Fast Parallel Model Combination Noise Adaptation Processing

Yasuhiro KOMORI, Tetsuo KOSAKA, Hiroki YAMAMOTO,
and Masayuki YAMADA
Media Technology Laboratory, Canon Inc.
890-12 Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa 211 JAPAN,
Email:komori@cis.canon.co.jp

Abstract

In this paper, a fast PMC (Parallel Model Combination) noise adaptation method is proposed for continuous HMM base speech recognizer. The proposed method is realized as a direct reduction of the number of PMC processing times by introducing the distribution composition with the spatial relation of distributions. The proposed method is compared with the basic PMC algorithm in recognition accuracy and adaptation processing time on telephone speech. The result showed that the proposed method saved around 65% (70.9% - 62.7%) of PMC computation amount with almost no degradation of recognition performance.

1 Introduction

To realize a real-world speech recognizer, a speech recognizer with high accuracy against the real-word environment is indispensable. And, many researches on noise adaptation are now on-going[1, 2, 3, 4, 5].

The noise environment greatly differs according to the place where we use the speech recognizer. Thus, a very short-time or an instant noise adaptation method with high performance is required.

Among the researches, the PMC[1] appears to be a good noise adaptation method which enables to adapt all recognition models with short-time noise data. Moreover, the PMC results good recognition performance because it enables to adapt the cepstrum feature and its dynamic feature, furthermore not only the mean but the variance, either. On the other hand, however, the PMC method requires to perform computationally heavy Gaussian-Integration on all distributions of the recognition models. And to achieve good performance in noise environment, context-dependent HMM is used which requires huge computation amount on using PMC noise adaptation.

Previously, a way to reduce the PMC computation amount, data-driven PMC (DPMC), is proposed[3]. In this paper, we also propose a way to reduce the PMC computation amount, but it quite differs to the DPMC. The proposed method is based on distribution composition which reduces the number of PMC process in direct. The computation amount of the DPMC is able to reduce more, if our method is applied to the DPMC.

In this paper, a new fast processing method for PMC noise adaptation is proposed. Furthermore, experimental result on telephone speech using the proposed method is reported.

2 PMC Noise Adaptation

The PMC generates the cepstrum-based noise corrupted HMM from the noise HMM and the speech HMM. each of which is separately modeled.

Here are the advantages of the PMC method:

- Short time noise data (eg. 1 sec) is required for noise adaptation.
- Direct adaptation of cepstrum domain HMM is possible.
- Not only the mean (μ) but the variance (σ^2) of Gaussian distribution are able to be adapted.
- Not only the static parameter but also the delta parameter are able to be adapted .

The PMC is realized by the parameter composition of cepstrum-based noise HMM and cepstrum-based speech HMM by transforming the output distribution of the HMM from the cepstral domain to the linear spectral domain.

Figure 1 shows the block diagram of the PMC noise adaptation process. O^c is noise adapted HMM, S^c is the speech HMM, N^c is the noise HMM represented in the cepstrum domain. O^l, S^l, N^l are the HMM transformed into the log spectrum domain, respectively.

The PMC of the static parameter and the delta parameter are defined as follows:

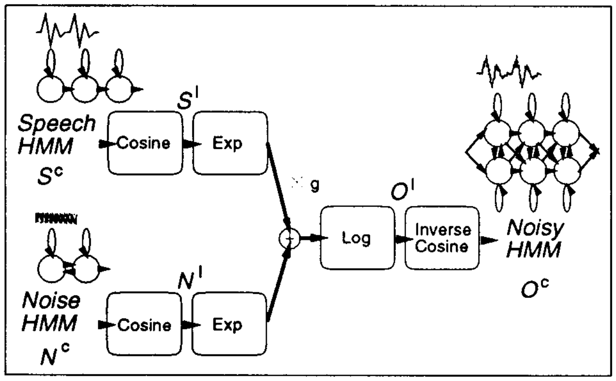

Figure 1: PMC Processing

PMC for Static Parameter

PMC for static parameter is realized by the next equation:

