FAST SPEECH RECOGNITION ALGORITHM UNDER NOISY ENVIRONMENT USING MODIFIED CMS-PMC AND IMPROVED IDMM+SQ

Hiroki Yamamoto, Tetsuo Kosaka, Masayuki Yamada, Yasuhiro Komori and Minoru Fujita

Media Technology Laboratory, Canon Inc.
890-12 Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa 211 Japan,
E-mail:hiroki@cis.canon.co.jp

ABSTRACT

In this paper, we describe a fast speech recognition algorithm under noisy environment. To achieve an accurate and fast speech recognition under noisy environment, a very fast speech recognition algorithm with well-adapted model against the noisy environment is required. First, for the model adaptation, we propose MCMS-PMC: a combination of the parallel model combination(PMC) and the modified cepstral mean subtraction(MCMS) which estimates the cepstrum mean by taking account of the additive noise. Then, for the fast speech recognition, we propose new techniques to create the noise-adapted scalar quantized codebook in order to introduce the MCMS-PMC into the IDMM+SQ which we proposed in ICASSP96 as fast speech recognition algorithm using scalar quantization approach. Finally, an effect of proposed method is shown through the speaker-independent telephone-bandwidth continuous speech recognition experiment.

1. INTRODUCTION

To realize a real-world speech recognizer, a real-time speech recognizer with high accuracy against the real-word environment is indispensable. This paper focuses on a fast speech recognition algorithm under noisy environment.

Generally, in order to achieve high accuracy under noisy environment, the speed of speech recognition gets slower. This derives from the use of detailed models which requires lots of computation and also from degradation of the beam-search mechanism according to the miss-match between the model and the environment. Thus, to achieve an accurate and fast speech recognition under noisy environment, a very fast speech recognition algorithm with well-adapted model against noisy environment is required. In this paper, first a new method for model adaptation against noisy environment, the modified CMS-PMC, is proposed. Then improvement on the fast speech recognition algorithm, IDMM+SQ [4] against noisy environment is discussed.

The primary factors of the variation of the environment are the additive noise and the channel distortion. The PMC [5] is well-known as a good mechanism for additive noise compensation and the CMS [6] is known as an effective method for channel distortion compensation. Thus, a combination of the PMC and the CMS seems to be an effective solution to adapt both additive noise and channel distortion. However, a simple combination of the PMC and the CMS has a problem against the low SNR noisy environment. Here, we propose a new HMM environment adaptation method named modified CMS-PMC(MCMS-PMC).

As for the fast speech recognition algorithm, many techniques were reported [1, 2, 3, 4]. All these methods lessen the cost of the output probability computation. These methods are realized by the table looking-up by means of scalar quantization (SQ) [1] or the calculation of the strict probability on the probable states[2, 3]. The IDMM+SQ[4] that we proposed in ICASSP96 uses these two techniques in effective. All these techniques attained good results under clean environment, however they were not examined against noisy environment. In this paper, we introduced the IDMM+SQ into the noisy environment speech recognition. In the IDMM+SQ, the SQ codebook(SQC) were created from clean speech which will not well-match to the noise environment. Thus, we propose a new technique to create the noise-adapted SQC and improve the IDMM+SQ against noisy environment.

With the modified CMS-PMC and the improved IDMM+SQ, a very fast real-word speech recognizer can be realize because: a) MCMS-PMC realizes a high accuracy recognizer against noisy environment. b) The IDMM+SQ realizes a very fast speech recognizer. c) Moreover, an efficient beam-search is achieved because the MCMS-PMC creates well-matched model against noisy environment.

In this paper, first the algorithm of the MCMS-PMC is proposed. Then the overview of the IDMM+SQ and its improvement are described. Evaluations on the telephone bandwidth speech, that showed the effectiveness of the proposed method, are also discussed.

Copyright 1997 IEEE

2. HMM ADAPTATION

It is well-known that Cepstral Mean Subtraction (CMS) is one of the accurate methods of the channel normalization. Parallel Model Combination (PMC) has been recently proposed for the additive noise compensation. A combination of the CMS and the PMC is expected to solve the problem of both additive noise and channel distortion. Such a simple combination, however, is not applicable when the SNR is low because the additive noise causes estimation errors. In this paper we propose a new environment adaptation method named "modified CMS-PMC (MCMS-PMC)" to compensate both additive noise and channel distortion in the low SNR conditions.

