# Development of a Door Lock System by Speaker Verification Using Weighted Cepstrum and Single Average Pattern

Younjeong Kyung*, Jongsoon Jung**, Seungho Choi*, Hwang-Soo Lee*

ABSTRACT

In this paper, we implement the door lock system based on pattern matching technique for speaker recognition using DTW. In this study, major features of our system are summarized as follows: (1) Make the average reference pattern using DTW. This method keeps the high recognition rate compared with the other systems whose performances degrade rapidly as time goes on. (2) Use F-ratio values as the weighting values of the cepstral coefficients. We find that the weighted cepstra reveals an effect on intensifying the difference between the customer and the imposter. The system hardware is composed of two parts: the door lock part and the speaker recognition processing part. We use an 8051 microprocessor in the door lock part for serial communication with host processor to open or close the lock. Using our system, we obtain speaker recognition rate of about 99.5%.

## Ⅰ. INTRODUCTION

Widely used means of individual verification for public security such as card key, seal, signature and identification cards have the risk of robbery or forgery. On the one hand, a security system using the verification of fingerprints has the demerit of higher cost of additional equipments. Speech signal includes "linguistic information for transfer of meaning" and "speaker information-who says?". Our study in this paper aims to develop a secure door lock system by using speaker verification techniques.

Traditional speaker verification systems have several problems. 1)They use several reference patterns for combatting intra-speaker variations. It results in heavy computational load. 2)Their performances degrade rapidly as time goes on.

We attempt to solve the problems as follows. In our system, we use only one reference pattern by averaging several patterns using DTW path information and updating the pattern as time passes on off-line. We use the weighted cepstrum by F-ratio value for the improvement of speaker recognition.

* KAIST, Department of information and Communication Engineering
** MunKyung Junior College

Manuscript Received April 30, 1996.

※This study was supported in part by the Korea Science and Engineering Foundation. (The Contract number: 95-0100-22-01-3)

This paper is organized as follows. In section II, we describe the speaker recognition system by pattern matching technique. This is our base system which is applied with betterments of section III and section IV. In section III, we provide the generation method of average reference pattern using DTW algorithm. We explain the weighted cepstrum which is the feature parameter of our recognition system in section IV. We present the overall structures of the system in section V. In section VI, we discuss the experimental results and the performance of the system. Conclusions are made in section VII.

## Ⅱ. THE SPEAKER RECOGNITION SYSTEM

The speaker recognition system is divided into various classes. According to the recognition techniques, the system is classed as pattern matching, neural nets, vector quantization and hidden Markov model, etc. The pattern matching system have an effect on the application part. The reason is that it needs relatively simple algorithm and minimum hardware. In this paper, we use the pattern matching technique for a simplicity. According to the recognition subject, the system is classified into the speaker identification and the speaker verification. Given a candidate speaker, a speaker identification system attempts to answer the question, "Who is he?". A speaker verification system addresses, "Is he who he says he is?". Whether the text of recognition system is fixed or not, the system is

classified into the text dependent system or the text independent system[1].

Our system is the text dependent speaker verification system where the text is the customer's name. The text independent system needs the large size of speech data for speaker recognition. This is not applicable to the door lock system's environment.

The Figure 2.1 shows the block diagram of the speaker verification system using pattern matching technique.

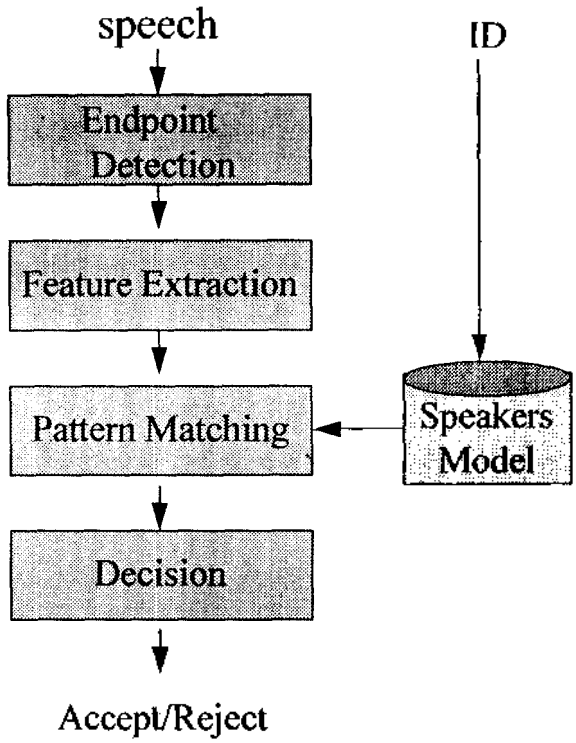



Figure 2.1 The block diagram of the speaker verification system using pattern matching technique.

