World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:3, No:11, 2009

Improvement of MLLR Speaker Adaptation Using a Novel Method

Ing-Jr Ding

***Abstract*—**This paper presents a technical speaker adaptation method called WMLLR, which is based on maximum likelihood linear regression (MLLR). In MLLR, a linear regression-based transform which adapted the HMM mean vectors was calculated to maximize the likelihood of adaptation data. In this paper, the prior knowledge of the initial model is adequately incorporated into the adaptation. A series of speaker adaptation experiments are carried out at a 30 famous city names database to investigate the efficiency of the proposed method. Experimental results show that the WMLLR method outperforms the conventional MLLR method, especially when only few utterances from a new speaker are available for adaptation.

***Keywords*—**hidden Markov model, maximum likelihood linear regression, speech recognition, speaker adaptation.

I. Introduction

SPEAKER adaptation techniques have been applied to speech recognition technology to get a good recognition performance over the last decade. A speaker-independent (SI) system is typically constructed using speech samples collected an as large as possible population of speakers. Nevertheless, in the speaker-dependent (SD) case, the large amount of required training data for each test speaker reduces the utility and portability of such system. For a given speech recognition task, a speaker-dependent system usually outperforms a speaker independent system by a factor of two to three as long as a sufficient amount of training data is available to acquire the speaker- dependent model. But when the amount of speaker-dependent data is limited, such a performance improvement may not be realized. Hence a speaker adaptive (SA) system is constructed to have desirable SD-like properties but require only a small fraction of the speaker-specific training data needed to build a full SD system. Speaker adaptation techniques that transform the SI acoustic models to obtain near speaker-dependent performance are sometimes named model adaptation techniques.

Model adaptation techniques mainly include three categories: Bayesian-based, transformation-based and eigenvoice-based. Bayesian-based model adaptation attempts to directly re-estimate the model parameters, for example, using maximum a posteriori (MAP) adaptation [1][2]. MAP adaptation only re-estimates model parameters of the corresponding units appearing in the adaptation data. Usually, MAP adaptation needs a large amount of data for adaptation and the performance improves as the amount of adaptation data increases. When the amount of data is sufficiently large, the MAP estimation yields recognition performances as good as that obtained using maximum-likelihood estimation [3]. As for transformation-based model adaptation, such the adaptation technique consists in applying some transformations estimated from a set of adaptation utterances to various clusters of hidden Markov models (HMM) parameters. Two kinds of transformation-based adaptations, the bias transformation and the affine transformation, are popular in the recent years. The bias transformation is usually applied by adding a cepstral bias to adapting the HMM parameters [4]. Sometimes, by adding a bias may not be sufficient for modeling the variations of test environments or test speakers. Thus, the affine transformation is suggested. Accordingly, the HMM parameters are linearly scaled by a proper transformation matrix and shifted by a bias. In [5], a maximum likelihood linear regression (MLLR) method was proposed for adapting the continuous-density hidden Markov model (CDHMM) parameters. In this study, the maximum likelihood (ML) theory for calculating the linear regression transform was employed to adapt the HMM mean vectors.

Besides, various methods for ensuring robust MLLR transformation estimation have been proposed. In [6][7], the maximum a posteriori (MAP) theory for estimating the transformation parameters was presented by maximizing the posterior density. In both [8] and [9], it is suggested that a prior distribution for the mean transformation matrix parameters be used (dubbing the technique MAPLR) and this improves performance when very small amounts of data are available. Alternatively a variant of the E-M algorithm that optimizes a discounted likelihood criterion was suggested in [10]. This technique also improved robustness for transformations being trained using small amounts of adaptation data.

Kuhn, et al. proposed the eigenvoice adaptation where a priori knowledge concerning the variations among all training speakers were represented as the set of SD model parameters in the form of eigenvectors named eigenvoices; a new speaker model was then expressed as the linear combination of the set of eigenvoices [11]. Following that, various extensions of such the eigenvoice-based adaptation have been developed recently, such as the eigenvoice-versions of conventional MLLR and MAPLR adaptation [12][13].

