Optimizing Feature Extraction for Speech Recognition

Chulhee Lee, Donghoon Hyun, Euisun Choi, Jinwook Go, and Chungyong Lee

***Abstract*—In this paper, we propose a method to minimize the loss of information during the feature extraction stage in speech recognition by optimizing the parameters of the mel-cepstrum transformation, a transform which is widely used in speech recognition. Typically, the mel-cepstrum is obtained by critical band filters whose characteristics play an important role in converting a speech signal into a sequence of vectors. First, we analyze the performance of the mel-cepstrum by changing the parameters of the filters such as shape, center frequency, and bandwidth. Then we propose an algorithm to optimize the parameters of the filters using the simplex method. Experiments with Korean digit words show that the recognition rate improved by about 4–7%.**

***Index Terms*—Critical band filters, feature extraction, mel-cepstrum, optimization, speech recognition.**

I. Introduction

Speech recognition can be roughly divided into two stages: feature extraction and classification. Although significant advances have been made in speech recognition technology, it is still a difficult problem to design a speech recognition system for speaker-independent, continuous speech. One of the fundamental questions is whether all of the information necessary to distinguish words is preserved during the feature extraction stage. If vital information is lost during this stage, the performance of the following classification stage is inherently crippled and can never measure up to human capability. Feature extraction can be understood as a step to reduce the dimensionality of the input data, a reduction which inevitably leads to some information loss. Typically, in speech recognition, we divide speech signals into frames and extract features from each frame. During feature extraction, speech signals are changed into a sequence of feature vectors. Then these vectors are transferred to the classification stage. For example, for the case of dynamic time warping (DTW), this sequence of feature vectors is compared with the reference data set. For the case of hidden Markov models (HMM), vector quantization may be applied to the feature vectors [1], [2], which can be viewed as a further step of feature extraction. In either case, information loss during the transition from speech signals to a sequence of feature vectors must be kept to a minimum.

There have been numerous efforts to develop good features for speech recognition in various circumstances [3]–[8]. Zahorian and Nossair proposed a pattern classification method for vowels using smoothed time/frequency features [6]. Their algorithm allows arbitrary nonlinear frequency, amplitude and time scales to represent the spectral and temporal characteristics of speech. Some researchers studied feature extraction for speech recognition [9]. For instance, Chengalvarayan and Deng proposed a method for dimensionality reduction in the feature space while preserving classification accuracies [5]. Biem and Katagiri proposed the discriminative feature extraction for cepstrum optimization [10], [11]. Gopinath investigated maximum likelihood modeling with some constraints and applied it to speech recognition [12]. Wand and Shamma studied spectro-temporal information in acoustic signals, providing insight into the human hearing mechanism [13].

Some of the most widely used features for speech recognition include LPC coefficients and the mel-cepstrum [14]. In particular, the mel-cepstrum, which is based on human auditory perception, has been used extensively for speech recognition [2], [14]. Although critical band filters are used to obtain the mel-cepstrum, it is not well understood how the parameters of the critical filters, such as center frequencies and bandwidths, affect the performance of speech recognition systems. More fundamentally, it is not clear whether all of the information necessary to distinguish between words is retained in the mel-cepstrum. The process of obtaining the mel-cepstrum is irreversible, i.e., we cannot reproduce speech signals from the mel-cepstrum. Thus, it is not possible for a human listener to determine whether the mel-cepstrum retains all of the information necessary to distinguish between words. In this paper, we first investigate how the center frequencies and bandwidths of the critical band filters influence the effectiveness of the mel-cepstrum and propose an algorithm to optimize the parameters.

II. Feature Extraction and Optimization

Typically, the mel-cepstrum is obtained using critical band filters as shown in Fig. 1 [2]. Thus, the parameters of the critical filters, such as center frequencies and bandwidths, may be optimized to reduce the error rate in speech recognition. The optimization of these parameters can be viewed as finding the minimum of a function whose inputs are the parameters of the filters and whose output is the error rate. Since the gradient of the function cannot be computed in this scenario, we propose to use the simplex method proposed by Nelder and Mead [15]. A simplex is a polyhedron consisting of $N+1$ points in an N dimensional space. For example, a simplex in a two-dimensional space is a triangle.

First, we briefly describe the simplex method, assuming that we want to find the minimum of a function [15], [16]. First, we

Manuscript received February 16, 2000; revised August 22, 2002. This work was supported in part by the Ministry of Information of Communication (MIC). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dirk van Compernolle.

The authors are with the Department of Electrical and Electronic Engineering, Yonsei University, 120-749 Seoul, Korea (e-mail: chulhee@yonsei.ac.kr).