The inputs of speaker verification system are customer's speech and their identification number. Input speech pass through the many procedures such as the end-point detection, the feature extraction and the pattern matching. The system declare the speaker as the customer or imposter by the distance value of the two patterns.

### 2.1 End-Points Detection

The correctness of speech end-points detection have an important effect on the performance of speech/speaker recognition system. In this paper, we use the short-term energy and zero crossing rate(ZCR) parameters. We control the threshold value by adaptive threshold value method. The short-term energy parameter consider the large energy frame as speech, the small energy frame as silence. This method fail to distinguish speech from silence in case of the small energy speech data or large background noise. To make up for the weak points in this energy parameter, we use the ZCR.

We also consider the pause of interword. For the case of the pause of interword, we look upon as the speech if starting point is detected within 40ms after endpoint detection.

### 2.2 Feature Extraction

The frame length is 30ms. The frame period is 10ms. We use the Hamming window. The 16 order LPC coefficients are extracted from speech data that pass through the preemphasis processing. We obtain the cepstra from LPC coefficients. The mel-scaled cepstra are extracted by warping module that transform the LPC coefficients on the basis of mel-scale.

The Figure 2.2 shows the feature extraction process.

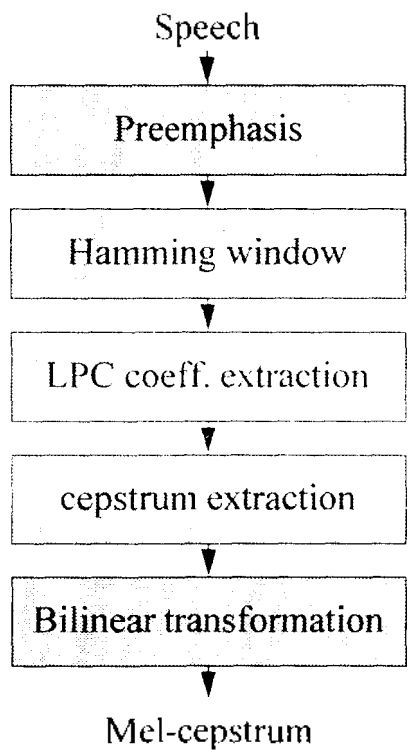



Figure 2.2 The feature extraction process.

### 2.3 Pattern Matching using DTW

Utterances are generally spoken at different rates, even for a single speaker repeating the same word. Thus, test and reference utterances normally have different durat-

ions. Most high-performance speaker recognizers address the problem of alignment by nonlinear warping one template onto another in an attempt to align similar acoustic segments in the test and reference templates. The procedure, called Dynamic Time Warping(DTW), combines alignment and distance computation through a dynamic programming procedure[2].

Deviations from a linear frame-by-frame comparison are allowed if the distance for such a frame pair is small compared to other local comparisons. In the absence of specified segment boundaries, DTW aligns templates by finding a time warping that minimizes the total distance measure, which sums the frame distances in the template comparison. The warping curve derives from the solution of an optimization problem $D = \min_{w(n)} [\sum_{n=1}^{T} d(T(n), R(w(n)))]$ where each d(,) term is a frame distance between the nth test frame and the w(n)th reference frame. D is the minimum distance measure corresponding to the 'best path' w(n) through a grid of T×R points. We select the ID of reference pattern that have the minimum D out of the whole speaker's reference pattern.

Many speaker recognition systems that use the pattern matching are published[3]-[5].

## Ⅲ. GENERATION OF THE AVERAGE REFERENCE PATTERN

Traditional speaker verification systems use several reference patterns for combatting intra-speaker variations. For example, in case that the number of reference patterns is N, we should compare with test pattern with them N times. It results in heavy computational load. Also in this case we should reconstruct the reference pattern as time passes.

The probability of false acceptance error increases, provided customers use the reference patterns that is made in too long time. It is due to the large allowable error of speech variation as time goes on. On the contrary, the probability of false rejection error increases, provided customers use reference patterns that is made in too short time. Both of them cause the increase in the recognition error rate.

In this system, we propose the method of making the reference pattern over a six-month period for the decrease in the recognition error rate. The method of constructing the reference pattern is shown in the Figure 3.1. We obtain the one reference pattern from six patterns using the DTW. That is, we average the corresponding frames by