I. -J. Ding is with the Graduate Institute of Automation and Control, National Taiwan University of Science and Technology, Taipei 106, Taiwan, R.O.C (corresponding author to provide phone: 886-2-27990171; fax: 886-2-27990171; e-mail: ingjr.csie90g@nctu.edu.tw).

To tackle the issue of unreliable MLLR model transformation due to the scantiness of training data without the daunting cost of MAPLR-like adaptation, an alternative MLLR-based speaker adaptation algorithm is presented in this paper in order to keep a good recognition performance when only a very limited amount of adaptation data is available. An adequate weight will be calculated for the transformation matrices so that the updated mean parameters for the initial model will not vary drastically, especially when only extremely a few adaptation data from a new speaker are available. The prior knowledge of the initial model will be adequately incorporated into the adaptation.

The rest of the paper is organized as follows. The theoretical formulation of the MLLR is briefly described in Section II. Besides, using the MAP theory for estimating the transformation parameters is then simply described after the MLLR. In Section III, the proposed method to adjust the parameters of the MLLR estimation is presented. Experiments on speaker adaptation by applying the proposed method to 30 famous city names database are carried out in Section IV. Finally, the conclusion is made in Section V.

II. Review of the Linear Regression Adaptation Technique

The maximum likelihood linear regression (MLLR) is a popular speaker adaptation scheme that uses model transforms to construct a more appropriate model [5]. Let Λ be a set of SI hidden Markov models. A transformation-based model adaptation consists of applying some transformations F to various clusters of HMM parameters. Given some adaptation data, Y, the objective of the adaptation is first to derive the parameters η of the transformations, and then use the transformed models $F_\eta(\Lambda)$ to recognize the incoming speech. The estimation of η is traditionally carried out using classical statistics that assume that η are some fixed but unknown parameters. Because of its simplicity, the maximum likelihood (ML) criterion is usually chosen, which states that $\hat{\eta}_{ML}$ should maximize the likelihood of the adaptation data given the transformed model, $p(Y \mid \Lambda, \eta)$:

$$\hat{\eta}_{ML} = \arg\max_{\eta} p(Y \mid \Lambda, \eta) . \tag{1}$$

In the MLLR technique, the ML criterion is used to estimate the transformation parameters. Meanwhile, a simple and linear transformation is utilized. The parameters η of the transformations are (A, b). The Gaussian mean parameters are updated according to

$$\hat{\mu} = A \cdot \mu + b , \tag{2}$$

where A is an $n \times n$ matrix and b is an n dimensional vector (and n is dimensionality of the observations). This equation is sometimes written as

$$\hat{\mu}_s = W_s \cdot \xi_s , \tag{3}$$

where W_s is the $n \times (n+1)$ transformation matrix (for n dimensional data) and ξ_s is the extended mean vector, which is defined as

$$\xi_s = [\omega, \mu_{s_1}, \ldots, \mu_{s_n}]' , \tag{4}$$

where ω is the offset term for the regression and usually set as 1.

The transformation matrix W_s is estimated such that the likelihood of the adaptation data is maximized. There is a closed form solution to the W_s estimation [5]. As usual, the Expectation-Maximization (E-M) algorithm is used to solve the W_s matrix estimation problem [14]. Accordingly, the W_s is derived by solving the following equation:

$$\begin{aligned} &\sum_{p=1}^{P} \sum_{t=1}^{T_p} \sum_{r=1}^{R} \gamma_{s_r}^{(p)}(t) \Sigma_{s_r}^{-1} o_t^{(p)} \xi_{s_r}' \\ &= \sum_{p=1}^{P} \sum_{t=1}^{T_p} \sum_{r=1}^{R} \gamma_{s_r}^{(p)}(t) \Sigma_{s_r}^{-1} W_s \xi_{s_r} \xi_{s_r}', \end{aligned} \tag{5}$$

where P is the number of adaptation data, $o_t^{(p)}$ is the observation vectors of adaptation data at time t, and $\gamma_{s_r}^{(p)}(t)$ is the total occupation probability for the r-th mixture of state s at time t given that the observation sequence $o^{(p)}$ is generated.

