EIGENSPACE-BASED MAXIMUM *A POSTERIORI* LINEAR REGRESSION FOR RAPID SPEAKER ADAPTATION

Kuan-ting Chen and Hsin-min Wang

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

ABSTRACT

In this paper, we present an eigenspace-based approach toward prior density selection for the MAPLR framework. The proposed eigenspace-based MAPLR approach 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-matrices. The priors of MAPLR transformations for each outside speaker are then chosen in the space spanned by the first *K* eigen-matrices. By incorporating the PPCA model into the MAPLR scheme, the number of free parameters in choosing the priors can be effectively 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. Both supervised and unsupervised adaptation experiments showed that the proposed approach significantly outperformed the conventional MLLR approach using either diagonal or full regression matrices.

1. INTRODUCTION

Various speaker adaptation techniques have been extensively studied in recent years to tackle the problem of speaker mismatch between the training set and the testing set of the speech recognition systems. According to [1], the popular model-based adaptation techniques can be classified into three families: the maximum *a posteriori* (MAP) adaptation family, the transformation-based adaptation family including maximum likelihood linear regression (MLLR) [2], and a family related to speaker clustering methods such as the eigenvoice approach [3]. In this paper, we will focus on the adaptation of mean parameters of the Gaussian mixture components in continuous density HMMs.

Among these techniques, MLLR approach has been widely used for rapid adaptation and unsupervised adaptation. In MLLR, the speaker independent (SI) mean parameters are adjusted according to one or more shared linear transformations. The transformation parameter tying mechanism based on the design of regression class tree can adequately adjust the level of regression matrix 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 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 diagonal regression matrices [2]. 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 [4]. Provided that good priors are chosen, the estimation of the regression matrices could be more robust.

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 [4] 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) [10] on a set of speaker dependent (SD) model parameters. Recently, the eigenvoice approach was further extended via the PPCA model [6] and incorporated into the Bayesian adaptation framework [7].

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 can also be obtained by analyzing the transformations for the training speakers. In our previous work [8], the eigenspace of the speaker-specific full regression matrices was utilized to improve the conventional full-matrix MLLR when the amount of adaptation data was strictly limited. To alleviate the problem of performance saturation as the amount of adaptation data increased, the eigenspace-based transformations were used as *a priori* for a smoothing procedure on the conventional MLLR transformations instead of used directly for adaptation.

In this paper, we will further extend such idea to the formulation of eigenspace-based MAPLR estimation, and propose a framework for priors choosing by employing the PPCA model. 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 MAPLR framework is presented in Section 4. Finally some experimental results for both supervised and unsupervised adaptation tested on a continuous Mandarin Chinese telephone speech database are discussed in Section 5, and concluding remarks are made in Section 6.

Thanks to Institute for Information Industry and National Science Council of the Republic of China for funding.

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 $m_m = [m_m(1),\ldots, m_m(n)]^{\mathrm{T}}$ of mixture component m be one of the M_c mean vectors in the regression class c, then the adapted mean vector $\hat{m}_m$ can be derived as

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

where $\tilde{m}_m = [1, m_m(1),\cdots,m_m(n)]^{\mathrm{T}}$ 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-matrices*, 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 $\mathbf{x} = [x_1,\ldots,x_K]$ 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 [6]. Let $\mathbf{y} = [y_1, \ldots, y_D]^{\mathrm{T}}$ be an observation vector of dimension D. In the PPCA model, $\mathbf{y}$ is assumed to be related to the latent variable $\mathbf{x} = [x_1,\ldots,x_K]^{\mathrm{T}}$ 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 e 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 e is also modeled by a multivariate Gaussian $N(\mathbf{0}, s^2\mathbf{I}_D)$, 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

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

Given an observation sequence $\mathbf{Y} = \{\mathbf{y}_1,\ldots,\mathbf{y}_T\}$, the PPCA model estimates the hidden variable sequence $\mathbf{X} = \{\mathbf{x}_1,\ldots,\mathbf{x}_T\}$ and finds the optimal parameter set $\hat{\lambda} = \{\hat{\mathbf{M}}, \hat{\overline{y}}, \hat{s}^{\,2}\}$ 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 [9] 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