Digital Object Identifier 10.1109/TSA.2002.805644

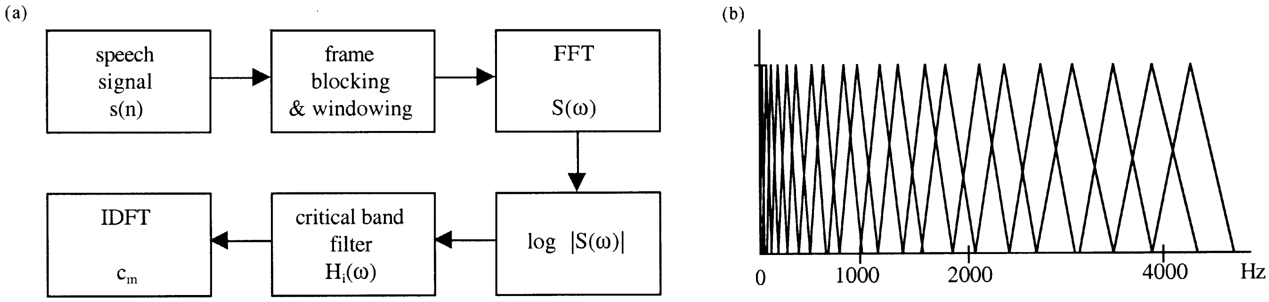


Fig. 1. (a) Block-diagram for obtaining the mel-cepstrum and (b) critical band filters for computing the mel-cepstrum.

need to create an initial simplex in the N-dimensional space. The $N+1$ points $(v_0, v_1, \ldots, v_n)$ that determine the initial simplex may be selected as follows:

$$v_i = v_0 + \alpha_i e_i, \quad i = 1, \ldots, N \tag{1}$$

$$\text{where} \begin{cases} e_1 = [1 \quad 0 \quad 0 \quad \ldots \quad 0]^T \\ e_2 = [0 \quad 1 \quad 0 \quad \ldots \quad 0]^T \\ \qquad \vdots \\ e_N = [0 \quad 0 \quad 0 \quad \ldots \quad 1]^T. \end{cases} \tag{2}$$

In the above equations, v_0 is the initial vector in the N dimensional space, $\{e_i\}$ is a basis and α_i is a constant that determines the length of each edge of the simplex. Then we compute values of the function on the corner points of the simplex and denote the maximum value as f_G and its corresponding point as G. Similarly, we denote the minimum and its corresponding point as f_S and S. Now we define

$$X = \frac{1}{N}\left(\sum_{i=0}^{N} v_i - G\right)$$

which can be considered as the center of a polyhedron consisting of all points except G. In order to move the simplex toward the region where the function has the minimum, we will replace G with a new point that can be obtained through three operations: reflection, expansion and contraction. If we denote the corresponding points as R, E, and C, respectively, they are obtained as follows:

$$\begin{aligned} R &= X + (X - G), \quad f_R = f(R) \\ E &= R + (X - G), \quad f_E = f(E) \\ C &= X + 0.5(G - X), \quad f_C = f(C) \end{aligned}$$

where f_R, f_E, and f_C, are the function values at the corresponding points. Another possibility is to move all of the points except S, which is called multiple contraction. In a multiple contraction, we update the simplex with the following points:

$$v_i = \frac{v_i + S}{2}, \quad i = 0, \ldots, N.$$

For a 2-D space, the above operations are illustrated in Fig. 2. Then, according to the rules in Fig. 3 [16], we select a new point and form a new simplex with the new points. We first compute the recognition accuracy at R and replace G with R if $f_R < f_G$. Then we compute the recognition accuracy at E and replace G with E if $f_E < f_R$. If $f_R \geq f_G$, we check a contraction and replace G with C if $f_C < f_G$. By repeating these operations, the

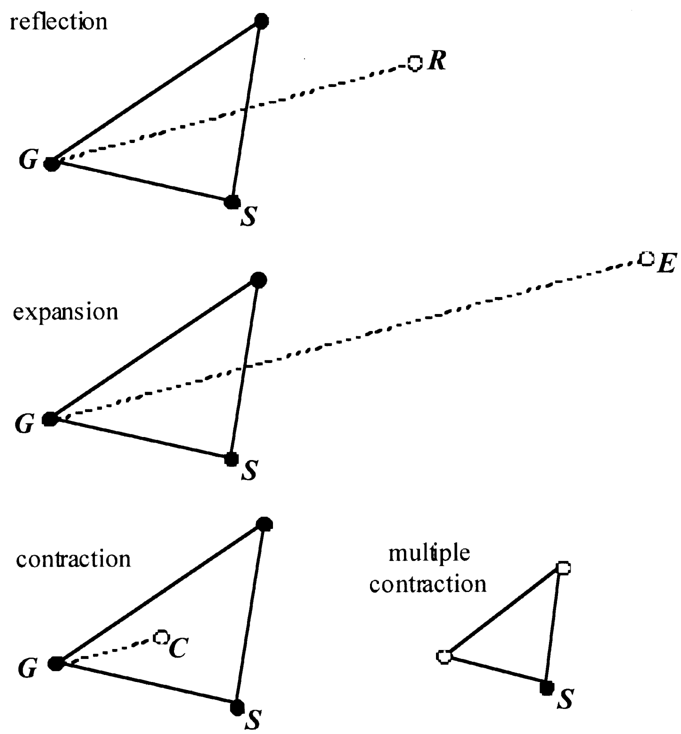


Fig. 2. Operations for changing a simplex.

simplex can be moved to the region that includes a minimum. The search stops if the edges of the simplex are smaller than a threshold

$$\max |v_i - v_j| < \varepsilon, \quad i, j = 1, \ldots, N$$

where ε is a constant.