Usually there are many Gaussians per matrix i.e., the transformation matrix is tied over a number of Gaussians. This transform sharing can allow all the Gaussians in a system to be updated. That is, all the Gaussian mean vectors can be adapted using Eq. (3) with the derived W_s from Eq. (5).

In general, the main difference between speakers is assumed to be characterized by the mean vectors. Therefore, usually only Gaussian mean vectors are adapted and the other HMM parameters such as Gaussian variance parameters are not adapted. That the Gaussian variances can also be updated is proposed in [15]. In this study, the variance transforms can be found after the mean transforms have been estimated. However, the improvement of the recognition performance is extremely limited compared with mean-only adaptation MLLR.

On the other hand, Chesta et al. suggests that the prior density can be taken into consideration in the estimation process of transformation parameters by using a maximum a *posteriori* criterion [8]:

$$\begin{aligned} \hat{\eta}_{MAP} &= \arg\max_{\eta} p(\eta \mid Y, \Lambda) \\ &\propto \arg\max_{\eta} p(Y \mid \eta, \Lambda) p(\eta). \end{aligned} \tag{6}$$

According to this estimation criterion, the adaptation scenario corresponds to the maximum a posterior linear regression, or MAPLR technique. The maximization of Eq. (6) can be carried out by means of the E-M algorithm [14]. After a series of derivations, the following system of $p \times (p+1)$ linear equation is obtained [8], where w_{ij} is the (i, j)-th component of the matrix W, γ_{ij}, m_{ij}, σ_{ij} and ϕ_{ij} are the (i, j)-th component of the matrices $R_{n,m}, M, \Sigma$ and Φ, and where $\tilde{\mu}_i$ is the i-th component of $\mu_{n,m}$:

$$\sum_{k=1}^{p}\sum_{l=1}^{p+1}\omega_{kl}\left[\sum_{n=1}^{N}\sum_{m=1}^{M}\left(\sum_{t=1}^{T}\gamma_t(n,m)\right)\gamma_{ik}\tilde{\mu}_l\tilde{\mu}_j+\frac{1}{2}\sigma_{ki}\phi_{jl}+\frac{1}{2}\sigma_{ik}\phi_{lj}\right] = z_{ij}, \quad \begin{matrix} 1\le i\le p \\ 1\le j\le p+1 \end{matrix} \tag{7}$$

where z_{ij} is defined as:

$$z_{ij}=\sum_{k=1}^{p}\sum_{l=1}^{p+1}\left[\sum_{n=1}^{N}\sum_{m=1}^{M}\left(\sum_{t=1}^{T}\gamma_t(n,m)y_k(t)\right)\gamma_{ik}\tilde{\mu}_j+\frac{1}{2}\sigma_{ki}m_{kl}\phi_{jl}+\frac{1}{2}\sigma_{ik}m_{kl}\phi_{lj}\right] \tag{8}$$

The matrix W can be obtained by solving the system of $p\times(p+1)$ linear equations described by Eq. (7) and Eq. (8). It is noted that the matrix W is much more difficult to be solved from this system of equations than from that of the standard MLLR due to the additional terms $\{M,\Sigma,\Phi\}$ related to the prior density.

III. The Proposed Fast MLLR-based Approach (WMLLR)

As described in the above section, when using the MLLR formulation, the mean vector μ is adapted using an affine transformation. However, when only a limited amount of adaptation data is available, the transformation matrix W_s derived from Eq. (5) can be not robust. Thus, all adapted mean vectors using the poor transformation through Eq. (3) would not be effective. In the following, an MLLR-based approach to ensure robust MLLR estimation is proposed. The approach is different from previous MLLR-based methods (e.g., [8][9]).

