EIGENSPACE-BASED LINEAR TRANSFORMATION APPROACH FOR RAPID SPEAKER ADAPTATION

Kuan-Ting Chen and Hsin-Min Wang

Institute of Information Science, Academia Sinica
Taipei, Taiwan, Republic of China
e-mail: {kenneth, whm}@iis.sinica.edu.tw

ABSTRACT

This paper presents our recent effort on the development of the eigenspace-based linear transformation approach for rapid speaker adaptation. The proposed approach toward prior density selection for the MAPLR framework was developed by introducing *a priori* knowledge analysis on the training speakers via probabilistic principal component analysis (PPCA), so as to construct an eigenspace for speaker-specific full regression matrices as well as to derive a set of bases called eigen-transformations. The prior densities of MAPLR transformations for each outside speaker are then chosen in the space spanned by the first few eigen-transformations. By incorporating the PPCA model of transformation parameters into the MAPLR scheme, the number of free parameters can be significantly reduced, while the underlying structure of the acoustic space as well as the precise modeling of the inter-dimensional correlation among the model parameters can be well preserved. Rapid supervised adaptation experiments showed that the proposed approach not only is superior to the conventional MLLR approach using either diagonal or block-diagonal regression matrices, but also outperformed by a great amount the full-matrix MLLR with either a global transformation or multiple transformations corresponding to different phonetic classes.

1. INTRODUCTION

Among the various model-based speaker adaptation techniques, the transformation-based approaches such as maximum likelihood linear regression (MLLR) [1] have been widely used for rapid adaptation and unsupervised adaptation. In MLLR for HMM mean adaptation, the speaker independent (SI) mean parameters are adapted according to one or more shared linear transformations. The transformation parameter tying mechanism based on the design of regression class tree can automatically adjust the level of parameter sharing according to the amount and content of data and, thus, can effectively improve the robustness of parameter estimation against the sparse data problem.

There are, however, several drawbacks of the conventional MLLR approach. In MLLR for mean adaptation, it is known that using full regression matrices can model the inter-dimensional correlation among the mean parameters more precisely and, thus, can provide superior description of speaker characteristics over the use of either block-diagonal or diagonal regression matrices. However, the large number of parameters makes robust estimation of full regression matrices very difficult, especially when the amount of adaptation data available is strictly limited. This problem can be alleviated by specifying a prior distribution for each of the regression matrices, and estimating the transformations in maximum *a posteriori* (MAP) sense. This leads to the maximum *a posteriori* linear regression (MAPLR) formulation [2]. Provided that good priors are chosen, the estimation of the regression matrices could be more robust.

Despite improved performance over MLLR, the MAPLR approach still suffers from some limitations. First, deriving the prior densities for the regression matrices in any node of the regression class tree is difficult and usually requires ad hoc techniques. Second, as in MLLR, all regression matrices are estimated independently of one another. Assuming a unified structure of the acoustic space for each outside speaker, a systematic approach toward prior density estimation for each and every transformation considering the correlation among different regression classes is, thus, highly desired. The structural MAPLR (SMAPLR) approach [4] addresses these issues by structuring the *a priori* information about the transformations in the regression class tree and suggests a method to derive prior densities for any of the transformations. However, the prior density selection for root node remains difficult, and the prior evolution scheme doesn't guarantee good priors for the lower layer when the nodes in upper layer are not estimated properly.

On the other hand, it is believed that the *a priori* knowledge about the inter-speaker variation can be explored by analyzing the training corpus. The eigenvoice approach [5] introduced for fast speaker adaptation is one of the examples that realize such concept. The eigenvoice technique finds the new speaker model as the linear combination of a set of canonical speaker models called eigenvoices. These eigenvoices, which characterize the *a priori* information of the training speakers, are constructed by performing principal component analysis (PCA) [6] on a set of speaker dependent (SD) model parameters. Recently, the eigenvoice approach was further extended via the PPCA model [7] and incorporated into the Bayesian adaptation framework [8].