2.1. MCMS-PMC

Figure 1 shows the blockdiagram of the MCMS-PMC method. The method is characterized by the estimation of the channel condition by taking account of the additive noise. First, the noise HMM of 1 state 1 mixture, is estimated in the cepstral domain from the input noise data. The cepstral mean(CM) with the additive noise is also calculated from the input speech data. The CM is compensated in the linear spectral domain as $\overline{x'}^{lin} = \overline{x}^{lin} - k_1\overline{n}^{lin}$ where lin means the linear domain. Next, the channel normalization on the mean parameters of HMMs derived from clean speech are performed as $y(t)^{cep} - \overline{y}^{cep} + \overline{x'}^{cep}$ where cep means cepstral domain. Finally, the PMC is carried out to eliminate the influence of the additive noise.

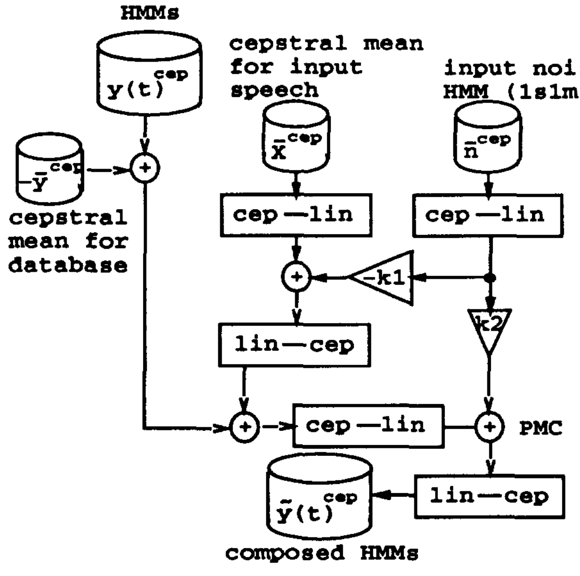
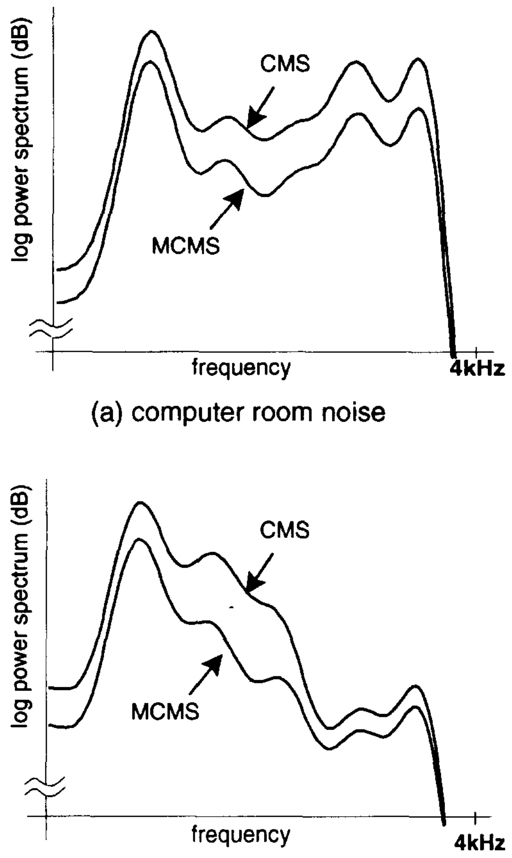


Figure 1. MCMS-PMC



(a) computer room noise

(b) colored noise

Figure 2. Spectral expression of the cepstral means used in the CMS and the MCMS

Figure 2 (a) shows the spectrum expression of the CMs with computer room noise, like white noise, used in the CMS and the MCMS process, (b) shows those with colored(non-white) noise, which low-pass filtered computer room noise. In the figure (a), the difference appears only on the bias and the shapes of those are almost the same. In contrast, the shapes of spectrum of (b) show the obvious difference. The effects are the same when the additive noise is white. Because the cepstral mean estimation errors, which are caused by white noise, influence equally on the whole frequency as the bias in the linear spectral domain. The modified part of the MCMS mechanism exhibits its effect against colored noise. The MCMS is an extended method of the CMS, which works either on the white and colored noise. From the above discussion, we can say that the MCMS is more effective than the CMS for noise compensation.

3. IMPROVEMENT OF IDMM+SQ

The IDMM+SQ is an algorithm for a fast output probability computation using the IDMM[1] and the SQ[2]. The IDMM is an approximate computation of the multi-mixture probability density function(pdf). De-

[1] Independent Dimension Multi-Mixture computation
[2] Scalar Quantization

Copyright 1997 IEEE

tails about the IDMM are described in [4]. The estimation error of the IDMM+SQ is compensated by the probability recalculation algorithm. The whole algorithm is as follows:

1. rough but quite fast computation of the output probability using the IDMM and the SQ,
2. finding some probable HMM states using the result of the rough computation in step 1,
3. recalculation of the strict probability on the probable states.

