Sequential Model Adaptation for Speaker Verification

Jun Wang[1], *Dong Wang*[1], *Xiaojun Wu*[1], *Thomas Fang Zheng*[1], *Javier Tejedor*[2]

1. Center for Speech and Language Technologies, Division of Technical Innovation and Development,
Tsinghua National Laboratory for Information Science and Technology
Center for Speech and Language Technologies, Research Institute of Information Technology
Department of Computer Science and Technology, Tsinghua University, Beijing, China
2. Human Computer Technology Laboratory, Universidad Autónoma de Madrid, Spain

wangjun@cslt.riit.tsinghua.edu.cn, wangdong99@mails.tsinghua.edu.cn
xjwu@tsinghua.edu.cn, fzheng@tsinghua.edu.cn, javier.tejedor@uam.es

Abstract

GMM-UBM-based speaker verification heavily relies on well-trained UBMs. In practice, it is not often easy to obtain a UBM that fully matches the acoustic channel in operation. In a previous study, we proposed to address this problem by a novel sequential UBM adaptation approach based on MAP. This work extends the study by applying the sequential approach to speaker model adaptation. In addition, we investigate a new feature-space sequential adaptation approach based on feature MAP linear regression (fMAPLR) and compare it with the previously proposed model-space MAP approach. We find that these two approaches are complementary and can be combined to deliver additional performance gains. The experiments conducted on a time-varying speech database demonstrate that the proposed MAP-fMAPLR approach leads to significant EER reduction with two mismatched UBMs (25% and 39% respectively).

Index Terms: MAP, fMAPLR, sequential adaptation, speaker verification

1. Introduction

The GMM-UBM framework is widely used in speaker verification [12]. This approach involves a well-trained universal background model (UBM) to represent general speakers, and each enrolled speaker is represented by a Gaussian mixture model (GMM) which is adapted from the UBM via maximum *a posteriori* (MAP) estimation [4].

A basic assumption of the GMM-UBM approach is that the UBM is able to represent all acoustic and phonetic variations in the speech data, so that the deviation of a speaker GMM from the UBM reflects and only reflects the speaker characteristics. On the one hand, this requires a large amount of data in UBM training, and on the other hand, the acoustic channels of the training data should be consistent with the operation environment. In practice, however, it is often difficult, if not impossible, to collect sufficient channel-matched data to train a fully consistent UBM. Furthermore, most of operation channels in practice are time-variant, which fails a pre-trained UBM anyway.

A multitude of researches have been conducted to address the channel mismatch or session variation problem within the GMM-UBM framework. These researches can be categorized into three directions: feature transformation [11, 19, 16], model compensation [5, 15] and score normalization [12, 1]. A comprehensive statistical approach was proposed in [6], where the authors modeled speaker and channel variation as independent variables spanning in low-rank subspace, and then inferred channel factors by factor analysis. [13] followed this line, but allowed only low-rank channel factors, leaving the speaker factors full-rank. This method is augmented in [10] where the authors presented a straightforward interpretation for the subspace method together with a simple implementation. In [2], various feature and model compensation approaches were investigated, and the conclusion was that adaptation based on low-rank channel subspace (eigen-channels) is highly effective to deal with channel mismatch.

Besides GMM-UBM, channel mismatch was also studied within other verification frameworks. For example, [7] proposed to reduce channel impact for verification systems based on neural networks by eliminating a proportion of hidden nodes; [13] presented some feature mapping functions to mitigate channel discrepancy for verification systems based on support vector machines (SVM).

All of the above researches require some training data to learn certain compensation structures (transforms or eigen-subspace). In situations where the working channel is totally new, or the channel is time-variant, it is usually difficult to collect such data, which in turn fails most of the existing methods.

In a previous study, we proposed a novel MAP-based sequential UBM adaptation approach [17]. We have shown that by adapting the UBM sequentially with the data of every new enrollment, the system will be gradually adapted to the new channel. The present paper follows this direction and extends the study by applying the sequential approach onto both the UBM and speaker models, which means that whenever a new speaker is enrolled, both the UBM and the early enrolled speaker models are adapted immediately. This leads to a full sequential adaptation that allows all the models to be adapted in an online manner.