In the transformation-based adaptation methods, each set of transformations for a specific speaker represents a mapping from the SI models to the SD models for that speaker and, thus, can be considered as a quantitative description of the speaker characteristics. It is clear that the *a priori* information about inter-speaker variation can also be obtained by analyzing the transformations for the training speakers, and a "speaker space" can thereby be characterized. In our previous work [9], the eigenspace of the speaker-specific full regression matrices, which was constructed by performing PCA on the set of training speaker-specific regression matrices, was utilized to improve the conventional full-matrix MLLR when the amount of adaptation data was strictly limited. Such idea was further extended by incorporating the PPCA model into the MAPLR formulation, leading to a unified framework of eigenspace-based MAPLR estimation as well as a procedure of prior density derivation [10]. Preliminary experiments showed that in the case of using a global transformation for HMM mean adaptation, the proposed eigenspace-based approach provided a relatively simple and reasonable solution to deriving the prior density of the MAPLR transformation, and led to superior performance on both supervised and unsupervised adaptation over conventional MLLR approaches.

In this paper, we will further examine the previous proposed eigenspace-based MAPLR approach in the case of multiple regression classes and discuss its property considering the correlation among the sets of transformation parameters associated with different regression classes. The rest of this paper is organized as follows. The eigenspace-based transformations and the PPCA model are introduced in Sections 2 and 3 respectively. The eigenspace-based MAPLR framework is presented in Section 4. Finally some experimental results for rapid supervised adaptation tested on a continuous Mandarin Chinese telephone speech database are discussed in Section 5, and concluding remarks are made in Section 6.

2. EIGENSPACE-BASED TRANSFORMATION

2.1. MLLR

MLLR finds the optimal affine transformations with respect to the mixture components in the SI model by maximizing the likelihood of adaptation data. SI Gaussian mean parameters are clustered into C regression classes, and each regression class c is associated with an $n \times (n+1)$ regression matrix $\mathbf{W}_c$, where n is the dimensionality of the feature vector. Let the mean vector $\boldsymbol{\mu}_m = [\mu_m(1),\ldots, \mu_m(n)]^{\mathrm{T}}$ of mixture component m be one of the M_c mean vectors in the regression class c, then the adapted mean vector $\hat{\boldsymbol{\mu}}_m$ can be derived as

$$\hat{\boldsymbol{\mu}}_m = \mathbf{W}_c \tilde{\boldsymbol{\mu}}_m = \mathbf{A}_c \boldsymbol{\mu}_m + \mathbf{b}_c,\ m = 1,\ldots, M_c;\ c = 1,\ldots, C\ , \quad (1)$$

where $\tilde{\boldsymbol{\mu}}_m = [1, \mu_m(1), \cdots, \mu_m(n)]^{\mathrm{T}}$ is the $(n+1)$-dimensional augmented mean vector. $\mathbf{A}_c$ and $\mathbf{b}_c$ are an $n \times n$ matrix and an n-dimensional vector, respectively, while $\mathbf{W}_c = [\mathbf{b}_c\ \mathbf{A}_c]$.

2.2. Eigenspace-based MLLR

Let $\{\mathbf{m}_{r,c}\}_{r\,=\,1,\ldots,R}$ be a set of supervectors, each of which consists of the c-th full regression matrix parameters for one of the R training speakers. To be more specific, $\mathbf{m}_{r,c} = [(\mathbf{W}_{r,c}^{(1)})^{\mathrm{T}},\ldots,(\mathbf{W}_{r,c}^{(l)})^{\mathrm{T}},\ldots,(\mathbf{W}_{r,c}^{(n+1)})^{\mathrm{T}}]^{\mathrm{T}}$, where $\mathbf{W}_{r,c}^{(l)}$ represents the l-th column of the c-th regression matrix for the r-th speaker. The eigenspace of the regression matrices is found as the linear subspace spanned by the K-dominant eigenvectors $\{\mathbf{e}_{1,c},\ldots,\mathbf{e}_{K,c}\}$ of the correlation matrix $\mathbf{P}_c$, where

