzy mapping. We developed three algorithms using fuzzy mapping concepts and evaluated their performances. For this, a speaker independent speech recognition system of which vocabulary is 75 phonetically-balanced Korean words is used. In what follows, we describe the three algorithms and the performance evaluation results.

First, a fuzzy segmental K-means (FSKM) algorithm for the HMM parameter reestimation have been studied[10]. A fuzzy vector quantization (FVQ)-based HMM(FVQ-HMM) scheme requires less training data than a vector quantization-based HMM(VQ/HMM) scheme. However, since the FVQ/HMM scheme estimates the HMM parameters by using the forward-backward algorithm, much computation time is required in the training process. On the other hand, the reestimation method using the FSKM algorithm requires much less computation time than the FVQ/HMM scheme, and also utilizes the state segment information. Furthermore, the FSKM algorithm can be simplified by limiting the codewords in calculating observation likelihoods without lowering the recognition performance. The error rate of the speaker-independent test system is reduced up to 15% when the HMM parameters are reestimated by the FSKM4 method which uses top 4 cnadidate codewords and smoothed by the fuzzy method.

Second, an HMM parameter smoothing method based on the fuzzy mapping concepts has been proposed[11]. In this method, HMM parameters are smoothed by the smoothing matrix obtained by the fuzzy relationship between output symbols and training data. The fuzzy smoothing method reduces the error rate of the test system by approximately 50 percent.

Third, a post-processor using FVQ has been proposed[12]. The recognition algorithm using the FVQ post-processor has much less computation time compared to the FVQ/HMM recognition algorithm. The computation time is reduced by the following three techniques. First, the post-processor use the most likely state sequence previously obtained by the Viterbi algorithm. Hence it does not need to search the optimal state sequence. Secondly, the post processor recalculates the likelihoods for only a few candidate words obtained by the Viterbi scorer. Lastly, the observation probability for an input feature vector in each state is obtained with a few candidate codewords most closely matched to the input feature vector. The error rate of the speker-independent test system is reduced up to 27% when the HMM parameters are reestimated by the FSKM16 method which uses top 16 candidate codewords, smoothed by the floor method, and recognized for top 2 candidate words and top 2 candidate codewords by the FVQ post-processor.

Another algorithm which can improve the speaker-independent performance of AP has been studied[13]. The purpose of this algorithm is the enhancement of discrimination ability of HMMs.

This method is enhancing the probability distriburion of HMM rather than blurring the probability distribution as in the algorithms using fuzzy mapping concepts. That is, with the VQ codebook design algorithm integrated with HMM, the discrimination ability and the robustness of HMM parameters can be improved. For this purpose, we extract codewords from the state segments of each recognition unit by an MKM algorithm or an LVQ2 algorithm, where the segmentation information is obtained by the Viterbi algorithm. For implementation, a unified estimation algorithm of VQ codebook and HMM parameters is used, in which a VQ codebook design procedure and an HMM parameter estimation procedure are alternated until convergence is obtained. The accuracy of the recognition system based on the proposed codebooks are evaluated and compared to those based on the conventional codebooks. The HMM integrated codebook using the LVQ2 method showed the highest recognition accuracy when HMM parameters are used directly. When HMM parameters are smoothed, the error rate of the test system is reduced up to 60 percent for an HMM-integrated codebook using the MKM method.

B. Improvement of Time Reduction Processor

Robust and improved time reduction algorithm has also been proposed for speaker independent recognition[14]. Since the time reduction algorithm described in B.3 of last section uses features which are very sensitive to speaker change, we devised an algorithm robust to variation of speaker and effective in terms of computational time. The proposed algorithm reduces recognition time by choosing candidate words for more detailed inspection according to the coarse likelihood score of every word in lexicon. To compute the coarse likelihood score in a short time, the duration information and speech spectra observation probability of recognition units are used. And the two smoothing methods for speech spectra observation probability have been proposed to improve the classification performance. The computational time of the proposed algorithm is only 7.5 percent of the time required to perform the Viterbi score computation for every word in lexicon. For a 1160-word recognition system, about 72 percent of recognition time can be saved by selecting 20 percent of the vocabulary as candidate words. In this case, the degradation of recognition accuracy is negligible.