The second contribution of this paper is that we study a new sequential adaptation approach based on feature maximum likelihood linear regression (fMLLR) [3]. fMLLR is a feature space adaptation approach and has been widely applied to compensate for channel mismatch. The disadvantage of fMLLR in our sequential adaptation framework is that the maximum likelihood estimation leads to highly aggressive adaption, which causes over-fitting to speakers instead of channels for the first few enrollments. We therefore consider the MAP-based fMLLR, or

This work was supported by the National Natural Science Foundation of China under Grant No. 61271389 and the National Basic Research Program (973 Program) of China under Grant No. 2013CB329302.

fMAPLR [9]. We compare the MAP-based and the fMAPLR-based sequential adaptation, and combine them to attain further improvement.

The rest of the paper is organized as follows: Section 2 and Section 3 present the MAP-based and the fMAPLR-based sequential adaptation respectively, followed by the combination approach in Section 4. The experiments are reported in Section 5, and the entire paper is concluded in Section 6 with some ideas for the future work.

2. MAP-based sequential adaptation

2.1. UBM sequential adaptation

MAP is a well established framework for model adaptation. For a GMM, assume a fixed (but need to estimate) diagonal covariance matrix $\mathbf{\Sigma}_c = diag(\boldsymbol{\sigma}_c)$ for the c-th component. Define a Gaussian prior on the mean vector $\boldsymbol{\mu}_c$:

$$p(\boldsymbol{\mu}_c) = \mathcal{N}(\boldsymbol{\mu}_c; \hat{\boldsymbol{\mu}}_c, diag(\hat{\boldsymbol{\sigma}}))$$

where $\hat{\boldsymbol{\mu}}_c$ is the mean vector. $\hat{\boldsymbol{\sigma}}$ is assumed to be shared among all the components. The MAP estimation for $\boldsymbol{\mu}_c$ with a set of training data $\{\boldsymbol{x}_i\}$ is given by:

$$\boldsymbol{\mu}_c = \frac{\boldsymbol{z}_c + \frac{\boldsymbol{\sigma}_c}{\hat{\boldsymbol{\sigma}}}\hat{\boldsymbol{\mu}}_c}{\boldsymbol{r}_c + \frac{\boldsymbol{\sigma}_c}{\hat{\boldsymbol{\sigma}}}}$$

where the division is element-wised, and $\boldsymbol{r}_c$ and $\boldsymbol{z}_c$ are two accumulative statistics defined as:

$$\boldsymbol{r}_c = \sum_i r_{c,i}[\mathbf{1}]$$

$$\boldsymbol{z}_c = \sum_i r_{c,i}\boldsymbol{x}_i$$

where $[\mathbf{1}]$ is a vector whose elements are all equal to 1, and $r_{c,i}$ is the effective occurrence of x_i on the c-th component, given by:

$$r_{c,i} = \frac{\boldsymbol{w}_c\mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_c, diag(\boldsymbol{\sigma}_c))}{\sum_m \boldsymbol{w}_m\mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_m, diag(\boldsymbol{\sigma}_m))} \quad (1)$$

In the conventional UBM-GMM framework, the above MAP approach is used to derive speaker models (GMMs) from the UBM. In the sequential UBM framework, it is used to adapt both the UBM and the speaker models. The basic process is simple: for the k-th enrollment, the statistics $\boldsymbol{r}_c^k$ and $\boldsymbol{z}_c^k$ are computed based on the current sequential UBM (denoted by UBM_{k-1}). These statistics are firstly pooled together with the statistics of the previously enrolled speakers to form the pooled statistics:

$$\bar{\boldsymbol{r}}_c^k = \sum_{i=1}^{i=k} \boldsymbol{r}_c^i$$

$$\bar{\boldsymbol{z}}_c^k = \sum_{i=1}^{i=k} \boldsymbol{z}_c^i.$$

$\bar{\boldsymbol{r}}_c$ and $\bar{\boldsymbol{z}}_c$ are then used to derive the k-th sequential UBM (UBM_k) from the original UBM (UBM_0) according to the following equation:

$$\boldsymbol{\mu}_c^{UBM_k} = \frac{\bar{\boldsymbol{z}}_c^k + \frac{\boldsymbol{\sigma}_c^{UBM_0}}{\hat{\boldsymbol{\sigma}}_{UBM}}\boldsymbol{\mu}_c^{UBM_0}}{\bar{\boldsymbol{r}}_c^k + \frac{\boldsymbol{\sigma}_c^{UBM_0}}{\hat{\boldsymbol{\sigma}}_{UBM}}} \quad (2)$$

where we denote the covariance of the prior by $\hat{\boldsymbol{\sigma}}_{UBM}$ to indicate that the prior is used for UBM update. Once UBM_k is derived, the speaker model GMM_k can be obtained by an additional MAP step over UBM_k:

$$\boldsymbol{\mu}_c^k = \frac{\boldsymbol{z}_c^k + \frac{\boldsymbol{\sigma}_c^{UBM_k}}{\hat{\boldsymbol{\sigma}}_{SPK}}\boldsymbol{\mu}_c^{UBM_k}}{\boldsymbol{r}_c^k + \frac{\boldsymbol{\sigma}_c^{UBM_k}}{\hat{\boldsymbol{\sigma}}_{SPK}}}$$

where $\hat{\boldsymbol{\sigma}}_{SPK}$ indicates that the prior is for speaker model adaptation. Finally, the pair (UBM_k, GMM_k) is preserved to use in verification. Note that in all the adaptation (either for UB-M or GMM), the covariance matrices remain fixed as adopted by most UBM-GMM systems. In addition, to guarantee learning channel characteristics rather than speaker particularities, we set a much stronger prior in the UBM MAP than in the speaker MAP ($\hat{\sigma}_{UBM} << \hat{\sigma}_{SPK}$). Our previous experiments show that this sequential UBM approach provides a substantial equal error rate (EER) reduction [17].

2.2. UBM-GMM sequential adaptation

A shortage of the UBM sequential adaptation is that each speaker reserves its own UBM-GMM pair and keeps it unchanged. This means that the early enrolled speakers cannot use the later updated UBM. We therefore extend the approach by re-adapting all the speaker GMMs whenever the UBM updated. This re-adapting, however, requires dumping the enrollment speech, which is storage expensive and usually not allowed. An approximated solution is to reserve the statistics $\boldsymbol{z}_c^k$ and $\boldsymbol{r}_c^k$ and use them to perform GMM update later on. This leads to the UBM-GMM sequential adaptation approach. Specifically, after the k-th enrollment, the original UBM is updated to UBM_k, and then for every early enrolled speaker m, its GMM is adapted as follows:

$$\boldsymbol{\mu}_c^m = \frac{\boldsymbol{z}_c^m + \frac{\boldsymbol{\sigma}_c^{UBM_k}}{\hat{\boldsymbol{\sigma}}_{SPK}}\boldsymbol{\mu}_c^{UBM_k}}{\boldsymbol{r}_c^m + \frac{\boldsymbol{\sigma}_c^{UBM_k}}{\hat{\boldsymbol{\sigma}}_{SPK}}}. \quad (3)$$

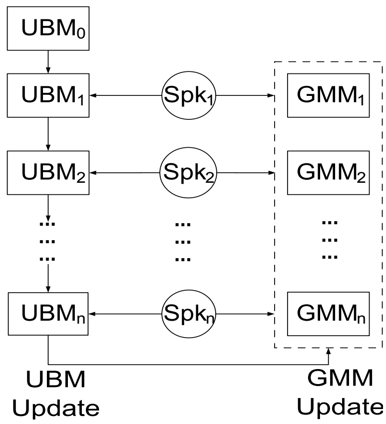

Figure 1: UBM-GMM sequential adaptation.

Figure 1 illustrates the sequential UBM-GMM process. For each enrollment k, the speaker statistics $\mathbf{r}_c^k$ and $\mathbf{z}_c^k$ are computed based on the current UBM and are accumulated with $\bar{\mathbf{r}}_c^k$ and $\bar{\mathbf{z}}_c^k$. The UBM is updated to UBM_k with $\bar{\mathbf{r}}_c^k$ and $\bar{\mathbf{z}}_c^k$ according to (2), and all the speaker models (including speaker k) are updated based on UBM_k with $\mathbf{r}_c^k$ and $\mathbf{z}_c^k$ according to (3).