$$\mathbf{P}_c = [\mathbf{m}_{1,c} - \overline{\mathbf{m}}_c,\ldots,\mathbf{m}_{R,c} - \overline{\mathbf{m}}_c][\mathbf{m}_{1,c} - \overline{\mathbf{m}}_c,\ldots,\mathbf{m}_{R,c} - \overline{\mathbf{m}}_c]^{\mathrm{T}} \quad (2)$$

and $\overline{\mathbf{m}}_c$ represents the sample mean. The bases $\{\mathbf{e}_{1,c},\ldots,\mathbf{e}_{K,c}\}$ are called *eigen-transformations*, since they represent the principal components of the transformation space. Let $\mathbf{M}_c = [\mathbf{e}_{1,c},\ldots,\mathbf{e}_{K,c}]$, then the supervector of the c-th full regression matrix for a new speaker, denoted as $\hat{\mathbf{m}}_c$, can be obtained by

$$\hat{\mathbf{m}}_c = \mathbf{M}_c \mathbf{x} + \overline{\mathbf{m}}_c\ , \quad (3)$$

where $\mathbf{x} = [\mathrm{x}_1,\ldots,\mathrm{x}_K]$ represents the coordinate vector for the new speaker and can be estimated by maximizing the likelihood of the adaptation data O as follows:

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} [\log p(O \mid \mathbf{M}_c \mathbf{x} + \overline{\mathbf{m}}_c)]\ . \quad (4)$$

The concept of employing the eigenspace of the full regression matrices for adaptation significantly reduced the number of free parameters and, thus, could improve the robustness of transformation estimation when only very limited adaptation data was available. However, we also found that adaptation performance of directly using the eigenspace-based regression matrices tended to saturate as the amount of adaptation data increased. Nevertheless, the eigenspace-based transformations provide a convenient way to determine the priors in the MAPLR framework, as will be discussed later on.

3. PROBABILISTIC PRINCIPAL COMPONENT ANALYSIS (PPCA)

In this section we briefly introduce the formulation of the PPCA model [7]. Let $\mathbf{y} = [\mathrm{y}_1, \ldots,\mathrm{y}_D]^{\mathrm{T}}$ be an observation vector of dimension D. In the PPCA model, $\mathbf{y}$ is assumed to be related to the latent variable $\mathbf{x} = [\mathrm{x}_1,\ldots,\mathrm{x}_K]^{\mathrm{T}}$ of dimension K by

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \overline{\mathbf{y}} + \varepsilon\ , \quad (5)$$

where $\overline{\mathbf{y}}$ is the mean vector of $\mathbf{y}$, $\mathbf{M}$ is a $D \times K$ matrix ($K<<D$) representing the principal subspace of the observation data, and ε is a normally distributed noise independent of $\mathbf{x}$. Conventionally, $\mathbf{x}$ is defined to be modeled by a K-dimensional multivariate Gaussian distribution $N(\mathbf{0},\mathbf{I}_K)$, and the noise ε is also modeled by a multivariate Gaussian $N(\mathbf{0},\sigma^2\mathbf{I}_D)$, where $\mathbf{I}_K$ and $\mathbf{I}_D$ are the $K \times K$ and $D \times D$ identity matrices respectively. Based on the above assumptions, the conditional density of $\mathbf{y}$ given a specific value of $\mathbf{x}$ can be derived as

$$f(\mathbf{y} \mid \mathbf{x}) = (2\pi\sigma^2)^{-D/2} \exp\left\{ -\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{M}\mathbf{x} - \overline{\mathbf{y}}\|^2 \right\}. \quad (6)$$

Given an observation sequence $\mathbf{Y} = \{\mathbf{y}_1,\ldots,\mathbf{y}_T\}$, the PPCA model estimates the latent variable sequence $\mathbf{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_T\}$ and finds the optimal parameter set $\hat{\lambda} = \{\hat{\mathbf{M}}, \hat{\overline{y}}, \hat{\sigma}^2\}$ according to the maximum likelihood (ML) criterion. However, because the variable sequence $\mathbf{X}$ is considered to be hidden, the ML estimate of the parameter set cannot be derived as a closed form solution. Therefore, the expectation maximization (EM) algorithm [11] is applied to iteratively update the model parameters. Let $\lambda^{(n)}$ be the parameter values obtained in the n-th iteration, then the new estimates $\lambda^{(n+1)}$ are obtained according to

$$\lambda^{(n+1)} = \arg\max_{\lambda} E\left[\log p(\mathbf{Y},\mathbf{X} \mid \lambda) \mid \mathbf{Y}, \lambda^{(n)}\right]. \quad (7)$$