III. Optimization of the Mel-Cepstrum

The optimization is performed assuming that the dynamic time warping algorithm (DTW) is used for speech recognition. Many researchers have proposed various local continuity constraints [13]. We first tried more than 10 local continuity constraints and chose the one that provided the best results for optimization. We used the Euclidean distance as a distance measure. However, it is noted that the proposed optimization method can be still used to optimize the performance of a system that uses a different local continuity constraint or a different distance measure.

The parameters of the critical band filter for the mel-cepstrum are shape, center frequencies, and bandwidths. We will optimize

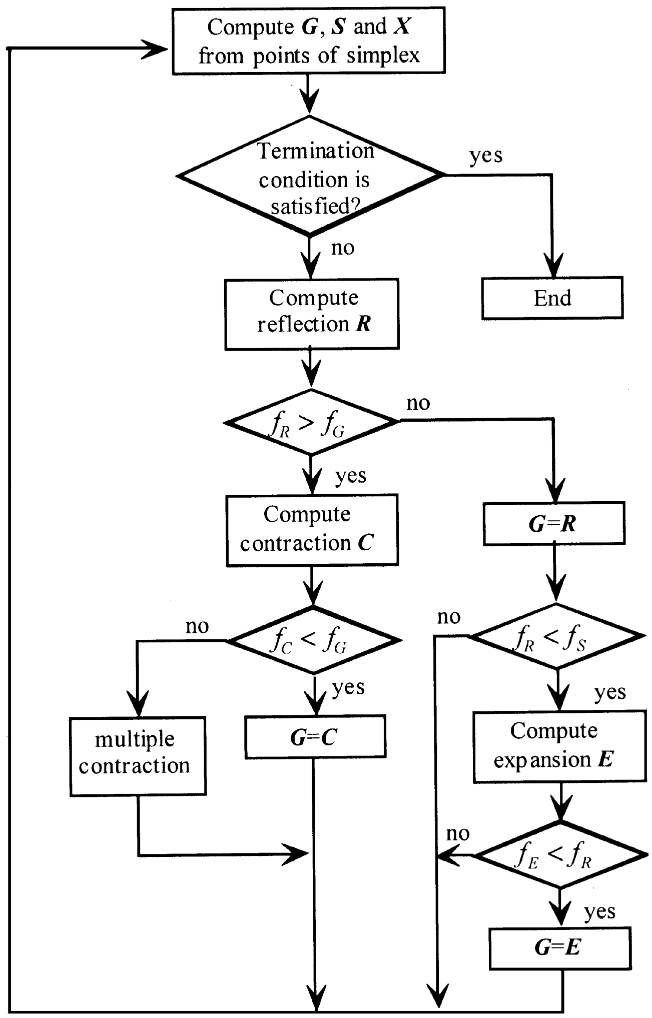


Fig. 3. Flowchart of the simplex method for finding a minimum.

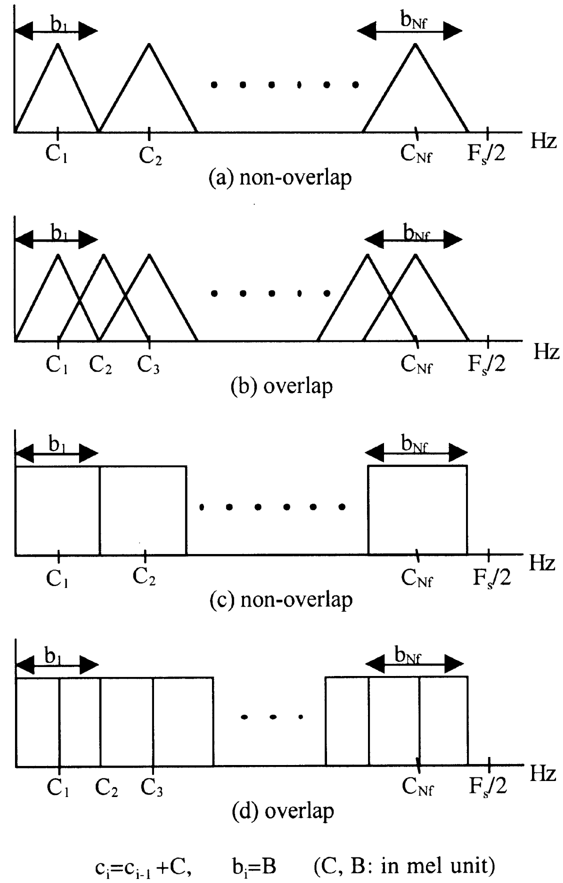


Fig. 4. Several critical band filters.

these parameters using the simplex method described in Section II. Assuming that the shapes of the filters are the same and there are N_f filters, there exist $2N_f$ parameters to adjust. We consider these parameters as the components of an input vector and the recognition error as an output of a function f. In other words, the input vector v is given by

$$v = [c_1, c_2, \ldots, c_{N_f}, b_1, b_2, \ldots, b_{N_f}]^T$$

where c_i and b_i are a center frequency and a bandwidth of the i-th filter respectively. In this paradigm, minimizing the recognition error is equivalent to finding the minimum of the function

$$f(c_1, c_2, \ldots, c_{N_f}, b_1, b_2, \ldots, b_{N_f})$$

where f is the classification error of a speech recognition system. In other words, we treat the classification error as a function of $c_1, c_2, \ldots, c_{N_f}, b_1, b_2, \ldots, b_{N_f}$. Since this f cannot be expressed in an analytical form, we use the simplex numerical method to find the minimum, as mentioned previously. It is noted that one may employ a gradient-based optimization procedure [10]. However, since the simplex method adjusts many factors simultaneously, it may converge faster to a solution.