3. fMAPLR-based sequential adaptation

MLLR was first proposed by the Cambridge group to deal with channel mismatch and session variability [8, 14]. Its constrained variant, feature MLLR (fMLLR), has been developed to learn transforms on feature vectors [3]. A major advantage of fMLLR is that the covariance matrices are implicitly adapted besides the mean vectors without increasing the number of training parameters, which often leads to additional gains.

Define a transformation matrix $\boldsymbol{W} = [\boldsymbol{b} \quad \boldsymbol{A}]$ that projects an input speech signal $\boldsymbol{x}_i$ as follows:

$$\hat{\boldsymbol{x}}_i = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b} = \boldsymbol{W}\boldsymbol{\xi}_i$$

where $\boldsymbol{A}$ is a rotation matrix and $\boldsymbol{b}$ is a bias term. $\boldsymbol{\xi}_i = [1 \quad \boldsymbol{x_i}^T]^T$ is the extended observation vector. Assuming that $\hat{\boldsymbol{x}}$ is modeled by a GMM given by:

$$\sum_c \gamma_c \mathcal{N}(\hat{\boldsymbol{x}}; \boldsymbol{\mu_c}, diag(\boldsymbol{\sigma_c}))$$

where γ_c is the weight of the c-th component, the optimal W can be attained by maximizing the likelihood function

$$\mathcal{L}(\boldsymbol{W}) = \sum_i log \sum_c \gamma_c \mathcal{N}(\boldsymbol{W}\boldsymbol{\xi}_i; \boldsymbol{\mu}_c, diag(\boldsymbol{\sigma}_c))$$

with respect to $\boldsymbol{W}$. This leads to the following iterative solution:

$$\boldsymbol{W}_l^T = \boldsymbol{G}^{(l)-1}(\alpha \boldsymbol{p}_l + \boldsymbol{k}^{(l)}) \quad l = 1, 2, ..., L \tag{4}$$

where $\boldsymbol{W}_l$ is the l-th column of $\boldsymbol{W}$, and $\boldsymbol{p}_l$ is the extended cofactor vector $[0 \; cof(\boldsymbol{A}_{l,1}) \; ... \; cof(\boldsymbol{A}_{l,L})]^T$. $\boldsymbol{G}^{(l)}$ and $\boldsymbol{k}^{(l)}$ are the accumulative statistics, defined by:

$$\boldsymbol{G}^{(l)} = \sum_i \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \sum_c \frac{r_{c,i}}{\sigma_{c,l}} \tag{5}$$

$$\boldsymbol{k}^{(l)} = \sum_i \boldsymbol{\xi}_i \sum_c \frac{r_{c,i}\mu_{c,l}}{\sigma_{c,l}} \tag{6}$$

where c indexes the Gaussian components, and $r_{c,i}$ is the effective occurrence defined in (1). $\mu_{c,l}$ and $\sigma_{c,l}$ are the l-th dimension of the mean and diagonal variance vectors of the c-th Gaussian component, respectively. The factor α is solved from the following equation and the root that maximizes the likelihood function is selected:

$$\alpha^2 \boldsymbol{p}_l^T \boldsymbol{G}^{(l)-1} \boldsymbol{p}_l = \alpha \boldsymbol{p}_l^T \boldsymbol{G}^{(l)-1} \boldsymbol{k}^{(l)} - \beta = 0$$

where

$$\beta = \sum_{c,i} r_{c,i}. \tag{7}$$

It is clear that the fMLLR adaptation is exclusively determined by the accumulative statistics $\boldsymbol{G} = [\boldsymbol{G}^{(1)}, ..., \boldsymbol{G}^{(L)}]$, $\boldsymbol{k} = [\boldsymbol{k}^{(l)}, ..., \boldsymbol{k}^{(L)}]$ and β, similar as in MAP where $\boldsymbol{r}_c$ and $\boldsymbol{z}_c$ determine the adaptation. Therefore, the sequential adaptation approach originally designed based on MAP can be migrated to work with fMLLR. Specifically, for each new enrollment m, the statistics, denoted by $\boldsymbol{G}^m$, $\boldsymbol{k}^m$ and β^m, are computed according to (5)-(7), and then are accumulated with the statistics of the early enrollments, simply by:

$$\bar{\boldsymbol{G}}^m = \sum_{i=1}^{i=m} \boldsymbol{G}^i$$

$$\bar{\boldsymbol{k}}^m = \sum_{i=1}^{i=m} \boldsymbol{k}^i$$

$$\bar{\beta}^m = \sum_{i=1}^{i=m} \beta^i.$$

Finally, the accumulated $\bar{\boldsymbol{G}}^m$, $\bar{\boldsymbol{k}}^m$ and $\bar{\beta}^m$ are used to estimate the transformation matrix $\boldsymbol{W}$ according to (4), denoted by $\boldsymbol{W}^m$. Once $\boldsymbol{W}^m$ is obtained, it is used to transform the input features for speaker m and derive the speaker model GMM$_m$ from the original UBM. In verification for speaker k, either the speaker-dependent $\boldsymbol{W}^k$ or the latest $\boldsymbol{W}^m$ is used to transfer the speech signal. The former case corresponds to the UBM sequential adaptation in the MAP approach, and the later corresponds to the UBM-GMM sequential adaptation.

A major problem of the fMLLR approach is that the adaptation is based on maximizing the likelihood of the enrollment data, which tends to result in aggressive adaptation to the enrollment data. For the first few enrollments, this may cause serious over-fitting and it is hard to tell whether the adaptation learns the channel characteristics or the speaker properties. We therefore consider to place a prior on the transformation matrix, which leads to the MAP-based fMLLR, or fMAPLR [9]. For simplicity, we choose the following prior in our study:

$$p(\boldsymbol{W}_l) \propto exp[r||\boldsymbol{W}_l - \boldsymbol{I}_l||_2^2]$$

where r is a scale factor and $\boldsymbol{I}_l$ is the l-th column of the identity matrix $\boldsymbol{I}$, and $||\cdot||_2$ is the ℓ_2 norm of a vector. This is a Gaussian prior with $\boldsymbol{I}_l$ as the mean and $\frac{1}{2r}\boldsymbol{I}$ as the covariance matrix. It can be shown that the maximum *a posteriori* solution takes the same form of the fMLLR solution (4)-(6), except a slight modification on the accumulative statistics:

$$\boldsymbol{G'}^{(l)} = \boldsymbol{G}^{(l)} + r\boldsymbol{I}$$

$$\boldsymbol{k'}^{(l)} = \boldsymbol{k}^{(l)} + r\boldsymbol{I}_l.$$

4. MAP-fMAPLR sequential adaptation

When comparing the MAP-based and fMAPLR-based sequential adaptation, we notice that the MAP approach is 'local', which means that each Gaussian component reserves its specific accumulative statistics and is adapted individually. The fMAPLR approach, in contrast, is 'global', which means that all the Gaussian components share the same transform (we do not consider component-based fMAPLR in this study). Another difference is that the MAP approach adapts the mean vectors only (at least in most UBM-GMM systems), while the fMAPLR adapts both the means and the covariances. These differences lead to unique behavior with each approach, and we assume they are complementary and thus can be combined. Specifically, the fMAPLR is used to incrementally update the transform which is then applied to transform the training/verification speech. Based on the transformed speech, the MAP-based sequential UBM-GMM adaptation is conducted. By this combination, we can utilize the advantage of fMAPLR in terms of its parsimonious parameters and the advantage of MAP in terms of its detailed adaptation. The experimental results demonstrate that this combination leads to considerable performance gains. The MAP-fMAPLR sequential adaptation is shown in Figure 2.