By applying the EM algorithm, it is shown that a set of closed form re-estimation formulae for the parameter values $\lambda^{(n+1)} = \{\mathbf{M}^{(n+1)}, \overline{y}^{(n+1)}, \sigma^{2,(n+1)}\}$ in the $(n+1)$-*th* iteration can be derived [7]. It can also be shown that for the global maximum of the likelihood, the ML estimate $\mathbf{M}_{\mathrm{ML}}$ contains the principal eigenvectors of the correlation matrix of the observation data [11], which implies that the PPCA model provides a unified framework for the training of the eigenspace parameters and can be utilized to derive the eigen-transformations in the proposed framework.

4. EIGENSPACE-BASED MAPLR FRAMEWORK

In this section we discuss the formulation of the proposed eigenspace-based MAPLR framework. Let $O = \{\mathbf{o}_1,\ldots,\mathbf{o}_T\}$ be a sequence of n-dimensional feature vectors generated by a CDHMM system with the parameter set Λ. The observation probability density function $p(\mathbf{o}_t \mid j)$ for state j is assumed to be a mixture of Gaussian distributions:

$$p(\mathbf{o}_t \mid j) = \sum_{k=1}^{M} w_{j,k} N(\mathbf{o}_t \mid \boldsymbol{\mu}_{j,k}, \Sigma_{j,k}) , \qquad (8)$$

where M is the mixture number in state j, $w_{j,k}$ is the weight of the k-th mixture, and $N(\bullet)$ represents the conventional n-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ.

Let the Gaussian mean vectors be divided into C disjoint clusters, $\{\lambda_c\}_{c=1,\ldots,C}$. Each cluster c is associated with a full regression matrix $\mathbf{W}_c$, and the parameter values in λ_c are updated according to (1). If $\mathbf{W}_c$, whose elements are considered to be random, is to be estimated from the likelihood function $f(O \mid \mathbf{W}, \lambda_c)$ and its prior density $g(\mathbf{W}_c \mid \Theta_c)$, then the MAP estimate is defined as

$$\hat{\mathbf{W}}_c = \arg\max_{\mathbf{W}} f(O \mid \mathbf{W}, \lambda_c) g(\mathbf{W} \mid \Theta_c) \ . \qquad (9)$$

Let each supervector $\mathbf{m}_c$ that augments the columns of $\mathbf{W}_c$ be generated by a PPCA model given by (5), which has a latent variable $\mathbf{x}_c$ with a prior parameter set $\theta_c = \{\mathbf{M}_c, \overline{\mathbf{m}}_c, \sigma^2\}$. Such assumption on the choice of prior density for $\mathbf{W}_c$, in fact, makes a special case of the elliptically symmetric matrix variate priors [3]. Furthermore, the prior parameter set θ_c can be derived in a data-driven manner by taking the transformation parameters for the training speakers as the sample data of the PPCA model. We can then apply the EM algorithm to iteratively obtain the optimal estimate of $\phi = \{\mathbf{m}_c\}$. Assume that $\phi^{(n)} = \{\mathbf{m}_c^{(n)}\}$ be the current fit, then the object auxiliary function to be maximized is given as

$$\begin{aligned} R(\phi, \phi^{(n)}) &= E\left[\log p(O, S, L \mid \Lambda, \phi) \mid O, \phi^{(n)}\right] \\ &\quad + E\left[\log p(\phi, \mathbf{x} \mid \theta) \mid \phi^{(n)}\right] \end{aligned} \qquad (10)$$

