ONLINE ADAPTATION USING TRANSFORMATION SPACE MODEL EVOLUTION

Dong Kook Kim, Young Joon Kim, Woo Hyung Lim and Nam Soo Kim

School of Electrical Engineering and INMC
Seoul National University, Seoul, Korea
email : {dkkim, kjun, whlim}@hi.snu.ac.kr, nskim@snu.ac.kr

ABSTRACT

This paper presents a new approach to online speaker adaptation based on transformation space model evolution. This approach extends the previous idea of speaker space model evolution [11] by applying the a priori knowledge of training speakers to the speaker-dependent maximum likelihood linear regression (MLLR) matrix parameters. A quasi-Bayes (QB) estimation algorithm is devised to incrementally update the hyperparameters of the transformation space model and the regression matrices simultaneously. Experiments on supervised speaker adaptation demonstrate that the proposed approach is more effective compared with the conventional quasi-Bayes linear regression (QBLR) technique when a small amount of adaptation data is available.

1. INTRODUCTION

Maximum likelihood linear regression (MLLR) [1] is considered as one of the most ef£cient methods for adapting continuous density hidden Markov model (CDHMM) to the current speaker and/or acoustic environment. MLLR is a transformation-based adaptation scheme where the overall CDHMM parameters are adapted via a set of linear regression functions estimated according to the maximum likelihood (ML) criterion. MLLR can be successful with a small amount of adaptation data and can operate in all adaptation modes including unsupervised adaptation as well as online adaptation.

Since it is necessary to have suf£cient adaptation data to robustly estimate the MLLR parameters, various adaptation techniques have been suggested to increase the robustness of MLLR for rapid adaptation. The maximum a posteriori linear regression (MAPLR) approach [2] is proposed to improve MLLR adaptation by incorporating a prior distribution for the transformation parameters and estimating the parameters according to the maximum a posteriori (MAP) criterion. Also, eigenspace-based MLLR and MAPLR [3] [4] techniques are developed by introducing the a priori knowledge of training speakers to the speaker-speci£c MLLR matrix parameters based on the principal component analysis (PCA) [5] and probabilistic PCA (PPCA) [6]. Compared with MLLR and MAPLR, These approach have been found to be more effective when a very small amount of adaptation data is available.

Huo and Lee [7] applied the quasi-Bayes (QB) learning framework to incrementally update both the CDHMM parameters and the hyperparameters through a prior evolution procedure. Chien [8] proposed the quasi-Bayes linear regression (QBLR) algorithm for online linear regression adaptation of the CDHMMs and showed that the QBLR is a general framework in which MLLR and MAPLR are treated as special cases. Experimental results demonstrated that the sequential adaptation technique using QBLR is ef£cient and asymptotically converge to the batch learning methods such as MLLR and MAPLR.

Recently, we proposed a rapid speaker adaptation technique based on the PPCA and extended this approach to the latent variable model such as factor analysis (FA) [9] to £nd the speaker space model [10] [11]. Also, we suggested a new approach to online adaptation of the CDHMM mean parameters based on the speaker space model evolution [11].

In this paper, we further extend the ideas of the previous speaker space model evolution technique by applying prior knowledge of the training speakers to the speaker-dependent MLLR matrix parameters. Similar to the speaker space, the transformation space model is established by extracting several principal components among a set of speaker-speci£c transformation matrices. Experiments on supervised speaker adaptation demonstrate that the proposed approach is more effective compared with the QBLR technique when a small amount of adaptation data is available.

