LETTER

Speaker Adaptation Based on PPCA of Acoustic Models in a Two-Way Array Representation

Yongwon JEONG†(a), Member

SUMMARY We propose a speaker adaptation method based on the probabilistic principal component analysis (PPCA) of acoustic models. We define a training matrix which is represented in a two-way array and decompose the training models by PPCA to construct bases. In the two-way array representation, each training model is represented as a matrix and the columns of each training matrix are treated as training vectors. We formulate the adaptation equation in the maximum a posteriori (MAP) framework using the bases and the prior.

key words: expectation-maximization, probabilistic principal component analysis, speaker adaptation, speech recognition

1. Introduction

In automatic speech recognition using hidden Markov models (HMMs) [1], speaker adaptation techniques [2] are used to improve the performance of a speaker-independent (SI) HMM using adaptation utterances from the target speaker. One class of speaker adaptation techniques for continuous density HMMs (CDHMMs), speaker clustering based techniques such as reference speaker weighting (RSW) [3], cluster adaptive training (CAT) [4], or eigenvoice adaptation [5], expresses the model of the target speaker as a weighted sum of canonical models or bases. In RSW, training models become canonical models and the weight for the canonical models during adaptation is estimated in the maximum likelihood (ML) framework with the nonnegativity constraint on the weight coefficients while the coefficients sum to one. In CAT, the canonical models are derived from the cluster analysis of individual speaker models and in the eigenvoice approach, the canonical models are derived from the principal component analysis (PCA) [6] of training speaker models. The eigenvoice approach is also used in transformation-based speaker adaptation techniques. A representative transformation-based adaptation technique is maximum likelihood linear regression (MLLR) adaptation [7]. In MLLR adaptation, the model for the target speaker is obtained by linearly transforming an SI model by maximizing the likelihood of adaptation data. Chen et al. [8] propose eigenspace-based MLLR where the eigenvoice approach is applied to MLLR adaptation; the transformation matrix for the target speaker is expressed as a linear combination of bases that are derived from the PCA of the transformation matrices of training speakers.

Probabilistic decomposition techniques such as probabilistic PCA (PPCA) [9] can be used to build bases in basis-based adaptation techniques. In Kim and Kim [10], the authors apply PPCA to decompose acoustic models in the eigenvoice framework and formulate the speaker adaptation equation in the maximum a posteriori (MAP) framework. In Chen and Wang [11], the authors apply PPCA to the eigenspace-based MLLR technique; the bases are derived by PPCA and they formulate the speaker adaptation equation in the MAP framework, thus the approach is called eigenspace-based MAP linear regression (MAPLR). In Schuster et al. [12], the authors apply mixtures of PPCA (MPPCA) [13] to acoustic modeling with the constrained full covariance whose complexity can be controlled. The MPPCA is a mixture of Gaussian models in the PCA framework.

In this letter, we improve the performance of eigenvoice adaptation by applying PPCA to acoustic models of training speakers expressed in a two-way array and formulating the adaptation equation in the maximum a posteriori (MAP) framework using the basis vectors and the prior distribution of the speaker weight. In representing the training models in a two-way array, each training model is expressed as a matrix and the columns of training models become training vectors. The motivation for representing acoustic models in a two-way array in the basis-based adaptation framework is that in building bases using PCA, the number of samples used to estimate the covariance matrix increases, whereas the dimension of samples decreases. Here, speaker adaptation is performed by updating the mean vectors of output distributions among CDHMM parameters.

This letter is organized as follows. Section 2 explains adaptation techniques based on the PCA of acoustic models. In Sect. 3, we explain the expectation-maximization (EM) approach to PCA. In Sect. 4, we propose the application of this approach to the decomposition of acoustic models and we present the speaker adaptation method in the MAP framework using the PPCA model. Section 5 presents the experiments, and Sect. 6 presents the conclusion of this work.