$$O_i^l = \log((g\exp(S_i^l(t))) + \exp(N_i^l(t)))$$

PMC for Delta Parameter

If the delta parameter is defined as:

$$\Delta \mathbf{O}^c(t) \;=\; \mathbf{O}^c(t-\omega) - \mathbf{O}^c(t+\omega)$$

then PMC for Delta parameter is realized by the next equation:

$$\begin{aligned} &\Delta O^l(t) = \\ &\quad \log(\exp(\Delta S_i^l(t) + S_i^l(t-\omega) + \log(g)) \\ &\quad + \exp(\Delta N_i^l(t) + N_i^l(t-\omega)) \\ &\quad - \log(\exp(S_i^l(t-\omega) + \log(g) + \exp(N_i^l(t-\omega)))) \end{aligned}$$

3 Fast Processing for PMC

In this section, the basic idea of the fast PMC method is first described, then the algorithm is explained.

3.1 Basic Idea

In order to realize a fast PMC noise adaptation, we make the following assumption:

> **"close distributions of the models are corrupted by noise in a close manner"**

And from this assumption, we lead the following sub-assumptions by introducing the distribution composition method:

- The relational position of the close distributions to the composite distribution will be kept before and after the noise corruption. In other words, the noise corrupted position of each distribution can be determined from the difference between the close distributions and the composite distribution before PMC by taking account of the area corruption.
- The area of the close distributions of the model is noise corrupted in the same manner to the area of a composite distribution of the close distributions. In other words, the noise corrupted area of each distribution can be determined from the area ratio of the composite distribution before PMC and after.

The image of the basic PMC noise adaptation is shown in figure 2 and the image of the proposed method is shown in figure 3. As it is shown in the figure, the PMC processing time is directly reduced. In the basic PMC, all distributions must perform the PMC-processing, while the proposed method requires a single PMC-processing par a composite distribution. Although there is an overhead computation for determining the spatial relation of distributions, the computational amount is lighter than that of Gaussian-Integration in the PMC process.

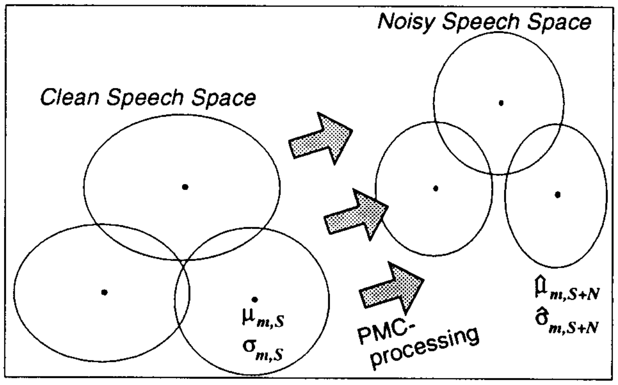

Figure 2: Image of Basic PMC Processing

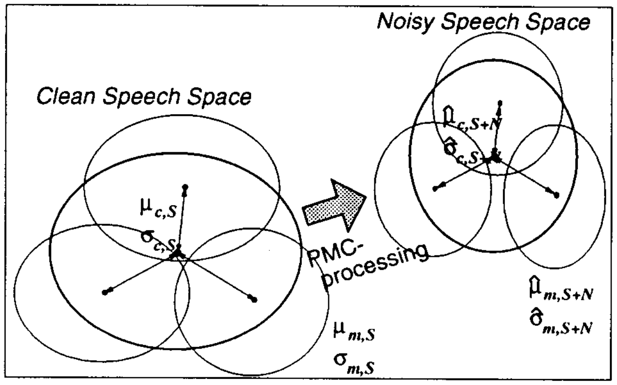

Figure 3: Image of Fast PMC Processing

3.2 Algorithm

1) Group close distributions (μ_m, σ_m^2) and create a composite distribution (μ_c, σ_c^2) par group:

$$\mu_c = \sum_{m \in G} w_m \mu_m$$

$$\sigma_c^2 = \sum_{m \in G} w_m \sigma_m^2 + \sum_{m \in G} w_m (\mu_m - \mu_c)^2$$

where w indicates weights and G indicates groups.