In the simplex method, the location of the initial simplex is critical. If the initial simplex is located near the region that includes the minimum, the search can be done quickly. Otherwise the search can take a long time or may even fail in the worst case. In order to find a good initial simplex, we tested several filters with different shapes, different center spacings C and different bandwidths B, as shown in Fig. 4. The center frequencies and bandwidths are determined as follows:

$$\begin{aligned} c_i &= c_{i-1} + C (c_0 = 0) \\ b_i &= B \end{aligned} \tag{3}$$

where c_i and b_i are the center frequency and the bandwidth of the ith filter, respectively. It is noted that once the center spacing C and bandwidth B are given, the number of filters and the center frequencies are determined. The center spacing C of adjacent filters and the bandwidth B are in mel units.

IV. EXPERIMENTS AND RESULTS

We tested the proposed optimization algorithm with Korean digit words (0–9) using the dynamic time warping algorithm. There were 20 speakers (ten males and ten females). Each speaker spoke each digit ten times and the sampling frequency was 11.025 kHz. Since there are 20 speakers and each speaker spoke each digit ten times, there are a total of 2000 speech files. We used a part of the speech files as reference data and the remaining speech files as test data and for filter optimization. In order to evaluate the performance of the optimized filters when they are applied to completely new speech data, we also conducted experiments using the *leave-one-out* method [17].

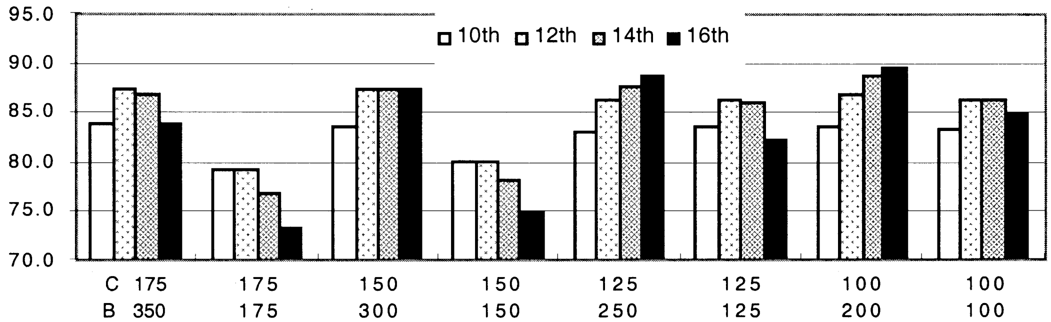


Fig. 5. Recognition rates of triangular filters with various (C, B)s (mel-cepstrum of the order of 10, 12, 14, and 16).

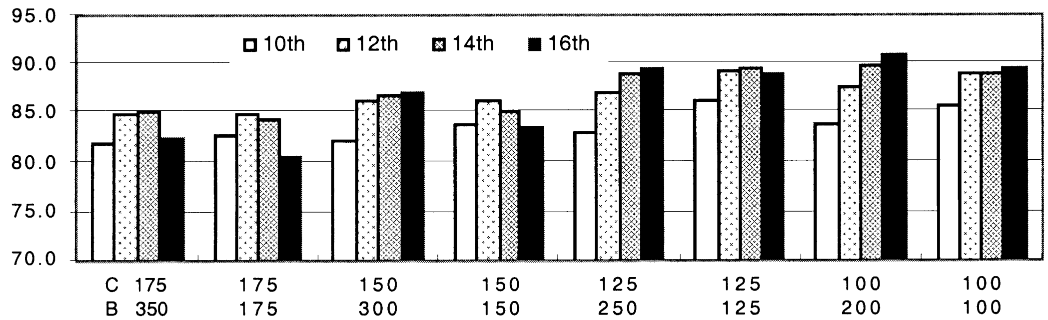


Fig. 6. Recognition rates of rectangular filters with various (C, B)s (mel-cepstrum of the order of 10, 12, 14, and 16).