A. Model Combination

It is expected that better performance will be achieved even when only a small training data available for adaptation with using a weighted sum of the initial mean vector and the MLLR adapted mean vector. This is reasonable because the appropriately chosen prior knowledge of the initial mean vector will adjust the ineffective MLLR adapted mean vector to the more reliable one. Accordingly, the combination approach of the initial model and the MLLR adapted model is proposed as follows:

$$\tilde{\mu}_s=\alpha\cdot\mu_s+(1-\alpha)\cdot W_s\cdot\xi_s, \tag{9}$$

where μ_s is the initial mean vector, α is a suggested weight obtained in a training procedure, and $\tilde{\mu}_s$ is the modified adapted mean vector. It is worth noting that this equation is similar to the standard MLLR adaptation solution in Eq. (3) except for the additional term related to the prior knowledge of the initial model.

Fig. 1 depicts the procedure of the proposed WMLLR adaptation method. We can get a set of transformation matrices W_s from observed adaptation data by using the standard MLLR method as described in the section 2. At the same time, according to the amount of data available for adaptation, we can choose the most proper α from those estimated well in advance. As soon as the transformation matrix is calculated and the weight α is appropriately chosen, the WMLLR adapted model can be obtained. The follows describes how those weights offered for model combination are trained in an offline supervised training procedure.

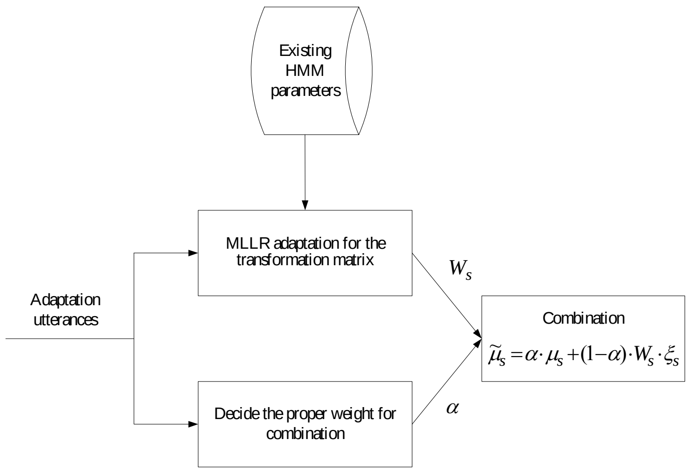

Fig. 1. Using α for model combination.

B. α Training

The training procedure of α is shown in Fig. 2. Suppose that there are total K training speakers whose baseline recognition rates are $R_b(1)$, $R_b(2)$,…, and $R_b(K)$. Then consider Eq. (9) with α replaced by $\alpha(k,j)$. As for $\alpha(k,j)$, ten intervals are equally partitioned between 0 and 1. Then for the *k*-th training speaker, the weight $\alpha(k,j)$, $j=1,2,......,9$, is sequentially assigned as 0.1, 0.2,…, and 0.9, respectively. And then, the performance, i.e., the recognition rate, of the *k*-th adapted speaker model with $\alpha(k,j)$ is $R(k,j)$. That is, for the *k*-th training speaker, the recognition set is composed of $R(k,1)$, $R(k,2)$,…, and $R(k,9)$, which are in correspondence to the adapted model with $\alpha(k,1)$, $\alpha(k,2)$,…, and $\alpha(k,9)$, respectively. If the performance $R(k,j)$ outperforms that of the *k*-th speaker's baseline model, i.e., $R_b(k)$, we take this effective weight $\alpha(k,j)$ into consideration. Conversely, if the performance of the adapted speaker model is worse than or equal to the baseline model, we discard the weight $\alpha(k,j)$, i.e., set $\alpha(k,j)=0$. The above process proceeds repeatedly until all K speakers are tested over.

When there are multiple effective weights $\alpha(k,j)$ available, it is of interest to determine to use what rule of combination to combine the useful information. A most common information combination approach suggested in [16] is a weighted linear combination. With this approach, the problem of information combination is to reduce multiple useful estimates to one optimal estimate.