In this paper, the group is the state of HMM, while distribution clustering is an alternative way of grouping. Figure 4 shows the image of "the state is the group".

2) Calculate the difference vectors between distributions (μ_m, σ_m^2) in the group and the composite distribution (μ_c, σ_c^2) of the group.

3) Perform PMC-processing on the composite distribution[1].

4) Calculate the noise corrupted position and area of each distribution by the difference between each distribution and the composite distribution before PMC, and the area ratio of the composite distribution before PMC and after, using the next equations:

$$\hat{\mu}_{m,S+N} = \hat{\mu}_{c,S+N} + (\mu_{m,S} - \mu_{c,S})\ (\hat{\sigma}_{c,S+N}/\sigma_{c,S})$$

$$\hat{\sigma}_{m,S+N} = \sigma_{m,S}\ (\hat{\sigma}_{c,S+N}/\sigma_{c,S})$$

while ${}_{S+N}$ indicates noisy speech and ${}_{S}$ indicates clean speech and μ, σ before adaptation and $\hat{\mu}, \hat{\sigma}$ after adaptation.

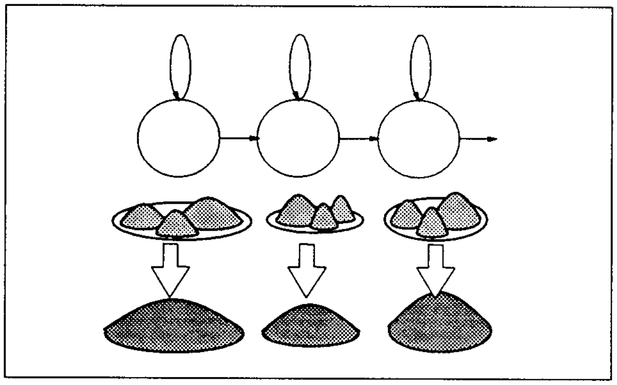

Figure 4: Grouping by State

4 Experiment

This section reports the comparison evaluation results, both adaptation processing time and in recognition accuracy, of the proposed fast PMC method with the basic PMC method using speaker independent telephone speech.

4.1 Condition

A speaker independent telephone speech recognition experiment is performed for evaluation. The tasks are 520 word recognition and 1,000 vocabulary size continuous speech recognition. Three types of HMMs are used in the experiment. 1) right context phone HMM of 3 state 6 mixtures and 2) right context phone HMM of 3 state 12 mixture, and 3) 785 shared-state triphone HMM of 3 state 6 mixture. These models are trained from the clean speech database and then they are adapted to the telephone environment by the MCMS-PMC[4]. Noise data of 1 second is used for PMC adaptation and 6 second telephone speech data is used for MCMS adaptation. Details of the MCMS-PMC environment adaptation is described in the [4]. Conditions are briefly shown in table 2.

Table 2: Experiment Condition

Acoustic Analysis	sampling rate: 8kHz, shift frame: 10ms, hamming window: 25.6ms, pre-emphasis: 0.97, LPC-Mel-Cep(12 dimension) ΔCep(12 dimension) and Δpower
Training Data	ASJ+ATR+CANON speech data 200 speakers, 72,000 utterances
Evaluation Data1 (20 speakers)	ATR speech database 520 words (104 words/speaker) telephone-line word utterance
Evaluation Data2 (10 speakers)	CANON speech database 1,004 words, perplexity 30.2 telephone-line continuous speech
model 1.	right context models of 3state6mix. 262 models
model 2.	right context models of 3state12mix. 262 models
model 3.	785 shared-state triphone of 3state6mix. 3334 models