2. ONLINE ADAPTATION BASED ON TRANSFORMATION SPACE MODEL EVOLUTION

2.1. Transformation Space Model

Consider an N-state CDHMM with K mixture components, $\lambda = \{\lambda_j\} = \{w_{jk}, \mu_{jk}, \Sigma_{jk}\}, j = 1, \cdots, N, k = 1, \cdots, K$. The state observation probability density function (pdf) of an observation vector $\mathbf{x}_t$ is de£ned to be a mixture of multivariate Gaussians

$$p(\mathbf{x}_t|\lambda_j) = \sum_{k=1}^{K} w_{jk}\mathcal{N}(\mathbf{x}_t; \mu_{jk}, \Sigma_{jk}) \tag{1}$$

where w_{jk} is the weight for the mixture component k in state j with $\sum_{j=1}^{K} w_{jk} = 1$, μ_{jk} is the d-dimensional mean vector and Σ_{jk} is the $d \times d$ covariance matrix.

With the MLLR adaptation technique, we try to adapt the CDHMM mean vector by applying a $d \times (d+1)$ regression matrix $\mathbf{W}$ to the $(d+1) \times 1$ extended mean vector $\xi_{jk} = [1, \mu_{jk}^T]^T$ as follows:

$$\hat{\mu}_{jk} = \mathbf{W}\xi_{jk} \ . \tag{2}$$

For notation simplicity, we assume that only a single global class associated with the regression matrix $\mathbf{W}$ is used for rapid adaptation. The Gaussian distributions can be clustered into several groups and a single transformation matrix is shared by all distributions which belong to the same group.

Let $\mathbf{W}_r, r \in \{1, \cdots, R\}$ be a set of R well trained speaker-dependent (SD) regression matrices, which can be obtained using the standard MLLR technique. Let $\mathbf{w}_r$ be a supervector of dimension $D\ (= d(d+1)\times 1)$ created by collecting the row vectors $\{W_{ri}\}$ of the regression matrix $\mathbf{W}$ such that $\mathbf{w}_r = [W_{r1}, \cdots, W_{rd}]^T$. We assume that a set of SD transformation parameters, $\{\mathbf{w}_r\}$ is generated by a latent variable model such as FA or PPCA with parameters $\phi = \{\mathbf{U}, \bar{\mathbf{w}}, \mathbf{\Lambda}\}$ such that

$$\mathbf{w} \quad = \quad \mathbf{U}\mathbf{v} + \bar{\mathbf{w}} + \epsilon \tag{3}$$

where $\bar{\mathbf{w}}$ is the mean of the supervectors, $\mathbf{v}$ is a latent variable of dimension P with $p(\mathbf{v}) \sim \mathcal{N}(0, \mathbf{I})$, $\mathbf{U}$ is a $D \times P$ matrix that represents the subspace of the transformation parameters, and ϵ is a Gaussian random noise, $p(\epsilon) \sim \mathcal{N}(0, \mathbf{\Lambda})$ independent of $\mathbf{v}$. The PPCA de£nes the noise covariance matrix to be isotropic, i.e., $\mathbf{\Lambda} = \sigma^2\mathbf{I}$. Based on the above assumption, we can derive the conditional distribution of $\mathbf{w}$ given $\mathbf{v}$ by $p(\mathbf{w}|\mathbf{v}) \sim \mathcal{N}(\mathbf{U}\mathbf{v} + \bar{\mathbf{w}}, \mathbf{\Lambda})$ and construct a prior pdf of $\mathbf{w}$ such that

$$g(\mathbf{w}|\phi) \sim \mathcal{N}(\bar{\mathbf{w}}, \mathbf{\Lambda} + \mathbf{U}\mathbf{U}^T)\ . \tag{4}$$

(3) de£nes the transformation space model which characterizes the a priori knowledge of the training speakers associated with a regression matrix $\mathbf{W}$. That is, it describes the prior information on the speaker variability by analyzing the transformation parameters related to each training speaker. It is noted that the transformation space model provides not only the key information of the speaker characteristics but also the prior pdf corresponding to the transformation parameters. This prior pdf enables us to employ the MAP-based speaker adaptation scheme like MAPLR [2]. The transformation space model parameters, ϕ can be estimated using the iterative expectation maximization (EM) algorithm [6] [9].