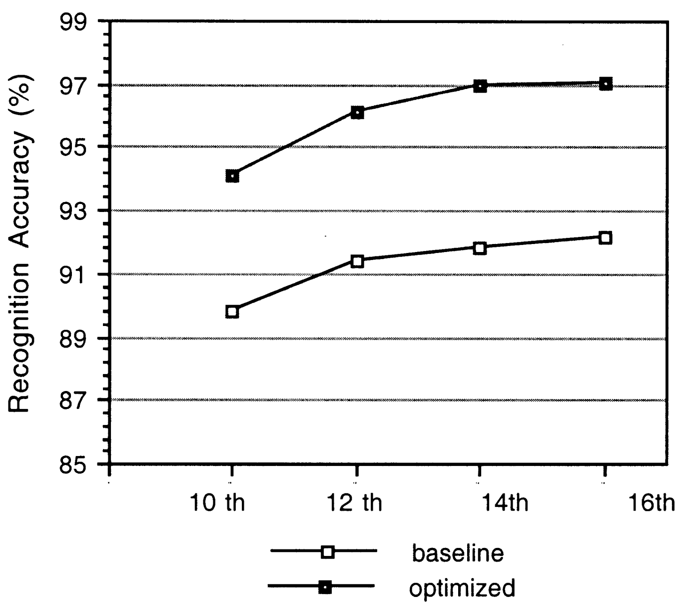


Fig. 7. Performance comparison after optimization.

In order to select a good initial simplex, we tested eight sets of critical band filters with the following center spacings and bandwidths (C, B):

$$(100, 100),\ (100, 200),\ (125, 125),\ (125, 150),\ (150, 150),\ (150, 300),\ (175, 175),\ (175, 350).$$

The actual center frequencies and bandwidths of all filters are determined using (3). Two filter shapes, triangular and rectangular, were tested. The recognition accuracies of the 16 filters (eight of triangular shape and eight of rectangular shape) are shown in Figs. 5 and 6. It appears that the rectangular shape provides better performances than the triangular shape. Based on these results, we chose the center spacing and bandwidth of (100, 100) to build the initial vector v_0 of (1)

$$v_0 = [100, 200, \ldots, 2300, 2400, 100, 100, \ldots, 100, 100]^T. \tag{4}$$

TABLE I
THE MINIMUM, MAXIMUM, AVERAGE, AND STANDARD DEVIATION OF THE IMPROVED RECOGNITION RATES OF THE 35 TRIALS USING THE FILTERS OPTIMIZED THE REFERENCE DATA OF 6 SPEAKERS

	Max	Min	Ave	Sd
10^{th}	5.4	1.3	3.1	1.1
12^{th}	5.8	-0.1	3.0	1.4
14^{th}	6.3	0.0	3.1	1.3
16^{th}	6.2	0.5	3.2	1.2

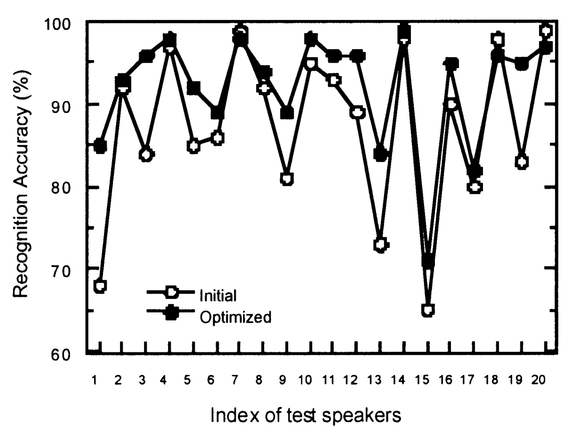


Fig. 8. Improved recognition rates of the 20 leave-one-out experiments.