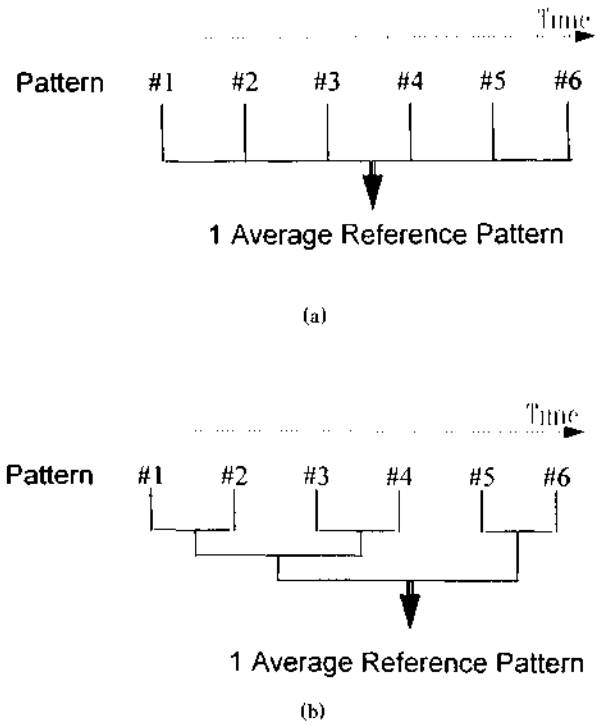



Figure 3.1 Construction of the average reference pattern.

path information obtained from DTW. The method of obtaining the one reference pattern from six speech utterences is various from the view point of computational methods. Figure 3.1(a) averages the six patterns by equal weight value. The Figure 3.1(b) averages the 6 patterns by more weighting to 5 and 6 pattern. This paper computes the reference pattern by the Figure 3.1(b) for the weighting to the latest speech pattern.

This method keeps the high recognition rate compared with the other systems whose performances degrade rapidly as time goes on. This method is also easy to update the reference pattern. In the case of updating the reference pattern, the least speech pattern reflects the change of speaker's voice. So, the least speech pattern is weighted in comparison with the already established patterns.

## Ⅳ. WEIGHTED CEPSTRUM BY F-RATIO

We propose the weighted cepstrum to maximize the inter-speaker discriminability. In this paper, we find the effective cepstral coefficient for each speaker and then weight the cepstral coefficient. This paper use the F-ratio value as the cepstral coefficients as the weighting value. Experimental results show that weighted cepstrum is good parameter for speaker recognition[6].

The F-ratio value is often used to the effectiveness me-

asurement of feature parameters. It is determined as the inter-speaker variance divided by intra-speaker variance. F-ratio values are defined as[7]

$$F\text{-}ratio = \frac{variance\ of\ means\ of\ different\ speakers}{mean\ intraspeaker\ variance} \qquad (4.1)$$

The numerator is large when values for the speaker averaged feature are widely spread for different speakers, and the denominator is small when feature values in the utterances of the same speaker vary little.

The feature of parameter is good when the intra-speaker variation is small and the inter-speaker variation is large.

In this paper, we compute the F-ratio value on each cepstral coefficient to measure the effectiveness of each cepstral coefficient in the inter-speaker's distinction. Our F-ratio value is computed by equation (4.2) according to the above definition. The equation (4.2) is used for the general F-ratio value and the equation (4.3) is used for the each speaker's F-ratio value. The results of equation (4.3) is used as weighting function $W(i)$.

$$F\text{-}ratio = \frac{Var(E(C_{ij})_j)_{whole\ speakers}}{E(Var(C_{ij})_j)_{whole\ speakers}}$$

$$i = 1, \cdots, I \qquad j = 1, \cdots, J \qquad (4.2)$$

$$W_j(i) = \frac{Var(E(C_{ij})_j)_{whole\ speaker}}{E(Var(C_i))_{each\ speaker}}$$

$$i = 1, 2, \cdots, I \qquad j = 1, 2, \cdots, J \qquad (4.3)$$

Where I is the order of cepstral coefficients and J is the number of speakers registered.

If the distribution of F-ratio values of each cepstrum order is equal or has little variation, the speaker information is distributed in the whole cepstrum order. On the contrary, if the F-ratio values of the cepstrum order have large variation, it shows that the specific cepstrum order has more speaker information. Figure 4.1 shows the general F-ratio values by equation (4.1). We can see the large variation of F-ratio values in each order of cepstrral coefficient in this figure. It says that some specific cepstral coefficients more effectiveness on speaker recognition. The results of F-ratio values on each speaker is shown in Figure 4.2 by equation (4.2). The example system have 4 customers. We find the different F-ratio distributions for all the speakers in the Figure 4.2. This fact explains that the above weighting function is appropriate to obtain more discriminability between speakers.

The performance of speaker verification is determined

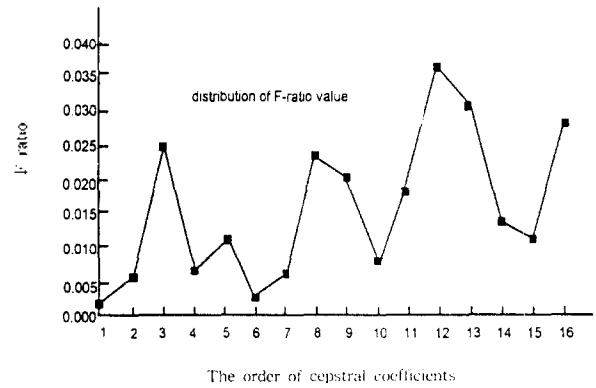



Figure 4.1 The distribution of F-ratio values of each cepstrum order.