2.2. QB Learning for Transformation Space Model

We brie¤y review the basic concept and formulation of QB learning for the linear regression parameters, $\mathbf{w}$ [7] [8]. Let $\mathcal{X}^n = \{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_n\}$ be a sequence of independent identically distributed observations statistically related to the regression matrix $\mathbf{w}$ and CDHMM parameters λ. A recursive expression for the a posteriori pdf of $\mathbf{w}$ is given by

$$p(\mathbf{w}|\mathcal{X}^n, \lambda) = \frac{p(\mathcal{X}_n|\mathbf{w}, \lambda)\cdot p(\mathbf{w}|\mathcal{X}^{n-1}, \lambda)}{\int p(\mathcal{X}_n|\mathbf{w}, \lambda)\cdot p(\mathbf{w}|\mathcal{X}^{n-1}, \lambda)\, d\mathbf{w}}\ . \tag{5}$$

This provides the basis for making a recursive Bayesian estimate of $\mathbf{w}$. However, the implementation of this type of recursive Bayesian estimation technique has been found very dif£cult. To alleviate this problem, an approach called the QB learning technique was proposed in [7]. The QB procedure, at each step of recursive Bayes learning, approximates the true posterior density $p(\mathbf{w}|\mathcal{X}^n, \lambda)$ by the closest tractable parametric prior density $g(\mathbf{w}|\phi^{(n)})$ under the criterion that both densities should have the same mode. Here, $\phi^{(n)}$ denotes the updated hyperparameters after observing $\mathcal{X}_n$.

Let us assume that at time instant n, we are given a set of observation vectors $\mathcal{X}_n = \{\mathbf{x}_1^{(n)}, \cdots, \mathbf{x}_{T_n}^{(n)}\}$ and the approximate prior pdf $g(\mathbf{w}|\phi^{(n-1)})$. Since we assume that the transformation parameter $\mathbf{w}$ is generated through a model given by (3), which has a hidden variable $\mathbf{v}$ with the hyperparameter $\phi^{(n-1)}$, the complete-data likelihood for $\mathbf{w}$ can be easily de£ned. Let $(\mathcal{X}_n, \mathbf{S}_n, \mathbf{L}_n)$ denote the complete-data for $\mathcal{X}_n$ in which $\mathbf{S}_n = \{s_t^{(n)}\}$ represents the state sequence and $\mathbf{L}_n = \{l_t^{(n)}\}$ is the mixture component sequence. We can update the approximate posterior density of the current estimate $\mathbf{w}^{(n)}$ and derive the new estimate $\mathbf{w}$ by repeating the following EM steps:

E-step: Compute the auxiliary function

$$\begin{aligned} R_{QB}(\mathbf{w}|\mathbf{w}^{(n)}, \phi^{(n-1)}) = &\ E\ [\ \log p(\mathcal{X}_n, \mathbf{S}_n, \mathbf{L}_n \mid \mathbf{w}, \lambda) \\ &+\ \rho \log g(\mathbf{w}, \mathbf{v} \mid \phi^{(n-1)}) \mid \mathcal{X}_n, \mathbf{w}^{(n)}]\ . \end{aligned} \tag{6}$$

M-step: Choose

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{argmax}\ R_{QB}(\mathbf{w}|\mathbf{w}^{(n)}, \phi^{(n-1)}) \tag{7}$$

where $0 < \rho \leq 1$ is a forgetting factor to reduce the effect of past observations $\mathcal{X}^{n-1}$ relative to the new data $\mathcal{X}_n$. By repeating the above EM iterations, we can get a series of approximate pdf $g(\mathbf{w}|\phi^{(n)})$ whose mode is approaching the mode of the true posterior pdf $p(\mathbf{w}|\mathcal{X}_n, \lambda)$. At the last EM iteration, the set of hyperparameters $\phi^{(n)}$ is computed to satisfy

$$g(\mathbf{w}|\phi^{(n)}) \propto \exp\left\{R_{QB}(\mathbf{w}|\mathbf{w}^{(n)}, \phi^{(n-1)})\right\}\ . \tag{8}$$

Finally the transformation parameters $\hat{\mathbf{w}}$ are updated by taking the mode of $g(\mathbf{w}|\phi^{(n)})$.