2. Speaker Adaptation Using the PCA Model

This section explains two speaker adaptation techniques using PCA in the ML framework, which are compared with the proposed method.
2.1 Decomposition of Acoustic Models Based on a Vector Representation of HMM Mean Parameters

First, we describe the speaker adaptation based on PCA of training models expressed in supervectors, i.e., eigenvoice adaptation [5]. Let $s$ be the index for training speakers ($s = 1, \cdots, S$). In eigenvoice adaptation, HMM mean parameters of the $s$th training speaker are expressed as an $RD \times 1$ vector:

$$
y_{s} = \begin{pmatrix}
y_{s,1} \\
\vdots \\
y_{s,r} \\
\vdots \\
y_{s,R}
\end{pmatrix}
$$

where $y_{s,r} \in \mathbb{R}^{D \times 1}$ denotes the HMM mean vector corresponding to the $r$th Gaussian mixture component ($r = 1, \cdots, R$ where $R$ is the number of tied states $\times$ number of mixtures/state) and $D$ is the dimension of the HMM mean vector. The set of training models $\{y_{1}, \cdots, y_{S}\}$ is decomposed by PCA to build basis vectors $\{\phi_{1}, \cdots, \phi_{K}\}$ ($\phi_{k} \in \mathbb{R}^{RD \times 1}$). The model of a new speaker is expressed as a linear combination of basis vectors plus the mean of the training models:

$$
y_{\text{new}} = \sum_{k=1}^{K} \phi_{k} x_{k} + \bar{y}
$$

where $\bar{y} = (1/S) \sum_{s} y_{s}$. Given adaptation data, the $K \times 1$ weight vector $x = [x_{1} \cdots x_{K}]^{\top}$ can be obtained in an ML criterion; for more details on eigenvoice adaptation, please refer to Kuhn et al. [5].

2.2 Decomposition of Acoustic Models Based on a Two-Way Array Representation of HMM Mean Parameters

Second, we describe the speaker adaptation based on PCA of training models expressed in matrices. HMM mean parameters of the $s$th training speaker are expressed in a matrix $Y_{s}^{R \times D}$. All the training models are arranged as

$$
Y_{\text{training}}^{R \times (SD)} = [Y_{1} Y_{2} \cdots Y_{S}]
$$

and each column vector is treated as a sample; we call $Y_{\text{training}}$ the “two-way array” representation of training models. Expressing a training model in a matrix has an advantage over expressing it in an $RD \times 1$ vector: the number of samples increases from $S$ to $SD$ while the dimension of the samples reduces from $RD$ to $R$. Applying PCA to the training models produces the basis vectors $W = [w_{1} \cdots w_{K}]$ where $w_{k} \in \mathbb{R}^{R \times 1}$ denotes the $k$th dominant eigenvector of the sample covariance matrix:

$$
C_{Y} = \frac{1}{S-1} \sum_{s=1}^{S} (Y_{s} - \bar{Y})(Y_{s} - \bar{Y})^{\top}
$$

where $\bar{Y} = (1/S) \sum_{s} y_{s}$. Because $(Y_{s} - \bar{Y})(Y_{s} - \bar{Y})^{\top} = \sum_{j=1}^{D} (y_{s,j} - \bar{y}_{j})(y_{s,j} - \bar{y}_{j})^{\top}$ where $y_{s,j}$ and $\bar{y}_{j}$ denote the $j$th column vectors of $Y_{s}$ and $\bar{Y}$, respectively, $C_{Y}$ can be rewritten as $C_{Y} = 1/(S-1) \sum_{s=1}^{S} \sum_{j=1}^{D} (y_{s,j} - \bar{y}_{j})(y_{s,j} - \bar{y}_{j})^{\top}$. Thus, the approach described here is PCA applied on vectors $y_{s,j}, s = 1, \cdots, S, j = 1, \cdots, D$. Using the bases, we define the model of a new speaker as

$$
Y_{\text{new}} = WX + Y_{SI}
$$

where $Y_{SI}$ denotes the mean vectors of an SI HMM expressed in an $R \times D$ matrix. Thus, the updated model for the new speaker is expressed as a linear combination of bases plus an SI model. As a result, the linear combination part, $WX$, expresses the deviation from an SI model. For given adaptation data $O = \{o_{t}, t = 1, \cdots, T\} (o_{t} \in \mathbb{R}^{RD \times 1})$, the weight matrix $X$ that maximizes the likelihood of the adaptation data can be estimated by iteratively maximizing the auxiliary function as in Leggetter and Woodland [7] using the expectation-maximization (EM) algorithm [14]:

$$
Q(\lambda, \bar{\lambda}) = \sum_{\theta \in \Theta} f(O, \theta|\lambda) \log f(O, \theta|\lambda)
$$

where $\lambda$ and $\bar{\lambda}$ denote the current and re-estimated sets of model parameters, respectively, $\theta = \{\theta_{1}, \cdots, \theta_{T}\}$ the state sequence, and $\Theta$ the set of all possible state sequences ($f(\cdot)$ denotes a probability distribution function (p.d.f.)). The auxiliary function can be written as [7]

$$
Q(\lambda, \bar{\lambda}) = -\frac{1}{2} f(O|\lambda) \times \sum_{t=1}^{T} \sum_{r=1}^{R} \gamma_{r}(t) (D \log(2\pi) + \log|C_{r}| + h(o_{t}, r))
$$

where $\gamma_{r}(t)$ is the a posteriori probability of occupying mixture component $r$ at $t$ given $O$:

$$
\gamma_{r}(t) = \frac{1}{f(O|\lambda)} \sum_{\theta \in \Theta} f(O, \theta_{t} = r|\lambda)
$$

and $C_{r}$ is the covariance matrix of Gaussian mixture component $r$ of an SI HMM (which is assumed to be a diagonal matrix). The part that contains the weight matrix is given by $h(o_{t}, r)$:

$$
h(o_{t}, r) = (o_{t} - s_{r}(X))^{\top} C_{r}^{-1}(o_{t} - s_{r}(X))
$$

where $s_{r}(X) = [w_{r}X + y_{SL,r}]^{\top}$ denotes the vector consisting of the elements of $Y_{\text{new}}$ that correspond to the $r$th mixture component, i.e., the $r$th row vector of $Y_{\text{new}}$ ($w_{r}$ and $y_{SL,r}$ denote the $r$th row vectors of $W$ and $Y_{SI}$, respectively). Discarding the terms that are independent of the model parameters, we obtain

$$
Q(\lambda, \bar{\lambda}) = -\frac{1}{2} f(O|\lambda) \sum_{t=1}^{T} \sum_{r=1}^{R} \gamma_{r}(t) (o_{t} - s_{r}(X))^{\top} C_{r}^{-1}(o_{t} - s_{r}(X)).
$$

The equation for finding the weight is obtained by setting
∂Q(λ, λ)/∂X = 0:

\[ \sum_{t=1}^{T} \sum_{r=1}^{R} \gamma_r(t) C_r^{-1} X^T w_r^j w_r \]  

\[ = \sum_{t=1}^{T} \sum_{r=1}^{R} \gamma_r(t) C_r^{-1} (o_r - y_{SL,r}) w_r. \]  

The above equation can be solved for X in a similar way as the technique used in MLLR adaptation [7]. We define the following: the right-hand side of Eq. (11) is defined as Z and

\[ V_r = \sum_{t=1}^{T} \gamma_r(t) C_r^{-1} \]  

\[ G_{(j)} = \sum_{r=1}^{R} \nu_r(j, j) D_r, \quad D_r = w_r^T w_r \]

where \( \nu_r(j, j) \) denotes the \((j, j)\) element of \( V_r \). Then, \( X \) can be computed by

\[ x_{(j)} = G_{(j)}^{-1} z_{(j)}^T, \quad j = 1, \ldots, D \]  

where \( x_{(j)} \) denotes the \( j\)th column vector of \( X \) (the weight matrix for the target speaker) and \( z_{(j)} \) the \( j\)th row vector of \( Z \). This method is similar to the 2DPCA-based approach in Jeong [15] where training matrices in their own forms are decomposed by PCA.