C. Improvement of Speaker Adaptation Processor

Another approach achieving high recognition accuracy for many speakers is the use of speaker adaptation. Consequently, the performance of speaker adaptation processor was again improved. For this purpose, three new algorithm are studied and combined into a new speaker adaptation processor whose performance is tested for 100 phoneticallybalanced words as follows[15]. First, we proposed a codebook adaptation scheme using learning vector quantization(LVQ)-a kind of neural network with highly discriminant ability[16]. By the proposed scheme, the codebook was generated to have the discriminant feature rather than the minimum distortion for adaptation speech. From the adaptation experiment, we found that the adaptation with LVQ codebook resulted in higher distortion error than with conventional codebook, but the recognition rate was better, and that LVQ2 codebook, in which, K-means-each codebook was used to initialize, yielded the best recognition rate.

Second, we investigated a modified corrective training algorithm as a method to improve the performance of HMM parameters adaptation. The observation probability parameters of HMM are reestimated by this algorithm after performaing the spectral mapping algorithm. From this experiment, we found that their performance of the speaker adaptation system was improved after adopting the modified corrective training (CT) algorithm, and the highest recognition rate was obtained when the modified CT algorithm was performed on the speaker adaptation system based on LVQ1 codebook in which the K-means-each codebook was used to initialize the LVQ1 codebook.

Third, we studied a hybrid normalization algorithm for feature normalization. Two kinds of normalization algorithms: the iterative DTW method and the mapped codebook algorithm were considered and combined into one for a hybrid normalization algorithm. The experiment indicated that the adaptation by the hybrid normalization method gave the highest recognition rate similar to that by the mapped codebook even though the average DTW distance by the hybrid normalization method is much smaller than that by the mapped codebook method.

Finally, we established the speaker-independent, speaker adaptive and speaker dependent recognition systems, and compared their performances with each other. From the experiment using the same male speaker, we obtained the recognition rate of 80.3% with the speaker-independent recognition system, and 96.3% with the speaker adaptation system. Then we established the speaker-dependent recognition system using the corrective training algorithm and obtained the recognition rate of 98.0%.

D. Miscellaneous Works

Also, there were reseaches which can be used for real-wold application of the automatic 114 service. First, the effect of noise and distortion of telephone channel on the speech recognition has been studied, and a method to improve the recognition rate has been proposed[17]. Second, a dedicated VLSI architecture is proposed for the Viterbi scoring procedure in an HMM-based real-time large-scale isolated word recognition system. The proposed architecture can update every state metric of HMM models up to vocabulary size of 1000 words with the average 50 states per word within a 10ms frame interval, thus achieving a real-time throughput[18].

Currnently, algorithms for continuous speech recogntion are under research. This research includes not only the modification of AP but also the developement of parsing algorithm suitable for Korean language. Another research is going on to improve recognition accuracy under noisy environment.

IV. Conclusion

Until now, we reviewed a Korean large vocabulary recognition system which consisting of acoustic processor, linguistic processor, time reduction processor and speaker adaptation processor. Before reviewing, we introduced some linguistic and phonetic features of Korean language. And then the total configuration of the Korean large vocabulary recognition is presented with algorithms of each processor. The performance evaluation results were also described. Through this review, language specific factors as well as general speech recognition algorithms are considered. Recent researches for improving the performance of the large vocabulary recognition system are also described. Especially, the acoustic processor algorithms for speaker-independent recognition were proposed and showed some promissing results. Also algorithms for another processors were presented. By combining these algorithms, more robust and accurate Korean large vocabulary recognition system will be implemented soon.

References

1. H. S. Kim and C. K. Un, "Improving discriminability among acoustically similar words by modified distance metric," Elect. Lett., vol.24, no.3, pp. 161-163, Feb. 1988.
2. H. S. Kim and C. K. Un, "Modified dynamic time warping algorithm based on transient information," Jour. Acous. Soc. of Korea, Vol.8, No.6, pp.82-90, Dec. 1989.
3. H. S. Kim and C. K. Un, "On the use of dual processing nature of speech perception for speech recognition," Jour. Acous. Soc. of Korea, Vol.8, No. 6, pp.91-105, Dec. 1989.
4. L. E. Baum, et al., "A maximization technique occuring in the statistical anaysis probabilistic functions of Markov chains," Ann. Math. Stat., vol.

41 pp.164-171, 1970.