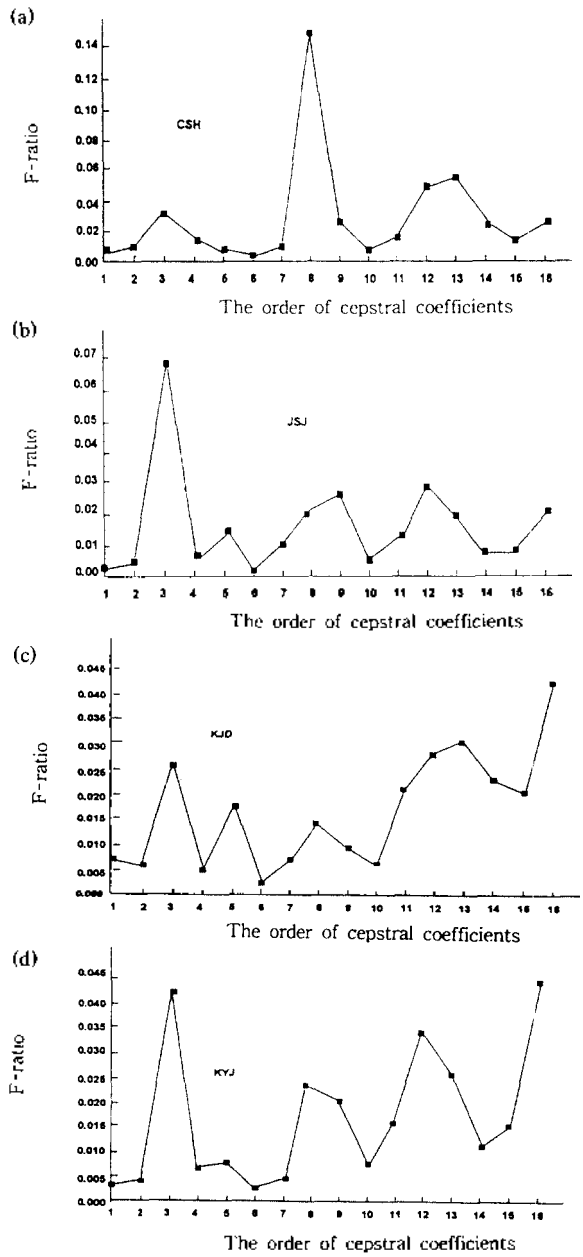
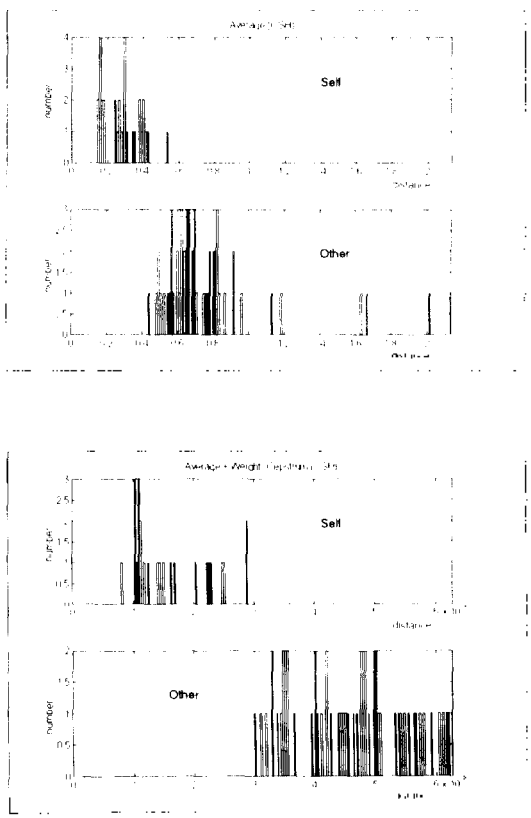


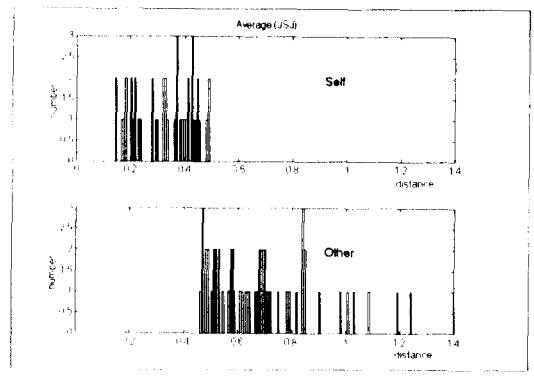

Figure 4.2 The distribution of each speaker's F-ratio values to the cepstrum order.