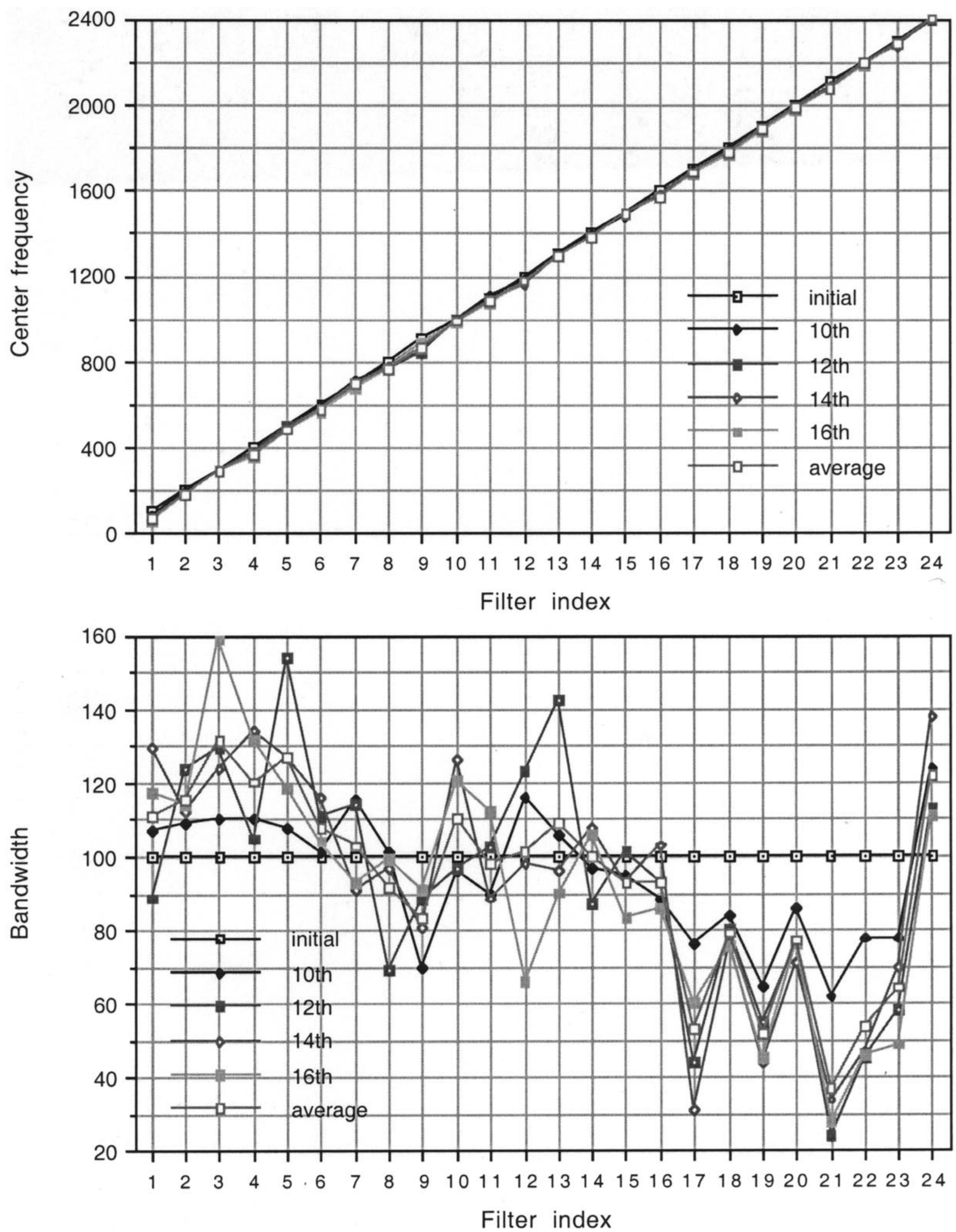


Fig. 9. Mel-center frequencies and mel-bandwidths of the optimized filters (in mel units).

Although the center spacing and bandwidth of (100, 200) provided better performances for mel-cepstrum using 14 and 16 coefficients, the center spacing and bandwidth of (100, 100) provided more uniform performances. In order to construct the initial simplex, we need N additional points. The remaining points of the initial simplex were computed as follows:

$$v_i = v_0 + 60e_i, \quad i = 1, 2, \ldots, N_f$$
$$v_i = v_0 + 150e_i, \quad i = N_f + 1, \ldots, 2N_f$$

where $\{e_i\}$ is given as in (2). Then we optimized these parameters using the simplex method. Although one may choose other values for the initial simplex, care should be taken in selecting the initial simplex. If the hyper-volume of the initial simplex is too small, it may take a long time to converge to an optimal point or the simplex can be stuck in a local minima. If the hyper-volume is too large, there can be a large variation within the simplex (several local minima and maxima), which is not desirable. In this paper, we tried several values and chose the above values since they provided the best performance in the test.

In the following experiments, the speech data of six speakers were chosen for reference data (half male and half female) and the remaining data, which include the unused data of the reference speakers, were used for the recognition test. In other words,

TABLE II
CENTER FREQUENCIES AND BANDWIDTHS OF THE OPTIMIZED FILTERS (MEL UNITS)

center freq	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
initial	100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200	2300	2400
10th MEL	89	185	298	378	486	588	703	776	841	997	1102	1190	1300	1393	1484	1577	1695	1785	1889	1992	2074	2194	2293	2395
12th MEL	64	181	299	376	502	567	700	769	844	997	1089	1176	1293	1384	1490	1573	1683	1775	1879	1993	2070	2195	2288	2394
14th MEL	77	180	295	373	490	578	689	777	882	989	1080	1153	1294	1375	1492	1574	1686	1777	1882	1986	2080	2199	2277	2402
16th MEL	58	181	297	362	479	573	678	767	880	981	1075	1178	1291	1374	1495	1565	1678	1768	1877	1983	2069	2191	2284	2398
average	72	182	297	372	489	577	693	772	862	991	1087	1174	1295	1382	1490	1572	1686	1776	1882	1989	2073	2195	2286	2397

bandwidth	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
initial	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
10th	107	109	110	110	108	101	115	101	70	96	90	116	106	97	95	88	76	84	65	86	62	78	78	124
12th	89	124	129	105	154	111	114	69	88	97	102	123	142	87	101	93	44	80	55	76	24	45	58	113
14th	129	112	124	134	127	116	91	97	81	126	89	98	96	108	93	103	31	76	44	71	34	47	70	138
16th	117	114	159	131	118	104	93	99	91	121	112	66	90	106	83	86	60	75	45	76	28	46	49	111
average	111	115	131	120	127	108	103	92	83	110	98	101	109	100	93	93	53	79	52	77	37	54	64	122

out of the 2000 speech files, 60 speech files were used for reference data and the remaining 1940 speech files were used for optimizing the filters. We optimized mel-cepstrum using 10, 12, 14, and 16 coefficients. The results are shown in Fig. 7. The baseline filters are given by (4). The recognition rates increased by 4.3–4.9%. It can be seen that the recognition rates improved significantly, indicating that some vital information was lost previously during the feature extraction stage.