3.1. Problem

In the previous paper[4], the SQ codebook (SQC) was made from the training data(clean speech) using the LBG algorithm. However, the SQC won't match to noisy environment because distributions of acoustic feature will be shifted by the noise. Moreover, it is not realistic to create a new SQC according to noisy environment because the LBG algorithm requires huge computation and large data collection.

In the next section, we will discuss how to make a good SQC according to noisy environment.

3.2. Construction of SQC

Sagayama et al. previously proposed the way to create an SQC from the distributions of the HMMs[1]. By introducing this method, the SQC for noisy environment can be created from the environment-adapted HMMs.

Here we propose some new manner for the SQC construction. In the IDMM+SQ, the SQC is created in each dimension independently. The steps of making the SQC from the HMMs are as follows:

1. Decide the range of SQ (SQ-Range).
2. Quantize with in the SQ-Range.

We proposed two ideas for each step.

Decision of SQ-Range

R1: Calculate the SQ-Range using the variance information as follows:

$$[\min_i(\mu_i - 3\sigma_i), \max_i(\mu_i + 3\sigma_i)] \qquad (1)$$

where μ_i and σ_i^2 denotes the mean and the variance of the i-th distribution, respectively.

R2: Define the SQ-Range as equation2 by merging all the distributions into one distribution.

$$[\hat{\mu} - 3\hat{\sigma}, \hat{\mu} + 3\hat{\sigma}] \qquad (2)$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ denotes the mean and the variance of the merged distribution, respectively. The distributions are merged as follows:

$$\hat{\mu} = \frac{\sum_i^M \mu_i}{M}, \quad \hat{\sigma}^2 = \frac{\sum_i^M \sigma_i^2 + \sum_i^M (\mu_i - \hat{\mu})^2}{M}$$

where M denotes the number of distributions.

Calculation of code value.

D1: Divide the SQ-Range equally by the SQC size.

D2: Divide the SQ-Range in order to equalize each integral of *pdf* between neighbor code values.

Figure 3 shows the way of dividing SQ-Range.

Thus, 4 kinds of SQC can be realized:

SQC1: R1 + D1
SQC2: R1 + D2
SQC3: R2 + D1
SQC4: R2 + D2

The SQC1 is proposed by Sagayama[1].

4. EVALUATION

We evaluated the proposed method on a speaker-independent telephone bandwidth continuous speech recognition. First, the MCMS-PMC was evaluated. Second, the robustness of the IDMM+SQ with the MCMS-PMC against the noise was verified.

Conditions

We used 72,000 utterances of 200 speakers for HMM training. The test set consisted of 500 sentences of 10 speakers. The average duration of the test set was 2.7[sec/utterance]. The recognition grammar consisted of 1,004 words with 30.2 word perplexity. The test set were artificially created into noisy and channel distorted speech by adding colored noise and by applying telephone bandwidth digital filtering. The speech HMM was a right context dependent HMM. The noise HMM for the PMC and the cepstral mean coefficients were calculated from 1-sec noise samples and 5.9-sec speech samples, respectively.

4.1. Experiment 1

The following methods were compared: 1) no adaptation(NONE), 2) a combination of the CMS and the Spectral Subtraction (CMS-SS), 3) a simple combination of the CMS and the PMC (CMS-PMC) and 4) the proposed method (MCMS-PMC).

Results

Table 1 shows the sentence accuracy. The column '∞' shows the result on the clean speech data.

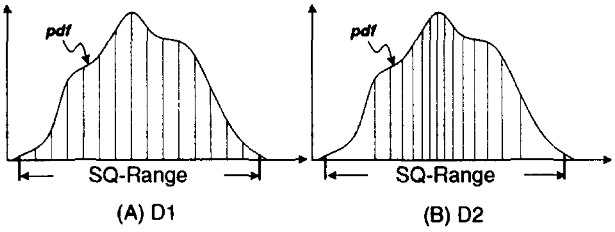


Figure 3. Dividing of the SQ-Range.

Copyright 1997 IEEE

The results show that the MCMS-PMC attained the best result. The difference between the MCMS-PMC and the CMS-PMC increases as the SNR becomes worse. This is because the estimation of the CM becomes worse by the additive noise.