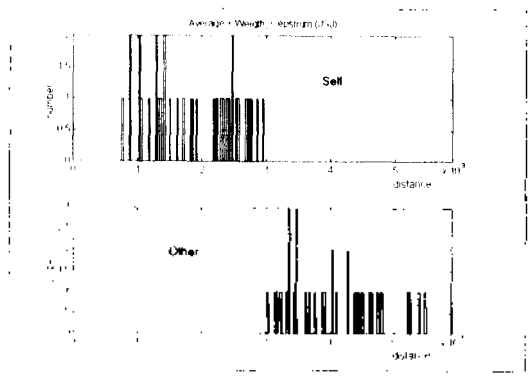
by comparing the distance of the input speech pattern and the reference pattern with the threshold value. That is, speaker verification requires comparing the test pattern against one reference pattern and involves a binary decision whether the test speech matches the template of the claimed speaker. The 2 types of errors are possible, the false rejection of customer and the false acceptance of imposter. We denote the probability of false rejection as P(N|s) and the probability of false acceptance as P(S|n). Generally, the threshold is selected by the ratio of both error. The case of too high a threshold value, the error rate of false acceptance decreases but the error rate of false rejection increases. On the contrary, the case of too low a threshold value, the error rate of false acceptance increases but the error rate of false rejection decreases. In consequence, it is able to select the lower threshold value or the higher threshold value according to the application. Mostly, the threshold value is determined as P(S|n) and P(N|s) have the same value.

In this paper, we determine the threshold value so that the both error rates have the same value. The example of setting the threshold value for each speaker is shown in Figure 4.3. We can see that the weighted cepstrum using F-ratio value is very effective parameter in Figure 4.3.

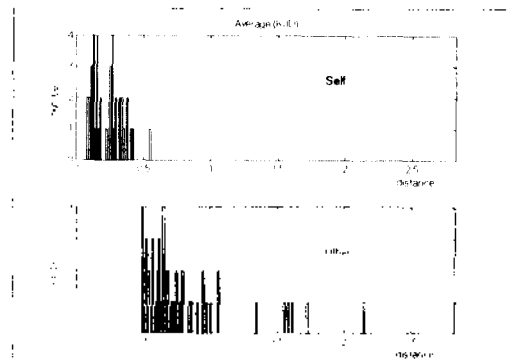


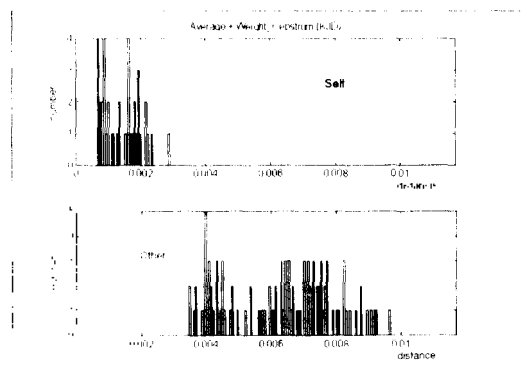

(a)



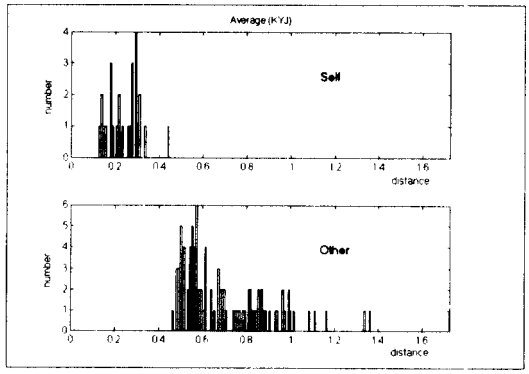



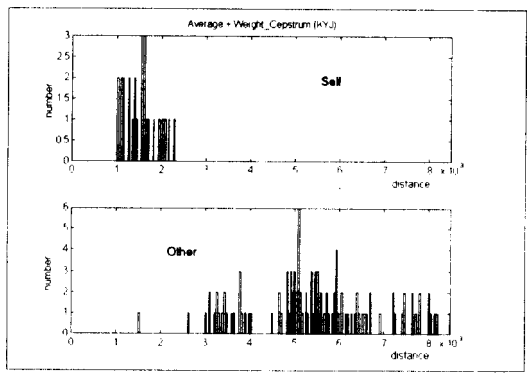

(b)




(c)




(d)

Figure 4.3 Establishment of the threshold value for each customer.
a)customer1 b)customer2 c)customer3 d)customer4

The Figure 4.3 a) upper shows the distance values which are computed using the general mel-scaled cepstra. The lower figure shows the distance values which are computed using the weighted cepstrum proposed in this paper. As shown in Figure 4.3 (a)-(d) lower, we can easily determine the threshold value which tell customers from imposters.