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

The re-estimation formulae for the parameter values can be derived as follows:

$$\overline{\mathbf{y}}^{(n+1)} = \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{y}_t - \mathbf{M}^{(n)} E\left[\mathbf{x}_t \mid \mathbf{y}_t, \lambda^{(n)} \right] \right) \quad (8)$$

$$\mathbf{M}^{(n+1)} = \left\{ \sum_{t=1}^{T} \left(\mathbf{y}_t - \overline{\mathbf{y}}^{(n+1)} \right) E\left[\mathbf{x}_t^{\mathrm{T}} \mid \mathbf{y}_t, \lambda^{(n)} \right] \right\} \cdot \left\{ \sum_{t=1}^{T} E\left[\mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \mid \mathbf{y}_t, \lambda^{(n)} \right] \right\}^{-1} \quad (9)$$

$$s^{2,(n+1)} = \frac{1}{DT} \sum_{t=1}^{T} \{ \left\| \mathbf{y}_t - \overline{\mathbf{y}}^{(n+1)} \right\|^2 - 2E\left[\mathbf{x}_t^{\mathrm{T}} \mid \mathbf{y}_t, \lambda^{(n)} \right] \mathbf{M}^{\mathrm{T},(n+1)} \cdot \left(\mathbf{y}_t - \overline{\mathbf{y}}^{(n+1)} \right) + tr\left(E\left[\mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \mid \mathbf{y}_t, \lambda^{(n)} \right] \mathbf{M}^{\mathrm{T},(n+1)} \mathbf{M}^{(n+1)} \right) \} \quad (10)$$

where

$$E[\mathbf{x}_t \mid \mathbf{y}_t, \lambda^{(n)}] = \Sigma_{\mathbf{x}}^{-1} \mathbf{M}^{\mathrm{T},(n)} \left(\mathbf{y}_t - \overline{\mathbf{y}}^{(n)} \right) \quad (11)$$

$$E[\mathbf{x}_t \mathbf{x}_t^{\mathrm{T}} \mid \mathbf{y}_t, \lambda^{(n)}] = s^{\,2,(n)} \Sigma_{\mathbf{x}}^{-1} + E[\mathbf{x}_t \mid \mathbf{y}_t, \lambda^{(n)}] E[\mathbf{x}_t^{\mathrm{T}} \mid \mathbf{y}_t, \lambda^{(n)}] \quad (12)$$

with $\Sigma_{\mathbf{x}} = s^2\mathbf{I}_K + \mathbf{M}^{\mathrm{T}}\mathbf{M}$, and tr denoting the trace of a matrix. It is shown in [9] 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.

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 pdf $p(\mathbf{o}_t \mid j)$ for state j is assumed to be a mixture of Gaussians:

$$p(\mathbf{o}_t \mid j) = \sum_{k=1}^{M} w_{j,k} N(\mathbf{o}_t \mid m_{j,k}, \Sigma_{j,k}) , \quad (13)$$

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.

In this paper, only the adaptation of Gaussian mean vectors is considered. The Gaussian means are divided into C disjoint clusters, $\{\lambda_c\}_{c=1,\dots,C}$. 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

$$\hat{\mathbf{W}}_c = \arg\max_{\mathbf{W}} f(O \mid \mathbf{W}, \lambda_c) g(\mathbf{W} \mid \Theta_c) \quad . \quad (14)$$

Here we assume that each supervector $\mathbf{m}_c$ that augments the columns of $\mathbf{W}_c$ is generated by a PPCA model given by (5), which has a hidden variable $\mathbf{x}_c$ with a prior parameter set $\theta_c = \{\mathbf{M}_c, \overline{\mathbf{m}}_c, s^2\}$. 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 [5]. Therefore, as expected, a closed form solution for $\mathbf{W}_c$ can be derived. We 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 follows:

$$\begin{aligned} R(\phi, \phi^{(n)}) &= E\left[\log p(O, S, L \mid \Lambda, \phi) \mid O, \phi^{(n)}\right] \\ &\quad + E\left[\log p(\phi, \mathbf{x} \mid \theta) \mid \phi^{(n)}\right] \end{aligned} \quad (15)$$

where S and L represent the state sequence and mixture sequence respectively. From (15) it can be derived that

$$\begin{aligned} R(\phi, \phi^{(n)}) &= \sum_S \sum_L p(S, L \mid O, \Lambda, \phi) \log p(O, S, L \mid \Lambda, \phi) \\ &\quad + E\left[\log\{p(\phi \mid \mathbf{x}, \theta) p(\mathbf{x})\} \mid \phi^{(n)}\right] \end{aligned} \quad (16)$$