2.3. Transformation Space Model Evolution

In this subsection, we derive a formulation for the prior evolution of the transformation space model. Under the speci£cation of the transformation space model for $\mathbf{w}$ in (3), the auxiliary function in the expectation step can be rewritten as

$$\begin{aligned} &R_{QB}(\mathbf{w}|\mathbf{w}^{(n)}, \phi^{(n-1)}) \propto \\ &\quad -\frac{1}{2}\sum_t \sum_{j,k} \gamma_t(j,k) \left[(\mathbf{x}_t^{(n)} - \mathbf{W}\xi_{jk})^T \Sigma_{jk}^{-1} (\mathbf{x}_t^{(n)} - \mathbf{W}\xi_{jk})\right] \\ &\quad + \rho E\ [-\frac{1}{2}(\mathbf{w} - \mathbf{U}^{(n-1)}\mathbf{v} - \bar{\mathbf{w}}^{(n-1)})^T \mathbf{\Lambda}^{-1,(n-1)} \\ &\quad \cdot (\mathbf{w} - \mathbf{U}^{(n-1)}\mathbf{v} - \bar{\mathbf{w}}^{(n-1)}) \mid \mathbf{w}^{(n)}\ ] \end{aligned} \tag{9}$$

where $\gamma_t(j,k) = P(s_t^{(n)} = j, l_t^{(n)} = k|\mathcal{X}_n, \mathbf{w}^{(n)})$ is the posterior probability of being in state j and mixture component k at time t conditioned on that current transformation parameter $\mathbf{w}^{(n)}$ generates $\mathcal{X}_n$. It is rewritten as

$$\begin{aligned} &R_{QB}(\mathbf{w}|\mathbf{w}^{(n)}, \phi^{(n-1)}) \propto \\ &\quad -\frac{1}{2}\sum_t \sum_{j,k} \gamma_t(j,k) \left[(\mathbf{x}_t^{(n)} - \mathbf{C}_{jk}\mathbf{w})^T \Sigma_{jk}^{-1} (\mathbf{x}_t^{(n)} - \mathbf{C}_{jk}\mathbf{w})\right] \\ &\quad + \rho E\ [-\frac{1}{2}(\mathbf{w} - \mathbf{U}^{(n-1)}\mathbf{v} - \bar{\mathbf{w}}^{(n-1)})^T \mathbf{\Lambda}^{-1,(n-1)} \\ &\quad \cdot (\mathbf{w} - \mathbf{U}^{(n-1)}\mathbf{v} - \bar{\mathbf{w}}^{(n-1)}) \mid \mathbf{w}^{(n)}\ ] \end{aligned} \tag{10}$$

where $\mathbf{C}_{jk}$ is a $d \times d(d+1)$ matrix composed of the components of the extended mean vector ξ_{jk} as shown at the top of the next page [8]:

After some manipulation, the exponential of expectation function multiplied by a normalization constant C, i.e., $C\ \cdot$

$$\mathbf{C}_{jk} = \begin{bmatrix} 1 & \mu_{jk1} & \cdots & \mu_{jkd} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \mu_{jk1} & \cdots & \mu_{jkd} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & \mu_{jk1} & \cdots & \mu_{jkd} \end{bmatrix}.$$