## V. OVERALL CONFIGURATION OF DOOR LOCK SYSTEM

We implement the door lock system based on pattern matching technique for speaker recognition using DTW. The Figure 5.1 shows the hardware configuration of our system.

The system hardware is composed of two parts: the door lock itself and the speaker recognition processor. We use an 8051 microprocessor in the door lock for serial communication with host processor.

## VI. EXPERIMENTAL RESULTS AND PERFORMANCE TESTS

In this section, we test the performance of speaker verification system and discuss the experimental results. The section 6.1 explains the speech DB and experimental environments. In section 6.2, we compare the proposed system with the traditional system.

### 6.1 DB and Experimental Environment

The speech DB consists of utterances extracted from ten speakers(male: 5, female: 5) in the laboratory environment. The customers are four persons(male: 2, female: 2) and the imposters are ten persons(male: 5, female: 5). All of the customers utter their names over a six-month period. Also, all of the imposters utter the customers' name over a six-month period. We choose the test patterns randomly from the speech DB.

The specification of the recording environment and the speech data are as follows.

▶ The environment of recording and testing: The common laboratory environment.

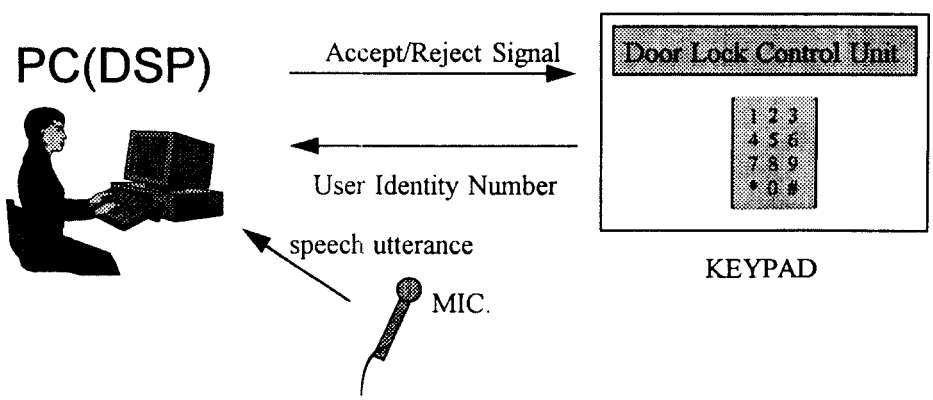



Figure 5.1 Overall door lock system

▶ A/D conversion : 8kHz sampling, 16 bit linear PCM. Use the ELF DSP Board of ASPI[8].

▶ The contents of speech DB : names consisting of 3 syllables.

▶ The number of total test utterances : 371.

In table 6.1, we show the generation of reference patterns for each speaker. We selected one or more reference patterns randomly out of the data over a six month period. Six patterns are selected for each speaker. We make the average reference pattern from these six patterns. The 'm' of customer name's parentheses means the male, the 'f' means the female.

Table 6.2 shows the distribution of test patterns.

Table 6.1 Construction of the reference pattern

| Customer | Month & No. of speech data for Average reference pattern | | | | | |
|---|---|---|---|---|---|---|
| | 1st Month | 2nd Month | 3rd Month | 4th Month | 5th Month | 6th Month |
| CSH(m) | | 2 | | | 2 | 2 |
| JSJ(f) | | | 2 | 1 | 1 | 2 |
| KJD(m) | 2 | 2 | | | | 2 |
| KYJ(f) | | 2 | | | 2 | 2 |

Table 6.2 Test patterns

a. Customer : CSH(82 data)

| Imposter | Month & No. of data | | | | | |
|---|---|---|---|---|---|---|
| | 1st Month | 2nd Month | 3rd Month | 4th Month | 5th Month | 6th Month |
| LYJ(m) | 3 | | 3 | 4 | 5 | |
| LCK(m) | | 4 | | | 6 | |
| SYH(m) | | 2 | | 3 | | 1 |
| KJD(m) | 2 | | 4 | | 5 | |
| JSJ(f) | | 6 | 3 | | | 5 |
| LMS1(f) | | 2 | | 3 | | 5 |
| LMS2(f) | 3 | | | 5 | | |
| JHS(f) | 5 | | 4 | | | |

b. Customer : JSJ(85 data)