where S and L represent the optimal posterior state sequence and mixture sequence, respectively. Based upon (6), it can be found that the relevant object function is given as

$$\begin{aligned} & R(\phi, \phi^{(n)}) = \\ & \sum_{t=1}^{T}\sum_{c=1}^{C}\sum_{k=1}^{M_c} \gamma_k(t) \left[-\frac{1}{2} (\mathbf{o}_t - \mathbf{W}_c \tilde{\boldsymbol{\mu}}_k)^{\mathrm{T}} \Sigma_k^{-1} (\mathbf{o}_t - \mathbf{W}_c \tilde{\boldsymbol{\mu}}_k) \right] \quad (11) \\ & + \sum_{c=1}^{C} E\left[-\frac{1}{2\sigma^2} \left\| \mathbf{m}_c - \mathbf{M}_c \mathbf{x}_c - \overline{\mathbf{m}}_c \right\|^2 \middle| \mathbf{m}_c^{(n)} \right], \end{aligned}$$

where $\gamma_k(t)$ represents the mixture occupation probability of the k-th Gaussian in the regression class c at time t given the observation sequence O, while M_c is the number of mixture components in class c. Starting from (11), it can be shown that by maximizing the relevant object function in a column-by-column manner, we have

$$\begin{aligned} & \sum_{t=1}^{T}\sum_{k=1}^{M_c} \gamma_k(t) \Sigma_k^{-1} \mathbf{W}_c \tilde{\boldsymbol{\mu}}_k \tilde{\boldsymbol{\mu}}_k^{\mathrm{T}} + \frac{1}{\sigma^2} \mathbf{W}_c \\ & = \sum_{t=1}^{T}\sum_{k=1}^{M_c} \gamma_k(t) \Sigma_k^{-1} \mathbf{o}_t \tilde{\boldsymbol{\mu}}_k^{\mathrm{T}} + \frac{1}{\sigma^2} \mathbf{W}_{c,\,\mathrm{PPCA}} \ , \end{aligned} \qquad (12)$$

where the elements in $\mathbf{W}_{c,\,\mathrm{PPCA}}$ can be derived according to the PPCA model parameters $\{\mathbf{M}_c, \overline{\mathbf{m}}_c, \sigma^2\}$, which can be obtained off-line by applying the re-estimation formulae for PPCA model on the set of training speaker-specific transformation parameters [10].

It is found that (12) can be further rewritten as

$$\sum_{k=1}^{M_c} \mathbf{V}_k \mathbf{W}_c \mathbf{D}_k + \frac{1}{\sigma^2} \mathbf{W}_c = \mathbf{Z} + \frac{1}{\sigma^2} \mathbf{W}_{c,\,\mathrm{PPCA}} , \qquad (13)$$

where $\mathbf{V}_k = \sum_{t=1}^{T} \gamma_k(t) \Sigma_k^{-1}$, and $\mathbf{D}_k = \tilde{\boldsymbol{\mu}}_k \tilde{\boldsymbol{\mu}}_k^{\mathrm{T}}$. Assume that the diagonal covariance prevailing conditions is used and let $\mathbf{G}_i = \sum_{k=1}^{M_c} \mathbf{V}_k(i,i) \mathbf{D}_k$, we have the re-estimation formula for $\mathbf{w}_{c,(i)}$, which is the i-th row of $\mathbf{W}_c$, as

$$\mathbf{w}_{c,(i)} = \left(\mathbf{G}_i + \frac{1}{\sigma^2} \mathbf{I}_{n+1} \right)^{-1} \left(\mathbf{z}_{(i)} + \frac{1}{\sigma^2} \mathbf{w}_{c,\,\mathrm{PPCA},(i)} \right) \qquad (14)$$

where $\mathbf{z}_{(i)}$ and $\mathbf{w}_{c,\mathrm{PPCA},(i)}$ are the i-th row of $\mathbf{Z}$ and $\mathbf{W}_{c,\mathrm{PPCA}}$ respectively. As compared to the standard MLLR re-estimation formula [1], (14) shows that, though not explicitly, the proposed MAPLR approach provides a unified framework that incorporates the eigenspace-based transformations $\mathbf{W}_{c,\mathrm{PPCA}}$ into the conventional MLLR re-estimation procedure.