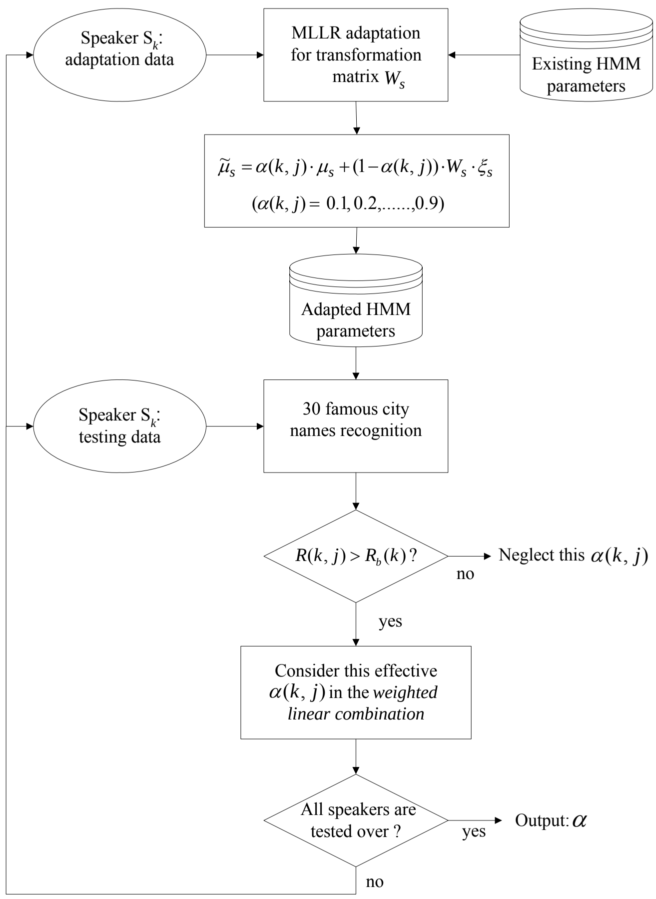

Fig. 2. The training procedure of α .

Then these various effective weights $\alpha(k, j)$ can be finally combined to a single term by taking the weighted average based on the recognition performance as follows:

$$\alpha = \sum_{k=1}^{K}\sum_{j=1}^{9} w(k,j)\cdot\alpha(k,j)\,, \qquad (10)$$

where $w(k, j)$ is defined as:

$$w(k,j) = \frac{R(k,j)}{\sum_{k=1}^{K}\sum_{j=1}^{9} R(k,j)}\,. \qquad (11)$$

The weight $w(k, j)$ assigned to each effective $\alpha(k, j)$ is proportional to the recognition performance $R(k, j)$.

C. Measurements and Analysis of α

For each specific amount of adaptation data, however, we can train a correspondent weight by the above training procedure. Fig. 3 depicts α values on this training procedure with training speakers K being set 150 under various amounts of adaptation data, starting at the size of one adaptation utterance. It is observed from Fig. 3 that the values of α decreases slightly when the number of adaptation utterances increases gradually. The weight α approaches to a saturated value when the adaptation utterances are over 4.

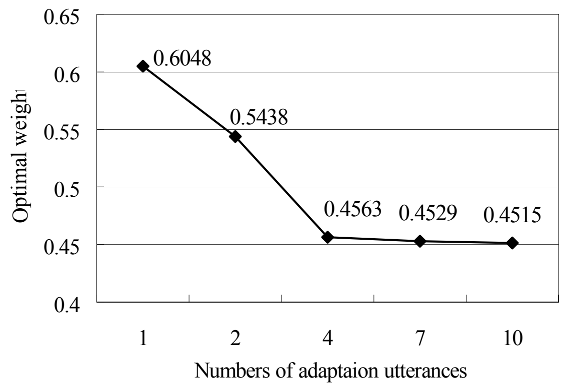

Fig. 3. The curve of the training values of α .