| Imposter | Month & No. of data | | | | | |
|---|---|---|---|---|---|---|
| | 1st Month | 2nd Month | 3rd Month | 4th Month | 5th Month | 6th Month |
| LYJ(m) | 3 | | 3 | | | 5 |
| LCK(m) | | 3 | | | 5 | |
| SYH(m) | | 2 | 4 | | | |
| KJD(m) | 3 | | | 4 | | 5 |
| CSH(m) | 5 | | 3 | | 3 | |
| LMS1(f) | | 3 | | 4 | | 5 |
| LMS2(f) | 1 | 4 | | | | |
| KYJ(f) | 3 | | 5 | | 2 | |
| JHS(f) | 4 | 3 | | | | |

c. Customer : KJD(97 data)

| Imposter | Month & No. of data | | | | | |
|---|---|---|---|---|---|---|
| | 1st Month | 2nd Month | 3rd Month | 4th Month | 5th Month | 6th Month |
| LYK(m) | | 4 | | | 5 | 3 |
| LKC(m) | 3 | | | | | |
| SYH(m) | 2 | 6 | 3 | | | |
| CSH(m) | 6 | | | | | |
| JSJ(f) | | 6 | | | 6 | 4 |
| LMS1(f) | 3 | | | | | |
| LMS2(f) | 6 | 5 | | | | |
| KYJ(f) | | | | 1 | | 4 |
| JHS(f) | 6 | | | | | |

d. Customer : KYJ(108 data)

| Imposter | Month & No. of data | | | | | |
|---|---|---|---|---|---|---|
| | 1st Month | 2nd Month | 3rd Month | 4th Month | 5th Month | 6th Month |
| LYK(m) | | 4 | | 5 | 2 | 2 |
| LKC(m) | | 4 | | | 3 | 1 |
| SYH(m) | 3 | 3 | 5 | | | |
| KJD(m) | 2 | | | | 5 | |
| CSH(m) | 6 | 5 | | | | |
| LMS1(f) | 3 | | 5 | 1 | | |
| LMS2(f) | 5 | | | | | |
| JSJ(f) | | | 4 | | 6 | 6 |
| JHS(f) | 5 | 9 | | | | |

## 6.2 The Performance Tests of the Speaker Verification System

This section shows the results which obtain from an average reference pattern, weighted cepstrum and both methods. We compare these results with the conventional method as following manner.

1) The recognition rate of the speaker verification system using the average reference pattern.

The results of the average reference pattern are shown in Figure 6.1. This figure shows the comparison results with the traditional system which have the $N$ reference patterns. We use test patterns in table 6.2.

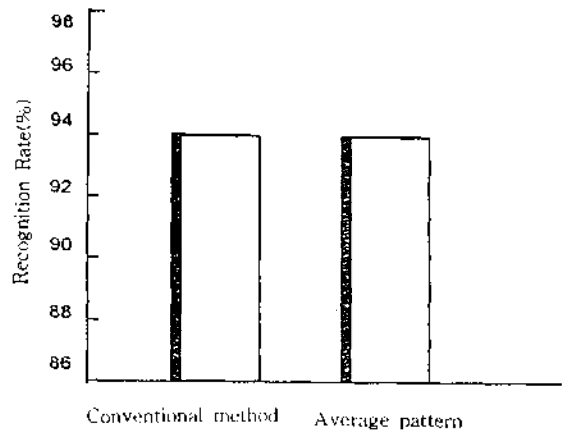



Figure 6.1 The result of use the one average pattern

In case of using the average reference pattern, the recognition rate is similar to the results of the traditional system. But the advantage is the low computational load comparing with the calculation of $N$ patterns.

2) The vecognition rate of the weighted cepstrum

We experiment the system with weighted cepstrum using F-ratio value for each speaker. The traditional system's recognition rate is about 94%. But in this experiment, we obtained the recognition rate at 98% using weighted cepstrum. Our system keeps the high recognition rate compared with the other systems whose performances is degraded rapidly as time goes on.

Figure 6.2 shows the experimental results of the weighted cepstrum.

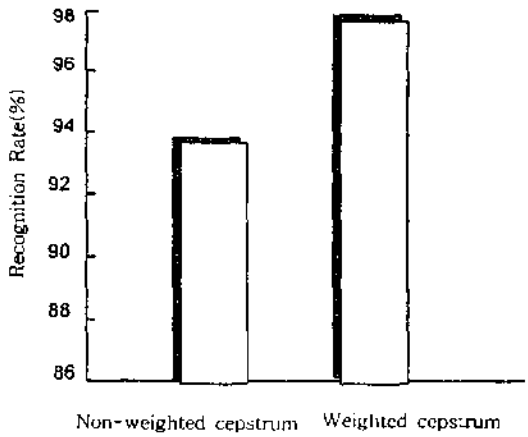



Figure 6.2 The result of the weighted cepstrum.

3) The recognition rate of the speaker verification system using both the average reference pattern and the weighted cepstrum.