$\exp\ \{R_{QB}(\mathbf{w}|\mathbf{w}^{(n)}, \phi^{(n-1)})\}$ can be expressed in the normal distribution as follows:

$$C \cdot \exp\left\{R_{QB}(\mathbf{w}|\mathbf{w}^{(n)}, \phi^{(n-1)})\right\} \propto \exp\left\{-\frac{1}{2}(\mathbf{w}-\mathbf{m})^T \mathbf{\Phi}^{-1}(\mathbf{w}-\mathbf{m})\right\} \tag{11}$$

where

$$\begin{aligned} \mathbf{m} = \ & \mathbf{\Phi} \sum_t \sum_{j,k} \gamma_t(j,k) \mathbf{C}_{jk}^T \Sigma_{jk}^{-1} \mathbf{C}_{jk} \mathbf{x}_t^{(n)} \\ & + \ \rho\, \mathbf{\Phi} \mathbf{\Lambda}^{-1,(n-1)} \left(\mathbf{U}^{(n-1)} \hat{\mathbf{v}} + \bar{\mathbf{w}}^{(n-1)}\right) \end{aligned} \tag{12}$$

$$\mathbf{\Phi} = \left(\sum_t \sum_{j,k} \gamma_t(j,k) \mathbf{C}_{jk}^T \Sigma_{jk}^{-1} \mathbf{C}_{jk} + \rho\, \mathbf{\Lambda}^{-1,(n-1)} \right)^{-1} \tag{13}$$

with

$$\begin{aligned} \hat{\mathbf{v}} = \ & E\left[\mathbf{v}|\mathbf{w}^{(n)}\right] \\ = \ & \left(\mathbf{I} + \mathbf{U}^{T,(n-1)} \mathbf{\Lambda}^{-1,(n-1)} \mathbf{U}^{(n-1)}\right)^{-1} \cdot \\ & \mathbf{U}^{T,(n-1)} \mathbf{\Lambda}^{-1,(n-1)} \left(\mathbf{w}^{(n)} - \bar{\mathbf{w}}^{(n-1)}\right) . \end{aligned} \tag{14}$$

(11) belongs to the same distribution family as $g(\mathbf{w}|\phi)$ in (4) with $\bar{\mathbf{w}} = \mathbf{m}$ and $\mathbf{\Lambda} + \mathbf{U}\mathbf{U}^T = \mathbf{\Phi}$. Here we assume that the hyperparameter $\mathbf{U}$ is £xed during prior evolution and the prior evolution for $\mathbf{\Lambda}$ is approximated by $\mathbf{\Phi}$. As a result, $g(\mathbf{w}|\phi^{(n)})$ in (8) can be denoted with the hyperparameters $\phi^{(n)}$ as follows:

$$\begin{aligned} \bar{\mathbf{w}}^{(n)} &= \mathbf{m} \\ \mathbf{\Lambda}^{(n)} &= \mathbf{\Phi} \\ \mathbf{U}^{(n)} &= \mathbf{U}^{(n-1)} \end{aligned} \tag{15}$$

such that the transformation space model is evolving accordingly. For the prior evolution of the PPCA model, similar results are obtained, $\bar{\mathbf{w}}^{(n)} = \mathbf{m}$ and $\sigma^{2,(n)} = trace\{\mathbf{\Phi}\}/D$ with $\mathbf{\Lambda}^{(n-1)} = \sigma^{2,(n-1)}\mathbf{I}$. After completing the transformation space model evolution procedure, the QB estimated transformation parameter $\hat{\mathbf{w}}^{(n)}$ at time instance n is obtained by just taking the mode of the evolved prior pdf as follows:

$$\hat{\mathbf{w}}^{(n)} = \mathbf{m}\,. \tag{16}$$

The proposed transformation space model evolution approach is similar to the QBLR technique [8]. Compared with the QBLR technique, it is noted in (12) that the parameter $\mathbf{m}$ is obtained by incorporating the prior transformation model estimated within the transformation space into the MLLR estimation procedure.