It is believed that the trained weight α should vary with different amounts of adaptation data. That is, the combination mean vector, i.e., the WMLLR adapted model, moves between the initial mean vector and the MLLR estimate mean vector according to the given adaptation data size. Fig. 4 shows an example of the movement of the combination mean vector. If the amount of given adaptation data is little, the combination mean vector after utilizing the above combination approach remains close to the initial mean vector. Conversely, if the amount of given adaptation data is large, the combination mean vector becomes close to the adapted mean vector by using the MLLR estimation.

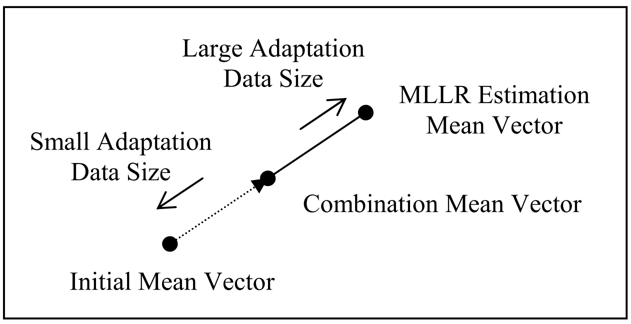

Fig. 4. An example of the movement of the combination mean vector.

D. Complexity Analysis of the Proposed Method

Moreover, compared with the MAPLR, this proposed combination approach is fast and simple. Observed from Eq. (9), the updated mean vector can be represented exactly in terms of a linear combination of the initial mean vector and the MLLR estimated mean vector. This combination action takes just $O(1)$. Furthermore, the combination factor, i.e., the weight α , is trained in an offline procedure and used for online adaptation. Thus, the task of acquiring α according to the adaptation data available takes also just $O(1)$. Concluded by the above statements, the proposed method has the same complexity as the standard MLLR formulation. Conversely, in the MAPLR, due to the additional terms related to the prior density, the required additional calculation to obtain the transformation matrix is rather expensive. This is a very adverse condition for online adaptation. Therefore, this proposed method is indeed superior to the MAPLR in the speed of calculation, which however also ensures the robustness of the MLLR estimation.

IV. Experiments and Results

A. System Description and Database

The speech signal was sampled at 8 kHz. The analysis frames were 30-ms wide with a 20-ms overlap. For each frame, a 24-dimensional feature vector was extracted. The feature vector for each frame consisted of a 12-dimensional (12-D) mel-cepstral vector and a 12-D delta-mel-cepstral vector. The initial models which were used as the speaker independent models were trained based on the hidden Markov network using a database, MAT400 sub-database DB3 [17]. The training set consisted of 4800 utterances from native Mandarin talkers. In general, each Mandarin syllable is composed of an initial part and a final part. In our experiment, the number of states was set 3 in the initial part and 6 in the final part.

In the recognition experiments, new adaptation and testing data from fifteen speakers were recorded by a close-talking microphone. The adaptation data consisted of 10 utterances from each speaker. For adaptation experiments, the number of utterances was 2, 4, 6, 8, and 10, respectively. The testing data consisted of 60 utterances from each speaker uttering twice for 30 city names. Full transformation matrices were used for MLLR, MAPLR, and the proposed WMLLR. Moreover, for MAPLR, the prior density was derived directly from the initial speaker independent models. In WMLLR, 150 speakers (not include these fifteen speakers in the adaptation testing experiment) were used to train α values.

B. Experimental Results

Table 1 and Fig. 5 show the performance and the performance curve of the proposed combination model adaptation method (WMLLR), the MAPLR method, and the conventional MLLR method in the unsupervised incremental adaptation testing experiment.

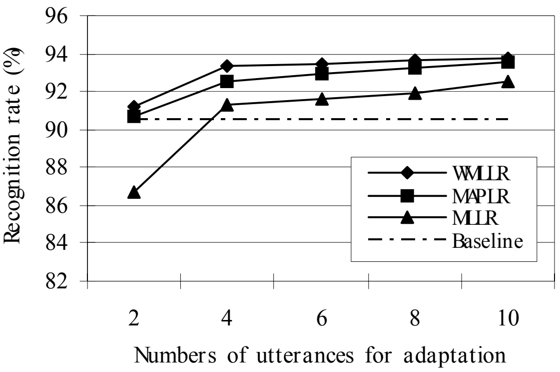

Fig. 5. The performance curves of the WMLLR method, the MAPLR method and the conventional MLLR method in the recognition testing experiments of the different amount of adaptation data.