We solve the heavy computational load on the traditional system by using the average reference pattern. Also the decrease problem of recognition rate as time goes on is solved by using the weighted cepstrum. Figure 6.3 shows the experimental results of using the traditional system, the average reference pattern, the weighted cepstrum and the both methods. We compare with the 4 methods in Figure 6.3. We confirm the high recognition rate that is obtained using the average reference pattern and the weighted cepstrum.

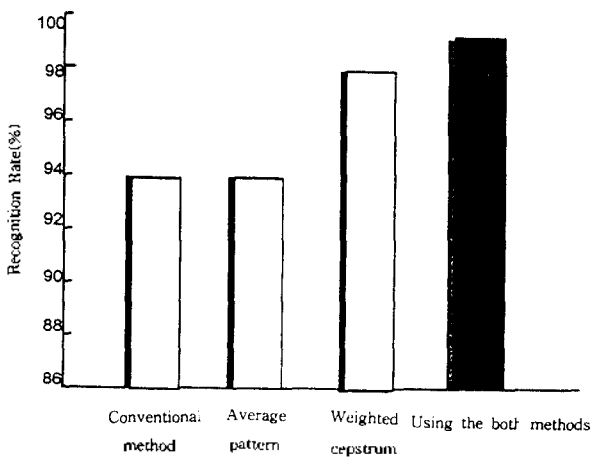



Figure 6.3 The result of the average pattern and the weighted cepstrum.

## Ⅵ. CONCLUSION

This paper improve the traditional text-dependent speaker verification system. The conventional system has the problem of heavy computational load and the performance degrade as time goes on. The suggested method which can improve the speaker verification system is made as follows. First, we make the average reference pattern from the several reference patterns by DTW. In this case, we weight more to the latest speech data. Second, we use the weighted cepstra to improve the recognition rate. The weighted values are obtained using F-ratio values. In the case of using the suggested two methods, the system's recognition rates are improved by about 5%-6%.

Conclusively, we implement the system according to the suggested method in this paper. The system is consisted of the door lock part and the speaker recognition processing part. The two parts are connected by serial port on PC. We use the general DSP board for high speed computation.

The further work includes the construction of stand-alone system.

## ACKNOWLEDGEMENT

This study was supported in part by the Korea Science and Engineering Foundation. The contract number is 95-0100-22-01-3.

## REFERENCES

1. H. S. Lee, "Speaker Recognition Technique," *Proc. of the Speech Comm. & Signal Processing Workshop*, pp. 42-46, Seoul, Korea, Aug., 1995.
2. H. Sakoe & S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. ASSP*, Vol. 26, pp. 43-49, Feb., 1978.
3. S. Furui & A. E. Rosenberg, "Experimental Studies in a New Automatic Speaker Verification System Using Telephone Speech," *Proc. of ICASSP*, pp. 1060-1062, 1980.
4. H. Ney & R. Gierloff, "Speaker Recognition Using a Feature Weighting Technique," *Proc. of ICASSP*, pp. 1645-1648, 1982.
5. J. Naik & G. Doddington, "High Performance Speaker Verification Using Principal Spectral Components," *Proc. of ICASSP*, pp. 881-884, 1986.
6. J. S. Jung et. al., "A study on the performance improvement of speaker recognition using average pattern and weighted cepstrum," *Proc. of the Speech Comm. & Signal Processing*

*Workshop*, pp. 179-183, Seoul, Korea, Aug., 1995.
7. D. O'Shaughnessy, "Speaker Recognition," *IEEE ASSP Magazine*, pp. 4-17, Oct., 1986.
8. *Elf DSP Platform Instruction Manual*, Atlanta Signal Processors Inc., 1993.

▲Youn Jeong Kyung

Youn Jeong Kyung was born in Korea on October 1, 1970. She received the B.S. degree in CS from Dongduck Women's University in 1992. She got the M.S. degree in 1994. She is currently enrolled in a Ph.D. degree of KAIST. Her research area includes the speech signal processing.

▲Jong Soon Jung

Jong Soon Jung was born in Kyungki-do, on Feb. 9, 1966. She received the B.S. degree in Electronics Engineering from Seoul Industrial University, in 1990, and the M. S. degree in Electronics Engineering from KAIST, in 1996. Since 1996 she has been with MunKyung Junior College, where she is a professor. Her research interests include speaker recognition.

▲Seung Ho Choi

Seung Ho Choi was born in Kyungki-do, on Feb. 8, 1969. He received the B.S. degree in Electronics Engineering from Hanyang University, Seoul, in 1991, and the M.S. degree in Electrical and Electronics Engineering from KAIST, Taejon, in 1993. He is currently pursuing his Ph.D. degree in Information and Communication Engineering at KAIST, Seoul. His research interests include speech recognition and speech coding.

▲Hwang-Soo Lee: Vol. 6, No. 3