4.2 Results

Table 1 shows the results with the time for PMC-processing (HP9000/J210) and the recognition accuracy. The "none" in the table indicates the results with no PMC-processing, the "basic" indicates the normal baseline PMC-processing and the "fast" indicates the proposed fast PMC-processing. The MCMS environment adaptation is carried out on all models for telephone band-width compensation. The "distributions" indicates the number of state × mixtures. This number is equal to the number of PMC-

Table 1: Recognition Performance on Telephone Speech

models		PMC	PMC time	Data1:word(%)		Data2:sentence(%)		
context	distributions		(sec)	top1	top5	word acc.	top1	top5
right	786 × 6	none	—	81.5	95.6	91.4	68.0	81.2
right	786 × 6	basic	11.0	89.3	98.0	94.9	75.8	90.2
right	786 × 6	fast	4.1	88.2	97.7	94.7	76.0	89.0
right	786 ×12	none	—	83.7	96.9	93.8	74.0	84.4
right	786 ×12	basic	20.6	89.7	98.4	95.9	79.4	93.6
right	786 ×12	fast	6.0	88.4	97.8	94.7	78.0	91.0
triphone	785 × 6	none	—	83.7	96.6	88.9	63.0	76.6
triphone	785 × 6	basic	12.0	90.6	98.2	94.4	77.6	88.2
triphone	785 × 6	fast	4.3	89.4	97.5	94.4	77.4	87.8

triphone : shared state model

processing times in the basic PMC. In the fast PMC, the number of PMC-processing times is equal to the number of states.

From the table, we can say the following things:

- In case of continuous speech recognition task, almost no degradation is seen on both sentence accuracy and word accuracy when the right model of 6 mixtures and the triphone model of 6 mixtures are utilized. However, in case of right model of 12 mixtures of the continuous speech recognition task or in every model of 520 word recognition task, slight degradation around 1% is seen. Although slight degradation of recognition accuracy is seen, the accuracy of the proposed method is almost comparable to that of the basic PMC and further better compared with that of the no PMC-adaptation "none".
- In case of right model of 6 mixtures, the processing time is reduced to 37.3%(4.1/11.0), in case of right model of 12 mixtures 29.1%(6.0/20.6), and in case of shared-state triphone 35.8%(4.3/12.0). The PMC-processing time of the propose fast PMC method is around 1/3 of the basic PMC method.
- The computation of the proposed method do not proportionally depend on the number of distribution. In the basic PMC, when the number of distributions is doubled the PMC-processing time is also double, however in the proposed fast PMC method the PMC-processing time is only 1.5 times even if the the number of distributions is doubled. (See the right models.)

From these results, we can conclude that the proposed fast PMC method is effective on the PMC computation reduction with almost no degradation of recognition performance.

5 Conclusion

A fast PMC noise adaptation method is proposed for continuous HMM base speech recognizer. The proposed method is evaluated on the telephone speech by integrating the basic or the proposed PMC into the MCMS-PMC environment adaptation [4]. The results showed that the proposed method saved around 65% (70.9% - 62.7%) of PMC computation amount compared to the basic PMC with almost no degradation of recognition performance.

Acknowledgment

The authors wish to thank Dr. Hideyuki TAMURA, Head of the Media Technology Laboratory at Canon Inc., Mr. Minoru FUJITA, the manager of the Intelligent Media Division, Mr. Tsuyoshi YAGISAWA, the manager of our department, for giving us the opportunity of this study.

References

[1] M.J.Gales, et al.: An improved approach to the hidden Markov model decomposition of speech and noise, ICASSP92, pp.233-236, 1992.

[2] K.Takagi, et al.: Rapid environment adaptation for robust speech recognition, ICASSP95, pp.149-152.

[3] M.J.Gales, et al.: A fast and flexible implementation of Parallel Model Combination, ICASSP95, pp.I-133-136,1995-5.

[4] H.Yamamoto, et al.: Fast speech recognition algorithm under noisy environment using modified CMS-PMC and improved IDMM+SQ, ICASSP97, 1997.(to be appeared)

[5] Y.Minani et al.: Adaptation method base on HMM composition and EM algorithm, ICASSP95, pp.327-330, 1996-5.