Based on (6), it can be found that the relevant object function is given as

$$\begin{aligned} & R(\phi, \phi^{(n)}) = \\ & \sum_{t=1}^{T} \sum_{c=1}^{C} \sum_{k=1}^{M_c} \gamma_k(t) \left[-\frac{1}{2} (\mathbf{o}_t - \mathbf{W}_c \tilde{m}_k)^{\mathrm{T}} \Sigma_k^{-1} (\mathbf{o}_t - \mathbf{W}_c \tilde{m}_k) \right] \quad (17) \\ & + \sum_{c=1}^{C} E\left[-\frac{1}{2s^2} \left\| \mathbf{m}_c - \mathbf{M}_c \mathbf{x}_c - \overline{\mathbf{m}}_c \right\|^2 \middle| \mathbf{m}_c^{(n)} \right] \end{aligned}$$

where $g_k(t)$ represents the mixture occupation probability of the k-th Gaussian in the regression class c at time t given the observation sequence O, and M_c is the number of mixture components in class c. Since the supervector $\mathbf{m}_c$ is not in matrix form, we differentiate the object function with respect to each column of $\mathbf{W}_c$ and setting the result equal to zero vector for each c. This leads to

$$\begin{aligned} & \sum_{t=1}^{T} \sum_{k=1}^{M_c} \gamma_k(t) \tilde{m}_k(j) \Sigma_k^{-1} \mathbf{W}_c \tilde{m}_k + \frac{1}{s^2} \mathbf{W}_c^{(j)} \\ & = \sum_{t=1}^{T} \sum_{k=1}^{M_c} \gamma_k(t) \tilde{m}_k(j) \Sigma_k^{-1} \mathbf{o}_t + \frac{1}{s^2} (\mathbf{M}_{c,j} E\left[\mathbf{x}_c \mid \mathbf{m}_c^{(n)}\right] + \overline{\mathbf{m}}_{c,j}) \quad (18) \\ & j = 1, \dots, n+1 \end{aligned}$$

where $\mathbf{W}_c^{(j)}$ represents the j-th column of $\mathbf{W}_c$, $\mathbf{M}_{c,j}$ denotes the sub-matrix associated to the $\mathbf{W}_c^{(j)}$ elements, and

$$E\left[\mathbf{x}_c \mid \mathbf{m}_c^{(n)}\right] = (s^2 \mathbf{I} + \mathbf{M}_c^{\mathrm{T}} \mathbf{M}_c)^{-1} \mathbf{M}_c^{\mathrm{T}} (\mathbf{m}_c^{(n)} - \overline{\mathbf{m}}_c) . \quad (19)$$

By observing (18), the equation in matrix form can be derived as

$$\begin{aligned} & \sum_{t=1}^{T} \sum_{k=1}^{M_c} \gamma_k(t) \Sigma_k^{-1} \mathbf{W}_c \tilde{m}_k \tilde{m}_k^{\mathrm{T}} + \frac{1}{s^2} \mathbf{W}_c \\ & = \sum_{t=1}^{T} \sum_{k=1}^{M_c} \gamma_k(t) \Sigma_k^{-1} \mathbf{o}_t \tilde{m}_k^{\mathrm{T}} + \frac{1}{s^2} \mathbf{W}_{c,\mathrm{PPCA}} \end{aligned} \quad (20)$$

It is found that (20) can be rewritten as

$$\sum_{k=1}^{M_c} \mathbf{V}_k \mathbf{W}_c \mathbf{D}_k + \frac{1}{s^2} \mathbf{W}_c = \mathbf{Z} + \frac{1}{s^2} \mathbf{W}_{c,\mathrm{PPCA}} \quad (21)$$

where $\mathbf{V}_k = \sum_{t=1}^{T} \gamma_k(t) \Sigma_k^{-1}$, and $\mathbf{D}_k = \tilde{m}_k \tilde{m}_k^{\mathrm{T}}$. 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

$$\mathbf{w}_{c,(i)} = \left(\mathbf{G}_i + \frac{1}{s^2} \mathbf{I}_{n+1} \right)^{-1} \left(\mathbf{z}_{(i)} + \frac{1}{s^2} \mathbf{w}_{c,\mathrm{PPCA},(i)} \right) \quad (22)$$