The advantages of the proposed algorithm can be explained as follows. First, the similarity of (14) as compared to the standard MAPLR re-estimation algorithm [3] indicates good asymptotic property as the amount of data increases. Second, the priors $\mathbf{W}_{c,\mathrm{PPCA}}$ can be simply determined according to the location of the new speaker within a lower dimensional eigenspace of transformations and, thus, can be effectively constrained with a much fewer number of parameters to be estimated. This implies not only the capability of rapid adaptation of the proposed approach, but also that the precise modeling of inter-dimensional correlation among the mean parameters may be well preserved. Third, in the case of multiple regression classes, by concatenating the parameters of all classes for each speaker to form a single supervector, the correlation among the sets of parameters associated with different regression classes can be further taken into account automatically in PPCA model training. This implies stronger constraints for the behavior of multiple regression classes as compared to estimating individual regression matrix independently.

5. EXPERIMENTAL RESULTS

The proposed approach was evaluated on a continuous Mandarin Chinese telephone speech database provided by Telecommunication Laboratories, Taiwan, Republic of China. The database consists of 59 female and 60 male speakers, each produced 120 sentences such that a total of 14,280 sentences (5.84 hrs) are included. The speech was sampled at 8 kHz, then parameterized into 12 MFCCs along with log-energy, and the first and second order time derivatives of these parameters, yielding a 39-dimensional feature vector. Cepstral mean subtraction (CMS) was performed on a per-speaker basis to remove the channel effect of the features. Baseline gender independent (GI) SI model was trained with the training set which contains 54 female and 55 male speakers. Considering the monosyllabic structure of Chinese language in which each syllable can be decomposed into an INITIAL/FINAL format, 112 INITIAL models and 38 FINAL models were used as the acoustic units for training. The total number of Gaussian mixture components in the SI model is approximately 2370.

For each testing speaker, the first 60 sentences were taken as the adaptation data while the rest were for recognition. The recognizer performed only free syllable decoding without any grammar constraints. The SI syllable recognition accuracy is 55.81%, averaged over 5 female and 5 male testing speakers.

Three types of conventional MLLR approaches using full, block-diagonal and diagonal regression matrices were conducted as baseline experiments. The full matrix and block-diagonal matrix MLLR were based on either a single global transformation or two transformations corresponding to INITIAL models and FINAL models, respectively. Diagonal matrix MLLR utilized a regression class tree for dynamic regression class generating with the maximum number of classes set to be 64. As for the proposed approach, instead of estimating the PPCA model parameters, the maximum likelihood eigen-decomposition (MLED) based transformations obtained by (3) and (4) were directly employed as the approximations of the prior parameters to simplify the off-line processing. This approximation allows us to conveniently adopt the framework proposed in our previous work [9]. In the training phase, for each training speaker all the 120 utterances were used to estimate a set of speaker-specific full regression matrices (each composed of $39\times 40 = 1560$ elements) and, then, PCA was performed to extract 109 bases. During adaptation, various number of eigen-transformations were selected to construct the principal subspace, and the PPCA model parameter σ was determined empirically.

We performed a series of batch supervised adaptation experiments, with various numbers of sentences extracted from the adaptation data. The results are summarized in Table 1, with "Global" and "I-F" representing using a global transformation and two transformations for INITIAL/FINAL, respectively. As for the proposed approach, the case of using two matrices for eigenspace training and adaptation (I-F) can be further classified into two methods: the "Layer-based" method corresponds to eigenspace treating the parameters of both transformations in the same layer of regression class tree as a single vector ($D = 1560\times 2 = 3120$), while the "Node-based" method corresponds to constructing two eigenspaces with respect to individual transformation ($D = 1560$). Each method was evaluated with different number (K) of eigen-transformations.