Finally, we need to examine whether this improvement is limited to the specific reference data or global. In other words, in order to examine the improvement is limited to the specific reference and test data or global, we selected reference data from six different speakers and applied the optimized filters. We repeated the experiment 35 times with different reference data and the averages of the 35 trials are shown in Table I. As can be seen, consistent improvements are observed for the optimized filters, indicating that the optimization is global. Table I also shows the minimum, maximum, and standard deviation of the 35 trials when the optimized filters are used. It can be seen that significant and consistent improvements were obtained through the proposed optimization procedure. We also performed the optimization for the reference data consisting of 2 and 4 speakers and obtained similar results. Although the detailed experimental results are not included in this paper, the recognition accuracy increased by 4.2–7.0%.

In some sense, the above optimization is biased since the same data were used for the optimization and test. In other words, the optimization is performed for a specific set of data. In the following experiments, the *leave-one-out* method was used to eliminate this bias. In the *leave-one-out* method, the reference data were chosen from six speakers (three males and three females) and the remaining data except for one speaker were used for optimizing the filters. In other words, speech data from 19 speakers were used for optimization. After optimization, the optimized filters were tested on the speech data of the speaker that was not used for optimization. We repeated this experiment 20 times, excluding a different speaker each time. In these experiments, we optimized 14 coefficients of the mel-cepstrum. Fig. 8 shows the improved recognition accuracies of the 20 experiments. The average improvement of recognition accuracy was 4.8%. As can be seen, consistent improvements were observed for the data that were not used for optimization.

Table II and Fig. 9 show the final center frequencies and bandwidths of the optimized filters. Although the curves in Fig. 9 are noisy, some global characteristics are still visible. The center frequencies did not change significantly. However, there are noticeable changes in the bandwidths. In particular, the bandwidths of the optimized filters for the high frequencies (filter index 17–23) decreased significantly. The optimized filters for the low frequencies (filter index 2–6) have wider bandwidths than the initial filters, though there are some exceptions. The optimized filters in mid frequencies (filter index 7–16) show somewhat mixed results. These results indicate that filters with fixed center frequencies and bandwidths might not retain all the information necessary to distinguish between words. Well-designed filters may provide critical information for some difficult words.

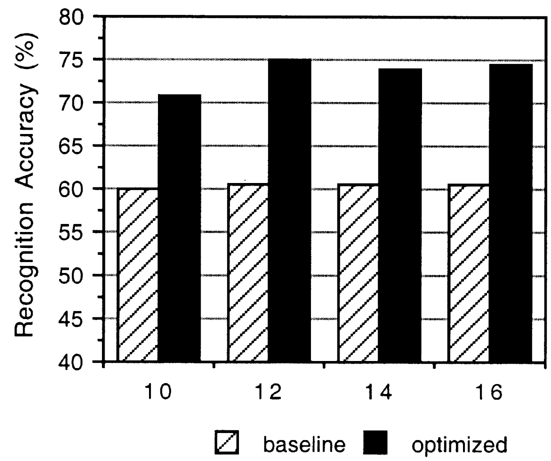


Fig. 10. Performance comparison of HMM systems. Ten speakers are used for training and the remaining ten speakers are used for test.

Finally, in order to examine whether the optimized filters can improve the performance of an HMM based system, we applied the optimized filters to an HMM system. However, due to the small size of speech data, the experiments are rather limited. In the first experiment, we used ten speakers (five males and five females) for training using the baseline filters. After the training, the remaining ten speakers were used for test. The same experiment was performed for the optimized filters. Fig. 10 shows the performance comparison. As can be seen, a 10–15% improvement was observed. In the next experiment, we used 19 speakers for training and the remaining speaker was used for testing. The experiment was repeated 4 times with a different speaker ex-

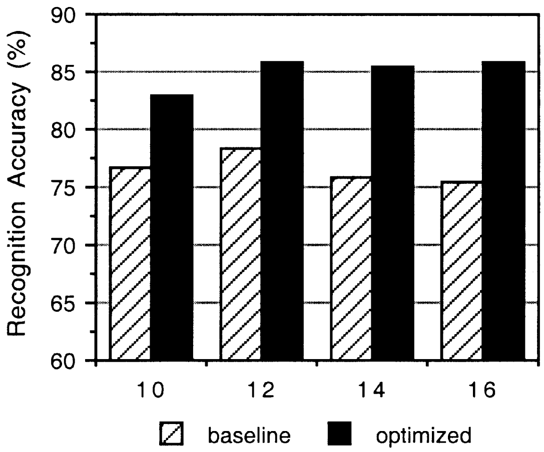


Fig. 11. Performance comparison of HMM systems. Nineteen speakers are used for training and the remaining speaker is used for test. The experiment was repeated four times and the average is shown.

cluded each time. Fig. 11 shows the average recognition accuracies for the baseline and optimized filters. In this case, a 6–10% improvement was observed. From these experiments, it appears that the optimized filters retain more information relevant for recognition from the original speech signal than the classical mel-cepstrum.

To conclude this section, we briefly discuss the complexity of the proposed algorithm. The complexity of the pre-operation, which was needed for the selection of the initial simplex, is negligible compared to that of the optimization procedure. The average number of iterations for the optimization is about 190. In other words, we had to classify the data set about 190 times for each optimization.