3. EXPERIMENTS AND RESULTS

3.1. Speech Database and Recognition System

Performance of the proposed online adaptation approach was evaluated with a number of supervised adaptation experiments on the task of continuous Korean digit recognition. All the training and test data used for building the baseline recognition system were recorded in a quite environment. Utterances from 105 speakers (68 males and 37 females) constructed the training data and those from the other 35 speakers (25 males and 13 females) were used for evaluation. Each speaker contributed 30∼40 sentences consisting of 3∼7 digits and each sentence had an average length of 1.3 seconds. Each digit was modeled by a seven-state left-to-right HMM without skips and the 3 silence types were modeled by a one-state HMM. We trained the CDHMM parameters by varying the number of Gaussian components in each state from one to two. The speech signal was sampled at 8 kHz and segmented into 30 ms long frame at every 10 ms with 20 ms overlap. Each frame was parameterized by a 24-dimensional feature vector consisting of 12 mel-frequency cepstral coef£cients and their £rst-order time derivatives. In the recognition experiments, we drew $1 \sim 10$ sentences from each target speaker for adaptation, and performed the recognition test on the remaining sentences. All the adaptation procedures were performed in a supervised manner using an exact transcription of each data. The speaker-independent (SI) system with a single Gaussian and two mixture Gaussians produced 87.58 % and 89.60 % of word recognition rates, respectively.

To obtain the transformation space model by latent variable models, we £rst trained a set of SI models over the speech from all the training speakers. To obtain the 105 SD regression matrices, the conventional MLLR adaptation approach with block-diagonal matrix was performed for each training speaker with the speaker-speci£c data. Considering the amount of adaptation data available, only a single global regression class was speci£ed for MLLR. Consequently, the transformation supervectors are of dimension $D = \{13 \times 12 \times 2\} = 312$. Before applying the latent variable model techniques, the supervectors were normalized by their standard deviation to prevent the variables with large absolute value from dominating the analysis. We obtained the estimates of the parameters for the latent variable model, $\{\bar{\mathbf{w}}, \mathbf{U}, \mathbf{\Lambda}\}$ with dimension $P = \{30, 40, 50\}$ for each transformation supervector using the EM algorithm.

3.2. Evaluation of Batch Adaptation

First, we performed supervised batch speaker adaptation experiments by using four different methods, ; 1) conventional MLLR adaptation method [1], 2) MAPLR method [2], 3) PPCA-based adaptation method (or eigenspace-based MAPLR) [4], 4) FA-based adaptation method. Tables I and II show the performance of the MLLR, MAPLR, PPCA and FA-based adaptation techniques for the single Gaussian and two mixture Gaussian HMM's, respectively. Word recognition rates are displayed against the number of adaptation sentences. Here "PPCA (P=40)" and "FA (P=30)" denote the PPCA and FA methods with subspace dimension 40 and 30, respectively.

The MLLR approach performed poorly for a small adaptation data size, but it improved the performance as more data became available. The MAPLR performed better than MLLR because incorporating a prior density is helpful in the regression matrix estimation when the adaptation data is very little. On the other hand, the PPCA and FA-based approaches possess both the rapid and consistent adaptation properties, and as a result they can ef£ciently perform speaker adaptation not only for very little adaptation data but also when a large amount of data is available.

Table I
Word recognition rate (%) for batch adaptation experiments in a single Gaussian system.

no. of sent.	1	2	4	6	8	10
MLLR	49.22	79.61	87.87	90.32	90.38	90.36
MAPLR	87.39	89.16	90.16	90.51	90.34	90.63
PPCA (P=40)	89.30	89.91	90.51	90.95	90.70	91.01
FA (P=30)	89.18	90.07	90.45	90.92	90.72	90.84

Table II
Word recognition rate (%) for batch adaptation experiments in two mixture Gaussian system.

no. of sent.	1	2	4	6	8	10
MLLR	68.25	85.82	90.41	91.22	91.65	91.73
MAPLR	89.55	90.72	91.30	91.51	91.90	92.03
PPCA (P=50)	90.84	91.24	91.65	91.92	91.92	92.11
FA (P=20)	90.68	91.09	91.57	91.80	92.03	92.07