As can be observed from Table 1, Conventional MLLR approach using either full matrix or block-diagonal matrix performed poorly when the amount of data was strictly limited. The situation became even worse when two regression matrices were used since the amount of data corresponding to individual class further decreased. On the other hand, with either a single transformation or two phonetic-class-based transformations used for adaptation, the proposed approach not only improved the recognition accuracy significantly but also outperformed the diagonal-matrix MLLR approach. Focusing on the proposed approach, we found that using multiple regression matrices performed slightly worse than using a single global matrix, which might be due to the fact that the number of transformation parameters in the former case is doubled. To be more specific, it can be found that the node-based approach performed worse than the layer-based approach in the multiple transformations case, which implied that constructing eigenspace with respect to all transformation parameters as a single vector is more reasonable, since the inter-correlation of different phonetic classes can be further characterized. This also suggests the advantage of the proposed approach over the MAPLR algorithm in terms of characterizing the inter-speaker variability by an overall processing of the transformations.

Average Data Length (sec.)			4.11	6.85	9.59	13.70
Diagonal MLLR			56.18	56.41	56.62	57.12
Block-diag. MLLR	Global		45.50	53.99	56.80	58.28
	I-F		21.99	43.31	51.84	56.82
Full MLLR	Global		5.85	21.81	38.15	48.02
	I-F		0.31	1.35	5.27	18.27
Eigenspace-based MAPLR	Global	K=10	57.51	57.93	58.00	58.60
		K=30	58.02	58.34	58.76	59.17
		K=50	58.16	57.95	58.41	59.41
		K=70	57.19	57.47	58.41	59.22
	I-F (Layer-based)	K=10	57.54	57.58	57.88	58.32
		K=30	58.00	58.28	58.51	58.60
		K=50	57.56	57.91	58.60	59.22
		K=70	57.10	57.84	58.64	59.13
	I-F (Node-based)	K=10	57.05	57.12	57.65	58.18
		K=30	57.24	57.63	58.32	58.58
		K=50	55.95	57.35	58.30	58.92
		K=70	54.91	57.08	58.09	58.71

Table 1: Supervised batch adaptation performance in syllable accuracy (%). The syllable accuracy of SI system is 55.81%.

6. CONCLUSION

In this paper, the eigenspace-based MAPLR approach was discussed. The formulation of the proposed adaptation framework was first reviewed, and the experiments regarding using multiple regression classes for adaptation were conducted. Significant improvements in batch supervised adaptation showed the effectiveness of the proposed approach for rapid model adaptation.

REFERENCES

[1] C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models," *Computer Speech and Language*, vol. 9, pp.171-185, 1995.

[2] C. Chesta, O. Siohan, and C.-H. Lee, "Maximum a Posteriori linear regression for hidden Markov model adaptation," *Proc. EuroSpeech'*99, pp.211-214, 1999.

[3] W. Chou, "Maximum a Posteriori linear regression with elliptically symmetric matrix variate priors," *Proc. EuroSpeech*'99, pp.1-4, 1999.

[4] O. Siohan, T.-A. Myrvoll, and C.-H. Lee, "Structural maximum a Posteriori linear regression for fast HMM adaptation," *Proc. ISCA ITRW ASR*2000, 2000.

[5] R. Kuhn, *et. al.*, "Eigenvoices for speaker adaptation," *Proc. ICSLP'*98, pp.1771-1774, 1998.

[6] I. T. Jolliffe, *Principal Component Analysis*. Springer- Verlag, 1986.

[7] M. Tipping and C. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol.11, pp.435-474, 1999.

[8] D. K. Kim and N.S. Kim, "Bayesian speaker adaptation based on probabilistic principal component analysis," *Proc. ICSLP'*2000, 2000.

[9] K. T. Chen, W. W. Liau, H. M. Wang and L. S. Lee, "Fast speaker adaptation using eigenspace-based maximum likelihood linear regression", *Proc. ICSLP*'2000, 2000.

[10] K. T. Chen and H. M. Wang, "Eigenspace-based maximum a Posteriori linear regression for rapid speaker adaptation", *Proc. ICASSP'*2001, 2001.

[11] A. P. Dempster, N. M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, vol. 39, pp.1-38, 1977.