5. Experiments

5.1. Database and configurations

We conduct the experiments on a time-varying database [18] which involves 60 speakers (30 males and 30 females) record-

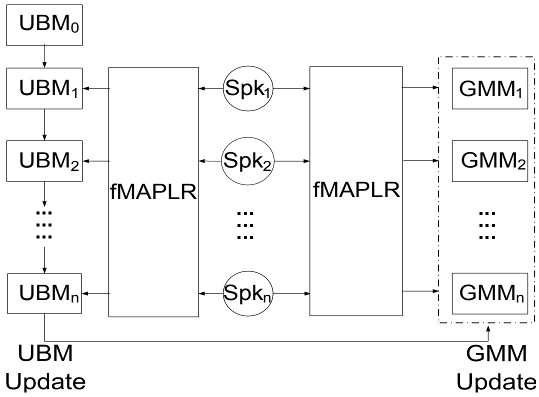

Figure 2: MAP-fMAPLR sequential adaptation.

ed with a desktop microphones from 2010 to 2012. The sampling rate of the signals is 8 kHz and the sample size is 16 bits. For each speaker, 100 Chinese utterances were recorded. The enrollment utterances are about 20 seconds in length, and the verification utterances are of 5-10 seconds. The 16-dimensional Mel frequency cepstral coefficients (MFCC) plus their first order derivatives are used as the acoustic features. Both the UBM and the speaker models involve 1024 Gaussian components.

In order to test the capability of the sequential approach in learning new channels, we conduct the experiments with two initial UBMs that were trained with databases that are acoustically inconsistent with the enrollment/verification speech: UBM$_a$ which was trained on 3 hours of desktop microphone speech data (45 males and 38 females), and UBM$_b$ which was trained on 6 hours of telephone speech data (150 males and 150 females). Clearly, UBM$_a$ is just slightly mismatched with the operation condition, while UBM$_b$ is more mismatched.

5.2. Experimental results

We treat the static UBM approach, i.e., without any adaptation once the UBM is delivered, as the baseline. We test the following sequential adaptation approaches:

- **SUBM:** The MAP-based UBM sequential adaptation, reported in [17].
- **SUBM-GMM:** The MAP-based UBM-GMM sequential adaptation, as presented in Section 2.
- **fMAPLR-SPK:** The fMAPLR-based sequential adaptation. Each speaker uses its own fMLLR matrix. This corresponds to the SUBM approach in the MAP style.
- **fMAPLR-LATEST:** The fMAPLR-based sequential adaptation. All speakers use the latest fMLLR matrix. This corresponds to the SUBM-GMM approach in the MAP style.
- **SUBM-GMM + fMAPLR-SPK:** The fMAPLR is used to transform speech signals, based on which SUBM-GMM is conducted. Each speaker uses its own fMLLR matrix.
- **SUBM-GMM + fMAPLR-LATEST:** The fMAPLR is used to transform speech signals, based on which SUBM-GMM is conducted. All speakers use the latest fMLLR matrix.

In our experiment, the elements of the prior covariance vector $\hat{\boldsymbol{\sigma}}_{SPK}$ for speaker model adaptation were chosen to be the same value 0.5. In SUBM and SUBM-GMM, the elements of the prior covariance vector $\hat{\boldsymbol{\sigma}}_{UBM}$ for the UBM adaptation were chosen to be the same value 0.003. In fMAPLR-SPK and fMAPLR-LATEST, the prior parameter r was chosen to be 10^5. These values were selected to optimize the performance. We see that the prior for the UBM adaptation, either with MAP or fMAPLR, is much stronger than the prior for the speaker model adaptation, which yields a slow but stable adaptation to the channel character.

	EER%
Baseline	11.77
SUBM	11.05
SUBM-GMM	9.24
fMAPLR-SPK	11.77
fMAPLR-LATEST	11.14
SUBM-GMM + fMAPLR-SPK	9.38
SUBM-GMM + fMAPLR-LATEST	**8.82**

Table 1: EER results with UBM$_a$ as the initial.

	EER%
baseline	10.44
SUBM	8.79
SUBM-GMM	6.93
fMAPLR-SPK	8.95
fMAPLR-LATEST	8.70
SUBM-GMM + fMAPLR-SPK	6.69
SUBM-GMM + fMAPLR-LATEST	**6.41**

Table 2: EER results with UBM$_b$ as the initial.