3. PCA by Expectation-Maximization

PCA can also be formulated in a probabilistic approach. PPCA assumes the following model for an observation vector \( y \in \mathbb{R}^{1 \times 1} \):

\[ y = W_{PPCA} x + y_{mean} + v \]  

where \( W_{PPCA} \in \mathbb{R}^{K \times 1} \) denotes the transformation, \( x \in \mathbb{R}^{K \times 1} \) the latent variable, \( y_{mean} \) the mean of the observation, and \( v \) the noise. The above equation is a compact representation of \( y \) using \( x \) (usually \( K \ll R \)). In PPCA, p.d.f.s are assumed for \( x \) and \( v \) [13]: \( x \) values are independent and identically distributed (i.i.d.) according to \( N(0, I) \), and \( v \) values are independent of \( x \) and are i.i.d. according to \( N(0, \sigma_v^2 I) \). Under these conditions,

\[ y|x \sim N(W_{PPCA} x + y_{mean}, \sigma_v^2 I). \]  

Due to the diagonal constraint on the covariance matrix of \( v \), \( y \) becomes conditionally independent given \( x \). Consequently, \( x \) represents the correlation between observation variables, and \( v \) expresses the independent noise present in \( y \). This is different from PCA, in which the latent variable and noise are not distinguished from one another. Integrating out \( x \) in Eq. (15) gives

\[ y \sim N(y_{mean}, G) \]  

where \( G = W_{PPCA} W_{PPCA}^T + \sigma_v^2 I \). Therefore, due to Eqs. (15) and (16) and using Bayes’ rule,

\[ x|y \sim N(H^{-1} W_{PPCA}^T (y - y_{mean}), \sigma_v^2 H^{-1}) \]  

where \( H = W_{PPCA} W_{PPCA}^T + \sigma_v^2 I \).

For given observations \( \{y_n, n = 1, \ldots, N\} \), the ML estimate of \( y_{mean} \) is given by the sample mean: \( \bar{y} = (1/N) \sum_n y_n \), and \( W_{PPCA} \) and \( \sigma_v^2 \) can be computed by jointly maximizing the likelihood of \( x \) and \( y \). With initial \( W_{PPCA} \) (the result from PCA is used in our experiments) and \( \sigma_v^2 \) value (0.01 is used in our experiments), the following steps are iterated until the model parameters converge.

E-step: Compute \( (x_n) \) and \( (x_n x_n^T) \) given \( W_{PPCA} \) and \( \sigma_v^2 \) (\( \cdot \) denotes the expectation): according to Eq. (17),

\[ (x_n) = H^{-1} W_{PPCA}^T (y_n - \bar{y}) \]  

\[ (x_n x_n^T) = \sigma_v^2 H^{-1} + (x_n)(x_n)^T. \]

M-step: Update \( W_{PPCA} \) and \( \sigma_v^2 \): the model parameters are updated by maximizing the expectation of the log-likelihood of complete data \( \langle L_c \rangle \) with respect to \( W_{PPCA} \) and \( \sigma_v^2 \). The expectation of \( L_c \) is given by

\[ \langle L_c \rangle = -N \sum_{n=1}^{N} \left( R \log(\sigma_v^2) + \frac{1}{2} |\text{tr}[x_n x_n^T]| + \frac{1}{2 \sigma_v^2} ||y_n - \bar{y}||^2 \right) \]

\[ -\frac{1}{\sigma_v^2} \langle x_n \rangle^T W_{PPCA}^T (y_n - \bar{y}) \]

\[ + \frac{1}{2 \sigma_v^2} \text{tr}[W_{PPCA} W_{PPCA}^T (x_n x_n^T)] \]  

where \( \text{tr}[\cdot] \) denotes the trace of a matrix. Setting \( \partial_{W} \langle L_c \rangle = 0 \) and \( \partial_{\sigma_v^2} \langle L_c \rangle = 0 \) yield:

\[ W_{PPCA,updated} = \left( \sum_{n=1}^{N} (y_n - \bar{y}) (x_n)^T \right) \left( \sum_{n=1}^{N} x_n x_n^T \right)^{-1} \]  

\[ \sigma_v^{2, updated} = \frac{1}{RN} \sum_{n=1}^{N} \left( ||y_n - \bar{y}||^2 - 2 \langle x_n \rangle^T W_{PPCA}^T y_n - \bar{y} \right) \]

\[ + \text{tr}[\langle x_n x_n^T \rangle W_{PPCA}^T W_{PPCA}]. \]

For more details on PPCA, please refer to Tipping and Bishop [13].