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. Compared to the standard MLLR re-estimation formula [2], (22) 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 estimation procedure. When the amount of adaptation data is limited, the eigenspace-based transformations dominate the computation of $\mathbf{W}_c$ and allow robust estimation to be made. On the other hand, as the amount of adaptation increases, the MAPLR solution simply converge to the standard MLLR regression matrices and, thus, can utilize the large amount of data more efficiently.

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. 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. It should be noted that each adaptation utterance is of an average length of 1.37 seconds and consists of 4.6 syllables on average and, therefore, may be regarded as a word. 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.

Both types of conventional MLLR approaches using full and diagonal regression matrices were conducted as baseline experiments. The full matrix MLLR was based on a global regression class, while the diagonal matrix MLLR utilized a regression class tree for dynamic regression class generating. As for the proposed MAPLR 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 [8], except that the smoothing procedure in [8] is now merged into the re-estimation process of the MAPLR transformations. In the training phase, for each training speaker all the 120 utterances were used to estimate a speaker-specific global full regression matrix and, then, PCA was performed to extract 109 bases. During adaptation, 50 eigen-transformations were selected to construct the principal subspace, and the PPCA model parameter s was determined empirically.

We performed a series of batch adaptation experiments, conducted in both supervised and unsupervised modes, with various numbers of sentences extracted from the adaptation data. The results for supervised and unsupervised adaptation are summarized in Table 1 and 2 respectively, where "Eigen" denotes the proposed eigenspace-based MAPLR approach. It can be observed that when the amount of adaptation data was strictly limited (3 to 10 sentences), full matrix MLLR gave very poor results, which implied the underlying structure of the acoustic space was severely corrupted by the poorly estimated full regression matrix. On the other hand, the proposed eigenspace-based MAPLR approach not only significantly improved the recognition accuracy, but also outperformed the diagonal matrix MLLR. This showed the advantage of the proposed approach: utilizing the eigenspace of full regression matrices to choose the priors for the MAPLR transformation not only significantly reduced the number of free parameters so that robust estimation could be easily made, but also maintained the precise modeling of the inter-dimensional correlation among the mean parameters as well as the acoustic space structure. It can also be seen that as the amount of adaptation data increased, the performance of the proposed approach converged to that of the conventional full-matrix MLLR. Furthermore, it can be observed from Table 2 that the constraints posed by the eigenspace made the proposed approach more robust to the recognition error, and led to the superior unsupervised adaptation performance over the conventional MLLR approaches.

Sent.	3	5	7	10	15	30	60
Diag	56.18	56.41	56.62	57.12	57.26	59.17	59.98
Full	5.85	21.81	38.15	48.02	55.00	58.67	61.41
Eigen	58.16	57.95	58.41	59.41	59.29	60.24	61.25

Table 1: Supervised batch adaptation performance in syllable accuracy (%). The syllable accuracy is 55.81% for the SI case.

Sent.	3	5	7	10	15	30	60
Diag	55.69	56.06	55.74	56.18	56.55	57.84	58.30
Full	5.16	25.80	39.00	48.43	53.48	57.12	58.92
Eigen	57.19	57.08	57.68	58.34	58.16	58.18	59.22

Table 2: Unsupervised batch adaptation performance.

6. CONCLUSION

In this paper, an eigenspace-based MAPLR approach was proposed. We introduced *a priori* knowledge analysis on training speakers via PPCA so as to construct an eigenspace in which the prior parameters of the MAPLR transformations can be chosen. The PPCA model was further incorporated into the MAPLR formulation and led to a unified framework for transformation estimation. Significant improvements in batch adaptation showed the effectiveness of the proposed approach for rapid model adaptation.

REFERENCES

[1] P. C. Woodland, "Speaker adaptation: techniques and challenges," *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, pp.85-90, 1999.

[2] 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.

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

[4] 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.

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

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

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

[8] 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.

[9] 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.

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