Observed from Fig. 5, the proposed WMLLR was continuously superior to the MAPLR and the conventional MLLR. The performance curve of the WMLLR method shows the recognition rates increased rapidly till about four adaptation utterances and saturated gradually. In contrast, for the conventional MLLR method without referencing to any knowledge of the initial model, the recognition performance was worst. The recognition rate was lower than baseline in only two adaptation utterances obtained, and increased gradually with adaptation data being increased, but these were still lower than those of the WMLLR method. Besides, the results from the MAPLR method were better than the MLLR method for all the testing condition, but they were still a little worse than the WMLLR method. It is observed that the WMLLR is more robust than the MAPLR and the conventional MLLR especially when the adaptation data size is small.

Table I The performance of the WMLLR method, the MAPLR method and the conventional MLLR method in the recognition testing experiments of the different amount of adaptation data.

Speaker	Adaptation Method	Recognition Rate (%)					
		Numbers of Adaptation Utterances					
		0	2	4	6	8	10
A	WMLLR	90	91.67	93.33	93.33	93.33	93.33
	MAPLR		90	91.67	93.33	93.33	93.33
	MLLR		86.67	90	91.67	91.67	93.33
B	WMLLR	90	90	91.67	93.33	93.33	93.33
	MAPLR		90	91.67	93.33	93.33	93.33
	MLLR		86.67	91.67	91.67	93.33	93.33
C	WMLLR	91.67	91.67	93.33	93.33	93.33	93.33
	MAPLR		91.67	93.33	93.33	93.33	93.33
	MLLR		88.33	93.33	93.33	93.33	93.33
D	WMLLR	91.67	91.67	93.33	93.33	93.33	93.33
	MAPLR		91.67	93.33	93.33	93.33	93.33
	MLLR		90	93.33	93.33	93.33	93.33
E	WMLLR	90	90	93.33	93.33	93.33	93.33
	MAPLR		90	93.33	93.33	93.33	93.33
	MLLR		85	91.67	91.67	91.67	91.67
F	WMLLR	90	91.67	93.33	93.33	93.33	93.33
	MAPLR		90	91.67	91.67	91.67	93.33
	MLLR		88.33	91.67	91.67	91.67	91.67
G	WMLLR	91.67	91.67	95	95	95	95
	MAPLR		91.67	93.33	93.33	93.33	93.33
	MLLR		86.67	91.67	91.67	91.67	93.33
H	WMLLR	90	91.67	93.33	93.33	93.33	93.33
	MAPLR		90	91.67	91.67	93.33	93.33
	MLLR		88.33	91.67	91.67	91.67	91.67
I	WMLLR	91.67	91.67	93.33	93.33	93.33	93.33
	MAPLR		91.67	93.33	93.33	93.33	93.33
	MLLR		88.33	91.67	91.67	91.67	93.33
J	WMLLR	86.67	88	91.67	91.67	93.33	93.33
	MAPLR		86.67	90	91.67	91.67	93.33
	MLLR		83.33	88.33	88.33	88.33	90
K	WMLLR	83.33	86.67	91.67	91.67	91.67	91.67
	MAPLR		85	88.33	88.33	90	90
	MLLR		81.67	85	85	86.67	86.67
L	WMLLR	98.33	98.33	98.33	98.33	100	100
	MAPLR		98.33	98.33	98.33	100	100
	MLLR		91.67	98.33	98.33	98.33	100
M	WMLLR	91.67	91.67	93.33	93.33	93.33	93.33
	MAPLR		91.67	93.33	93.33	93.33	93.33
	MLLR		85	90	91.67	91.67	93.33
N	WMLLR	90	90	91.67	91.67	91.67	93.33
	MAPLR		90	91.67	91.67	91.67	93.33
	MLLR		86.67	90	90	91.67	91.67
O	WMLLR	90	91.67	93.33	93.33	93.33	93.33
	MAPLR		91.67	93.33	93.33	93.33	93.33
	MLLR		83.33	91.67	91.67	91.67	91.67
Average	WMLLR	90.45	91.20	93.33	93.44	93.66	93.78
	MAPLR		90.67	92.55	92.89	93.22	93.55
	MLLR		86.67	91.33	91.56	91.89	92.56