4. Speaker Adaptation Using the PPCA Model

We apply PPCA to \( Y_{training} \) of Eq. (3), treating each column vector as a sample. After obtaining \( W_{PPCA} \) and \( \sigma_v^2 \), the weight for each column vector, \( x_{s,j} \), is given by

\[ x_{s,j} = H^{-1} W_{PPCA}^T (y_{s,j} - \bar{y}), \]  

\[ s = 1, \ldots, S, \quad j = 1, \ldots, D \]

where \( y_{s,j} \) denotes the \((s, j)\) column vector of \( Y_{training} \) and \( \bar{y} = 1/(SD) \sum_s \sum_j y_{s,j} \). Unlike PCA, the uncertainty associated with the latent variable is modeled. Grouping \( x_{s,j} \)’s for
Thus, using the bases, the $s$th training model is approximated as
\[ Y_s = W_{\text{PPCA}}X_s + [\hat{y}_1 \cdots \hat{y}]. \]  
(23)

Based on Eqs. (5) and (23), we express the model of a new speaker as
\[ Y_{\text{new}} = W_{\text{PPCA}}X + Y_{\text{SL}}. \]  
(24)

That is, the model of a new speaker is expressed as the transformation of the latent matrix $X$ for the new speaker plus an SI model. For adaptation data $O$, we estimate the weight in a MAP criterion:
\[ \lambda_{\text{MAP}} = \arg\max_{\lambda} g(\lambda|O) \]  
\[ = \arg\max_{\lambda} f(O|\lambda) g(\lambda). \]  
(25)

The MAP estimate of the weight matrix can be obtained by maximizing the following auxiliary function [16]:
\[ Q(\lambda, \hat{\lambda}) = \log p(X) - \frac{1}{2} \sum_{t=1}^{T} \sum_{r=1}^{R} \gamma_r(t)(o_r - s_r(X))^\top C_r^{-1}(o_r - s_r(X)). \]  
(26)

where $s_r(X) = [w_{\text{PPCA}, r} X + y_{\text{SL}, r}]^\top$ (the $r$th row vector of $W_{\text{PPCA}}$). Let the prior density in Eq. (26) be a matrix-variate normal [17]:
\[ p(X) \propto \frac{1}{|\Psi|^{K/2}|\Sigma|^{D/2}} \times \exp\left\{ -\frac{1}{2} \text{tr}\left[ (X - X_{\text{mean}})^\top \Sigma^{-1} (X - X_{\text{mean}}) \Psi^{-1} \right] \right\} \]  
(27)

where $X_{\text{mean}} \in \mathbb{R}^{K \times D}$ is the mean, and $\Psi \in \mathbb{R}^{D \times D}$ and $\Sigma \in \mathbb{R}^{K \times K}$ are the covariance matrices. If $X$ follows a matrix-variate normal as above, then $p(\text{vec}(X)) \sim N(\text{vec}(X_{\text{mean}}), \Psi \otimes \Sigma)$ where vec(·) denotes the vectorization of a matrix and $\otimes$ the Kronecker product. Hyperparameters are estimated from $\{X_1, \cdots, X_S\}$ as follows. If $X$ is a matrix-variate normal and $\Psi$ is known, then $X_{\text{mean}}$ and $\Sigma$ can be easily estimated [17]. Therefore, $\Psi$ is set to the identity matrix as in Siohan et al. [18], and the remaining hyperparameters are estimated as
\[ X_{\text{mean}} = \bar{X} = \frac{1}{S} \sum_{s=1}^{S} X_s, \]  
(28)
\[ \hat{\Sigma} = \frac{1}{S - 1} \sum_{s=1}^{S} (X_s - \bar{X}) (X_s - \bar{X})^\top. \]  

Then, Eq. (26) becomes
\[ Q(\lambda, \hat{\lambda}) \]  
(29)

\[ = -\frac{1}{2} \sum_{t=1}^{T} \sum_{r=1}^{R} y_r(t)(o_r - s_r(X))^\top C_r^{-1}(o_r - s_r(X)) \]  
\[ - \frac{1}{2} \text{tr}[(X - \bar{X})^\top \hat{\Sigma}^{-1} (X - \bar{X})]. \]  