3.3. Evaluation of Online Adaptation

We implemented the QBLR scheme to compare the performance of the proposed online adaptation schemes. Fig. 1 shows the performance of the QBLR, PPCA and FA-based online adaptation techniques with various amount of adaptation data. For online adaptation, the parameters were updated for each adaptation sentence and the forgetting mechanism with $\rho = 1$ was applied. From Fig. 1, we can £nd that the proposed online approaches perform better than the QBLR and achieves a similar rapid adaptation performance as that of the batch PPCA and FA approaches for a small amount of adaptation data. Also it maintains a good asymptotic convergence property as the data size grows. The experimental results show that the online adaptation approaches which are based on the latent variable models perform well especially for a small adaptation data size.

4. CONCLUSION

This paper has presented a novel online adaptation algorithm based on transformation space model evolution. We have applied the latent variable model to £nd the transformation space model associated with the MLLR matrix parameters as well as to obtain their prior pdfs. We have extended the recursive QB learning approach to online speaker adaptation using the transformation space model. From the results of a set of online speaker adaptation experiments, we can conclude that the proposed approach is effective in speaker adaptation especially for sparse data.

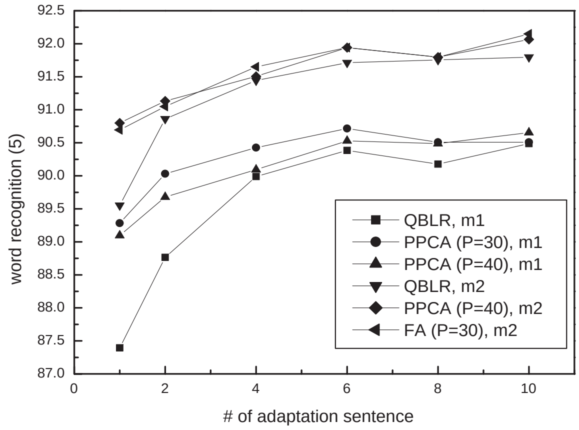

Fig. 1. Word recognition rate for online adaptation experiments in a single Gaussian (m1) and two mixture Gaussian system (m2).

5. REFERENCES

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

[2] W. Chou, "Maximum a posteriori linear regression with elliptically symmetric matrix variate priors," in *Proc. Eur. Conf. Speech Communication Technology*, Budapest, Hungary, Sep. 1999.

[3] K.-T. Chen, W.-W. Liau, H.-M. Wang, and L.-S. Lee, "Fast speaker adaptation using eigenspace-based maximum likelihood linear regression," in *Proc. Int. Conf. Spoken Language Processing*, Beijing, China, pp. 742-745, Oct. 2000.

[4] K.-T. Chen and H.-M. Wang, "Eigenspace-based maximum a posteriori linear regression for rapid speaker adaptation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Singnal Processing*, Salt Lake City, USA, May 2001.

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

[6] M. Tipping and C. Bishop, "Mixtures of probabilistic principal comonent analysers," Neural *Computation*, vol. 11, no. 2, pp. 443-482, 1999.

[7] Q. Huo and C.-H. Lee, "On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate," *IEEE Trans. Speech and Audio Proc.*, vol. 5, pp. 161-172, Mar. 1997.

[8] J.-T. Chien, "Quasi-Bayes linear regression for sequential learning of hidden Markov models," *IEEE Trans. Speech and Audio Proc.*, vol. 10, pp. 268-278, July 2002.

[9] D. Rubin and D. Thayer, "EM algorithms for factor analysis," *Psychometrika*, vol. 47, pp. 69-76, 1982.

[10] D. K. Kim and N. S. Kim, "Baysian speaker adaptation based on probabilistic principal component analysis", in *Proc. Int. Conf. Spoken Language Processing*, Beijing, China, pp. 734-737, Oct. 2000.

[11] D. K. Kim and N. S. Kim, "Online adaptation of continuous density hidden Markov models based on speaker space model evolution", in *Proc. Int. Conf. Spoken Language Processing*, Denver, USA, 2002.