V. CONCLUSIONS

This paper has presented the concept of the MLLR adaptation method combined with the initial model using a direct combination factor. From the performance evaluation, the proposed WMLLR adaptation method was significantly superior to the conventional MLLR adaptation method without any knowledge of the initial model. Besides, unlike the MAPLR, the proposed method is relatively simple and fast. More advanced and efficient combination approaches have been studying for further improvements on speaker adaptation.

REFERENCES

[1] J. L. Gauvain, and C. H. Lee, "Maximum A *Posteriori* Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 291-298, 1994.

[2] G. Zavaliagkos, R. Schwartz, and J. Makhoul, "Batch, Incremental and Instantaneous Adaptation Techniques for Speech Recognition," in *Proc. the International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 1995, pp. 676-679.

[3] C. H. Lee, C. H. Lin, and B. H. Juang, "A Study on Speaker Adaptation of the Parameters of Continuous Density Hidden Markov Models," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 39, no. 4, pp. 806–814, 1991.

[4] S. J. Cox, and J. S. Bridle, "Unsupervised Speaker Adaptation by Probabilistic Spectrum Fitting," in *Proc. the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 1989, pp. 294-297.

[5] 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.

[6] J. T. Chien, L. M. Lee, and H. C. Wang, "Estimation of Channel Bias for Telephone Speech Recognition," in *Proc. the International Conference on Spoken Language Processing (ICSLP)*, 1996, pp. 1840-1843.

[7] J. T. Chien, and H. C. Wang, "Telephone Speech Recognition Based on Bayesian Adaptation of Hidden Markov Models," *Speech Communication*, vol. 22, pp. 369-384, 1997.

[8] C. Chesta, O. Siohan, and C. H. Lee, "Maximum A Posteriori Linear Regression for Hidden Markov Model Adaptation," in *Proc. the European Conference on Speech Communication and Technology (EUROSPEECH)*, 1999, pp. 211-214.

[9] W. Chou, "Maximum A Posteriori Linear Regression with Elliptically Symmetric Matrix Priors," in *Proc. the European Conference on Speech Communication and Technology (EUROSPEECH)*, 1999, pp. 1-4.

[10] W. Byrne, and A. Gunawardana, "Discounted Likelihood Linear Regression for Rapid Adaptation," in *Proc. the European Conference on Speech Communication and Technology (EUROSPEECH)*, 1999, pp. 203-206.

[11] R. Kuhn, J. -C. Junqua, P. Nguyen, and N. Niedzielski, "Rapid Speaker Adaptation in Eigenvoice Space," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 6, pp. 695–707, 2000.

[12] 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. the International Conference on Spoken Language Processing (ICSLP)*, 2000, pp. 742–745.

[13] K. T. Chen, and H. M. Wang, "Eigenspace-based Maximum A *Posteriori* Linear Regression for Rapid Speaker Adaptation," in *Proc. the IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 2001, pp. 917-920.

[14] 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.

[15] M. J. F. Gales, and P. C. Woodland, "Mean and Variance Adaptation Within the MLLR Framework," *Computer and Speech Language*, vol. 10, pp. 249-264, 1996.

[16] W. G. Cochran, "Problems Arising in the Analysis of A Series of Similar Experiments," *Journal of the Royal Statistical Society*, vol. 4 (Suppl.), pp. 102-118, 1937.

[17] H. C. Wang, "MAT – A Project to Collect Mandarin Speech Data through Telephone Networks in Taiwan," *Comput. Linguist. Chinese Lang. Process.*, vol. 2, pp. 73-89, 1997.