5. G. D. Forney, Jr., "The Viterbi algorithm," Proc. IEEE, vol.61, no.3, Mar. 1973.
6. S. E. Levinson, "The effect of syntatic analysis on word recognition accuracy," Bell Syst. Tech. Jour., Vol. 57, No. 5, pp.1627-1644, May-June 1978.
7. C. K. Un, et al, "Phone-class and word-group recognition for recognition time reduction" in Fifth-Year Final Report on Development of Speech Recognition System(in Korean), CRL-P-8901, pp.73-125, KAIST, Jan. 1989.
8. R. Schwarttz, et al., "Rapid speaker adaptation using a probabilistic spectral mapping," Proc. Int. Conf. on Acoust., Speech, Signal Processing, pp. 633-636, 1987.
9. J. C. Bezdek, Pattern recognition with fuzzy objective function algorithm, Plenum Press, New York, 1981.
10. H. R. Kim and H. S. Lee, "Performance improvement of HMM-based speech recognition system using FVQ and segmental information," in writing.
11. J. M. Koo and C. K. Un, "Fuzzy smoothing of HMM parameters in speech recognition," IEE Elect. Lett., vol. 26, no. 11, pp.743-744, May 1990.
12. H. R. Kim and H. S. Lee, "Postprocessor using fuzzy vector quantization in HMM-base speech recognition," IEE Elect. Lett., vol. 27, no. 22, pp. 1998-2000, Oct. 1991.
13. J. M. Koo, H. S. Lee and C. K. Un, "An improved VQ codebook design algorithm for HMM," Proc. of Int. Conf. on Acoust., Speech and Signal Processing, pp.S45.2.1-4, San Francisco, Mar. 1992.
14. J. M. Koo and C. K. Un, "A recognition time reduction algorithm for largevocabulary speech recognition," Speech Communication, vol. 11, pp. 45-50, 1992.
15. M. W. Koo, "A study on speaker adaptation for a large vocabulary speech recognition system," Ph.D dissertation, KAIST, 1991.
16. M. W. Koo and C. K. Un, "Speaker adaptation using learning vector quantization," IEE Elect. Lett., vol. 26, no. 20, pp.1731-1732, Sept. 1990.
17. S. J. Do and C. K. Un, "A study on the improvement of isolated word recognition for telephone speech," Jour. Acous. Soc. of Korea, vol. 9, no. 4, pp.66-76, Aug. 1990.
18. J. Y. Kim, H. S. Lee and C. K. Un, "An efficient Viterbi scoring architecture for HMM-base isolated word recognition system," Proc. of Int. Conf. on Spoken Language Processing, pp.13.6.1-4, Kobe, Nov. 1990.

▲Koo Jun Mo

Koo, Jun no was born in 1963. He received the B.S degree in electronics engineering from Seoul National University in 1985, the M.S. degree and Ph.D degree in electrical engineering from Korea Advanced Institute of Science and Technology in 1987 and 1991, respectively. He has been with Digicom Institute of Telematics since July 1987, where he is currently in change of speech processing laboratory. His research interest includes speech recognition, speech synthesis, speech coding and application of digital signal processing technology.

▲Hyung Soon Kim : VOL.8 NO.6 1989

▲Dr. Chong Kwan Un received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Delaware, Newyork, Delaware, in 1964, 1968 and 1969, respectively. From 1969 to 1973 he was Assistant Professor of Electrical Engineering at the University of Maine, Portland, where he taugth communications and did research on synchronization problems. In May 1973, he joined the Staff of the Telecommunication Sciences Center, SRI International, Menlo Park, CA, where he did research on voice digitization and bandwidth compression systems. Since June 1977 he has been with KAIST, where he is Professor of Electrical Engineering and Head of the Communications Research Laboratory, teaching and doing research in the areas of digital communications and digital signal processing. He has authored or coauthored over 200 papers and 70 reports in speech coding and processing, packet voice / data transmission, synchronization, and digital filtering. Also, he holds 7 patents granted or pending. From February 1982 to June 1983 he served as Dean of Engineering at KAIST.

Dr. Un has received a number of awards including the 1976 Best Paper Award from the IEEE Communications Society, the National Order of Merits (Dong Baik Jang) from the Government of Korea, and Achievement Awards from KITE, KICS and ASK. He is a Fellow of IEEE and a member of Tau Beta Pi and Eta Kappa Nu Honor Societies.