Setting $\partial Q(\lambda, \hat{\lambda})/\partial X = 0$ yields
\[ \sum_{t=1}^{T} \sum_{r=1}^{R} y_r(t) C_r^{-1} X^\top w_{\text{PPCA}, r}^\top + (X - \bar{X})^\top \hat{\Sigma}^{-1} \]  
\[ = \sum_{t=1}^{T} \sum_{r=1}^{R} y_r(t) C_r^{-1} (o_r - y_{\text{SL}, r}) w_{\text{PPCA}, r}. \]  
(30)

The above equation can be solved for $X$ in a similar way as in Chou [16]. We define
\[ \Sigma_{(j)} = \frac{1}{S - 1} \sum_{s=1}^{S} (x_{s,j} - \bar{x}_j)(x_{s,j} - \bar{x}_j)^\top \]  
(31)

where $x_{s,j} = (1/S) \sum_{i=1}^{S} x_{s,i,j}$. $V_r$, $D_r$, $G_{(j)}$, and $Z$ are defined in the same way as Eq. (12) with $w_r$ replaced by $w_{\text{PPCA}, r}$. Then, $X$ can be computed by
\[ x_{(j)} = (G_{(j)} + \Sigma_{(j)}^{-1}) z_{(j)}, \quad j = 1, \cdots, D \]  
(32)

where $x_{(j)}$ denotes the $j$th column vector of $X$, the weight matrix for the target speaker.

The marginalization of $x$ in Eq. (21) can be done as in Hahn et al. [19], where the authors apply variational Bayesian linear regression (VBLR) to feature space speaker adaptation (i.e., transformation-based speaker adaptation), extending the model space VBLR [20]. They propose feature space VBLR (FVBLR) and additionally present the speaker adaptation method that combines the feature space and model space VBLRs, which showed better performance. However, in our case, the relationship between the prior distributions of $x$ and $X$ are not clear and for the purpose of speaker adaptation, the prior distribution in Eq. (27) gives a good estimate in our experiments.

5. Experiments

We used the Wall Street Journal (WSJ) corpus WSJ0 with 5k vocabulary [21] for continuous speech recognition experiments. In the training stage, we used 12,754 utterances of 101 training speakers from the corpus. We used the 39-D vector as the acoustic feature vector consisting of 13-D mel-frequency cepstral coefficients (MFCCs) (9th-order to 12th-order coefficients), derivative coefficients, and acceleration coefficients. We built tied-state triphone HMMs with 3,473 tied states and a mixture of eight Gaussian distributions ($R = 27,784$). We obtained speaker-adapted (SA) models for each training speaker by transforming the SI HMM by MLLR adaptation + MAP adaptation. These 101 SA models were used as training models to build bases ($S = 101$).

For adaptation tests, we used the November 92 5k
Table 1  Word recognition accuracy (%) of updated models.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of adaptation sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
</tr>
<tr>
<td>Two-way &amp; PPCA (MAP)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Two-way &amp; PCA (ML)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td>PCA (eigenvoice)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
</tbody>
</table>

As the baseline, the word recognition accuracy of the SI model is 91.54%. Table 1 shows the word recognition accuracy of adapted models. All three methods exhibit good performance compared to the SI model, and the proposed method shows the best performance for 1–3 adaptation sentences. The table shows that the performance is improved by using the basis vectors derived from PPCA in the MAP framework, and the improvement in performance is particularly great for small amounts of adaptation data. This is due to the fact that the obtained weight is constrained by the prior distribution in the MAP framework; the weight cannot reliably be estimated using only adaptation data when the amount of adaptation data is small. As the amount of adaptation data increases, the performance of the model based on a two-way array representation in the ML framework is comparable with that of the proposed method. For large amounts of adaptation data, the weight estimated using only the adaptation data is reliable, and the ML-based adaptation method approaches the performance of the MAP counterpart.

6. Conclusions

We proposed a speaker adaptation technique using the two-way array representation of training models in the PPCA framework. Using the basis vectors derived from PPCA of the training models, we formulated speaker adaptation in the MAP framework. The proposed method performed better than the ML counterpart and the eigenvoice method did, especially for small amounts of adaptation data.

References