V. Conclusion

In this paper, we proposed an algorithm to minimize the information loss when transforming the speech signal to a sequence of feature vectors. By applying the algorithm to the mel-cepstrum, we were able to improve the recognition rate by 4–7%, which indicates that some of the vital information had been lost previously. In particular, in order to optimize the parameters of the mel-cepstrum, the simplex numerical method was used. Since this kind of problem cannot be solved analytically, it would be also possible to find a solution using a gradient-based optimization procedure. However, it is reported that the simplex method often finds a solution more rapidly and efficiently, though the performance may vary depending on applications. From our experiments, it can be said that the simplex method is an efficient tool to optimize the parameters of the filter bank.

The proposed algorithm provides a means to extract the maximum information for a given problem. The proposed algorithm can be applied to other speech recognition systems such as HMM, though the optimization may take a much longer time.

References

[1] Y. Zhang, R. Togneri, and M. Alder, "Phoneme-based vector quantization in a discrete HMM speech recognizer," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 26–32, Jan. 1997.

[2] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. New York: Macmillan, 1993.

[3] S. Moon and J. Hwang, "Robust speech recognition based on joint model and feature space optimization of hidden Markov models," *IEEE Trans. Neural Networks*, vol. 8, pp. 194–204, Mar. 1997.

[4] K. Wang, C. Lee, and B. Juang, "Selective feature extraction via signal decomposition," *IEEE Signal Processing Lett.*, vol. 4, pp. 8–11, 1997.

[5] R. Chengalvarayan and L. Deng, "HMM-based speech recognition using state-dependent, discriminatively derived transforms on mel-warped DFT features," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 243–256, May 1997.

[6] S. A. Zahorian and Z. B. Nossair, "A partitioned neural network approach for vowel classification using smoothed time/frequency features," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 414–425, July 1999.

[7] S. Chang, Y. Kwon, and S. Yang, "Speech feature extracted from adaptive wavelet for speech recognition," *Electron. Lett.*, vol. 34, pp. 2211–2213, 1998.

[8] D. J. Mashao, Y. Gotoh, and H. F. Silverman, "Analysis of LPC/DFT features for an HMM-based alphadigit recognizer," *IEEE Signal Processing Lett.*, vol. 3, pp. 103–106, Apr. 1996.

[9] L. Gu and R. Liu, "The application of optimization in feature extraction of speech recognition," in *Proc. Int. Conf. Signal Processing*, 1996, pp. 745–748.

[10] A. Biem and S. Katagiri, "Cepstrum-based filter-bank design using discriminative feature extraction training at various levels," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1997, pp. 1503–1506.

[11] ——, "Filter bank design based on discriminative feature extraction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1994, pp. 485–489.

[12] R. A. Gopinath, "Maximum likelihood modeling with Gaussian distributions for classification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 1998, pp. 661–664.

[13] K. Wang and S. Shamma, "Auditory analysis of spectro-temporal information in acoustic signals," *IEEE Eng. Med. Biol.*, vol. 14, pp. 186–194, Mar./Apr. 1995.

[14] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[15] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*. Cambridge, U.K.: Cambridge Univ. Press, 1986.

[16] J. L. Buchanan and P. R. Turner, *Numerical Methods and Analysis*. New York: McGraw-Hill, 1992.

[17] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.

Chulhee Lee received the B.S. and M.S degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1984 and 1986, respectively, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1992.

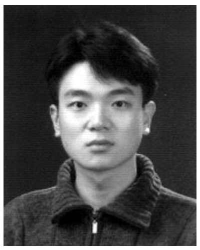
From 1986 to 1987, he was a Researcher in the Acoustic Laboratory at Technical University of Denmark (DTH). From 1993 to 1996, he worked with National Institutes of Health, Bethesda, MD. In 1996, he joined the faculty of the Department of Electrical and Computer Engineering, Yonsei University, Seoul. His research interests include image/signal processing, pattern recognition, and neural networks.

Dr. Lee is a member of Tau Beta Pi, Eta Kappa Nu, and KSEA.

Donghoon Hyun received the B.S. and M.S. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1997 and 1999, respectively.

He is currently developing GPRS mobile phones at Samsung Electronics.

Euisun Choi received the B.S. and M.S. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1998 and 2000, respectively. He is currently pursuing the Ph.D. degree in electronic engineering.



Jinwook Go received the B.S. degree in electronics engineering from Konkuk University, Seoul, Korea, in 1996 and the M.S. degree in electronics engineering from Yonsei University, Seoul, in 1999. He is currently pursuing the Ph.D. degree in electrical and electronic engineering at Yonsei University.

His current research interests include speech recognition, pattern recognition, and neural networks.

Chungyong Lee received the B.S. and M.S. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1987 and 1989, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 1995.

He was a Senior Engineer of Samsung Electronics Co., Ltd., Kiheung, Korea, from 1996 to 1997. In 1997, he joined the faculty of the Department of Electrical and Electronic Engineering, Yonsei University, where he is currently an Associate Professor. His research interests include array signal processing and communication signal processing.