Table 1. Comparison of Adaptation Methods (%)

Adaptation Method	SNR[dB]			
	∞	20	15	10
NONE	84.6	8.0	0.6	0.4
CMS-SS	—	61.6	45.8	27.4
CMS-PMC	—	79.8	74.2	64.6
MCMS-PMC	—	79.6	76.4	70.4

4.2. Experiment 2

At first in this experiment, we compared 5 types of the SQC: LBG, SQC1-4. LBG was made from the clean speech. Then performance of the IDMM+SQ with the MCMS-PMC was evaluated from a point of sentence accuracy and recognition time.

Results

Table 2 shows the sentence accuracy at SNR 10 [dB]. The codebook size was changed from 8 to 64. Most results using SQC made from the HMMs are better than those of using LBG, and the results of SQC4 are stable in every codebook size.

Table 3 shows the sentence accuracy and the recognition time using SQC4 at codebook size 64. The table also shows the results without IDMM+SQ as a baseline. The number in the bracket denotes the average recognition time[3] par one utterance. The IDMM+SQ reduces about 60% of recognition time while the degradation of accuracy is less than 0.8%.

Table 2. Comparison of SQC types (%)

SQC	codebook size			
	64	32	16	8
LBG	62.0	62.8	60.2	59.8
SQC1	69.6	67.2	62.0	49.4
SQC2	69.0	67.6	68.0	59.6
SQC3	69.4	69.4	67.2	60.0
SQC4	69.0	68.2	68.0	67.0

Table 3. Robustness of IDMM+SQ against noise (%)

	SNR[dB]		
	20	15	10
baseline (time[sec])	79.6 (14.12)	76.4 (14.30)	70.4 (14.51)
IDMM+SQ (time[sec])	80.0 (5.76)	75.6 (5.72)	69.6 (5.76)

Finally, we performed an experiment on telephone speech. In the experiment, we changed the amount of recalculation from 0% to 50%. Figure 4 shows the sentence accuracy and the recognition time. A dotted line denotes the result without IDMM+SQ as a baseline. The figure shows that IDMM+SQ can reduce 50% of the recognition time with no degradation of accuracy when the amount of recalculation is 27.4%.

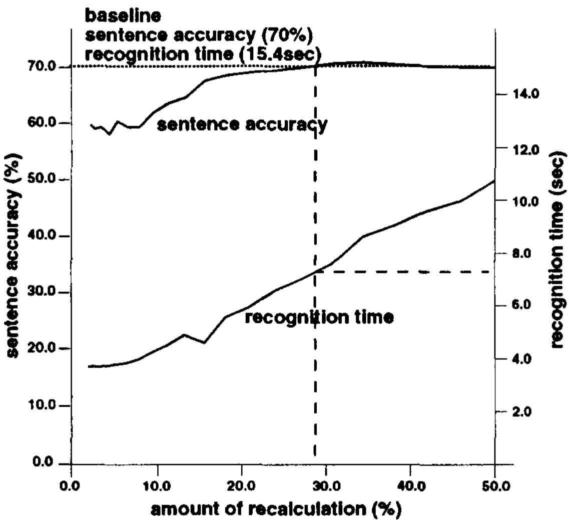


Figure 4. The sentence accuracy on telephone speech and the recognition time.

5. CONCLUSION

In this paper, we described a fast speech recognition algorithm under noisy environment. First, we proposed a new environment adaptation method: MCMS-PMC. Then, the improved IDMM+SQ was proposed whose SQC was made by new techniques. Experiments were carried out on the telephone bandwidth speech and showed the effectiveness of the MCMS-PMC. The combination of the improved IDMM+SQ and the MCMS-PMC achieved 60% reduction of recognition time with almost no degradation of accuracy.

REFERENCES

[1] Sagayama S. et al. : "On the Use of Scalar Quantization for Fast HMM Computation", ICASSP95, pp.213-216 (1995).

[2] Watanabe T. et al. : "High Speed Speech Recognition Using Tree-structured Probability Density Function", ICASSP95, pp.556-557 (1995).

[3] Bocchieri E. : "Vector Quantization for the Efficient Computation of Continuous Density Likelihoods", ICASSP93, II, pp. 692-69 (1993).

[4] Yamada M. et al.: "Fast output probability computation using scalar quantization and independent dimension multi-mixture," Proc. ICASSP 96, vol. II, pp.893-896 (1996).

[5] M.J.Gales, S.Young: An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise, IEEE, ICASSP'92, I-233-236, (1992).

[6] Rahim, et al.: Signal Bias Removal for Robust Telephone Based Speech Recognition in Adverse Environments, ICASSP94, (1994).

[3] The recognition time was measured by Sun Sparcstation 20(60MHz) using profiler (gprof).

Copyright 1997 IEEE