The verification performance is evaluated in terms of the EER. The results with UBM$_a$ and UBM$_b$ as the initial models are presented in Table 1 and Table 2 respectively. We first observe that for both UBM$_a$ and UBM$_b$, the MAP-based sequential adaptation approaches, both SUBM and SUBM-GMM, lead to significant EER reduction, particularly with the SUBM-GMM approach. The fMAPLR approach is less effective, particularly for UBM$_a$ the improvement is marginal. This might be attributed to the fact that UBM$_a$ was trained on data recorded with desktop microphones and therefore is less mismatched with the operation environment than UBM$_b$. This less mismatch leads to less effectiveness of the sequential adaptation technique, particularly with fMAPLR which is global and thus limited in thorough adaptation. Nevertheless, combining the fMAPLR with the MAP leads to the best performance, confirming our conjecture that these two approaches are complementary.

6. Conclusions

This paper extended our previous study on sequential UBM adaptation. We proposed to apply the sequential approach to adapt speaker models, and comparatively studied two adaptation methods based on MAP and fMAPLR respectively. We find that the MAP-based approach is more effective than the fMAPLR-based approach, and they are complementary so can be combined. This study leads to a relative EER reduction of 25% for a slightly mismatched UBM and 39% for a more mismatched UBM. Further work involves study of sequential adaptation based on component-dependent fMAPLR.

7. References

[1] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 42–54, 2000.

[2] L. Burget, P. Matejka, O. Glembek, and P. Schwarz, "Analysis of feature extraction and channel compensation in GMM speaker recognition system," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 7, pp. 1979–1986, 2007.

[3] M. J. F. Gales and P. C. Woodland, "Mean and variance adaptation within the MLLR framework," *Computer Speech and Language*, vol. 10, no. 4, pp. 249–264, 1996.

[4] J. L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. on Speech and Audio Processing*, vol. 2, no. 2, pp. 291–298, 1994.

[5] L. Heck and M. Weintraub, "Handset-dependent background models for robust text-independent speaker recognition," in *Proc. ICASSP'97*, vol. 2, 1997, pp. 1071–1074.

[6] P. Kenny and P. Dumouchel, "Disentangling speaker and channel effects in speaker verification," in *Proc. ICASSP'04*, vol. 1, 2004, pp. 37–40.

[7] S. P. Kishore and B. Yegnanarayana, "Speaker verification: minimizing the channel effects using autoassociative neural network models," in *Proc. ICASSP'00*, vol. 2, 2000, pp. 1101–1104.

[8] 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, no. 2, pp. 171–185, 1995.

[9] X. Lei, J. Hamaker, and X. He, "Robust feature space adaptation for telephony speech recognition," in *Proc. Interspeech'06*, 2006, pp. 773–776.

[10] D. Matrouf, N. Scheffer, B. Fauve, and J.-F. Bonastre, "A straightforward and efficient implementation of the factor analysis model for speaker verification," in *Proc. Interspeech'07*, 2007, pp. 1242–1245.

[11] D. A. Reynolds, "Channel robust speaker verification via feature mapping," in *Proc. ICASSP'03*, vol. 2, 2003, pp. 53–56.

[12] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[13] A. Solomonoff, C. Quillen, and W. M. Campbell, "Channel compensation for SVM speaker recognition," in *Proc. Odyssey Speaker and Language Recognition Workshop*, 2004, pp. 57–62.

[14] A. Stolcke, S. S. Kajarekar, L. Ferrer, and E. Shrinberg, "Speaker recognition with session variability normalization based on MLLR adaptation transforms," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 7, pp. 1987–1998, 2007.

[15] R. Teunen, B. Shahshahani, and L. Heck, "A model-based transformational approach to robust speaker recognition," in *Proc. ICSLP'00*, 2000, pp. 495–498.

[16] C. Vair, D. Colibro, and P. Laface, "Channel factors compensation in model and feature domain for speaker recognition," in *Proc. Odyssey Speaker and Language Recognition Workshop*, 2006, pp. 1–6.

[17] J. Wang, D. Wang, X.-J. Wu, and T. F. Zheng, "Sequential UBM adaptation for speaker verification," in *Submitted to ChinaSIP'13*, 2013.

[18] L.-L. Wang and T. F. Zheng, "Creation of time-varying voiceprint database," in *Proc. Oriental-COCOSDA*, 2010.

[19] D. Zhu, B. Ma, H. Li, and Q. Huo, "A generalized feature transformation approach for channel robust speaker verification," in *Proc. ICASSP'07*, vol. 4, 2007, pp. 61–64.