SUBSPACE GAUSSIAN MIXTURE MODEL WITH STATE-DEPENDENT SUBSPACE DIMENSIONS

Tom Ko, Brian Mak

Department of Computer Science & Engineering
Hong Kong University of Science & Technology
{tomko,mak}@cse.ust.hk

Cheung-Chi Leung

Institute for Infocomm Research
A*STAR, Singapore
ccleung@i2r.a-star.edu.sg

ABSTRACT

In recent years, under the hidden Markov modeling (HMM) framework, the use of subspace Gaussian mixture models (SGMMs) has demonstrated better recognition performance than traditional Gaussian mixture models (GMMs) in automatic speech recognition. In state-of-the-art SGMM formulation, a fixed subspace dimension is assigned to every phone states. While a constant subspace dimension is easier to implement, it may, however, lead to overfitting or underfitting of some state models as the data is usually distributed unevenly among the states. In a later extension of SGMM, states are split to sub-states with an appropriate objective function so that the problem is eased by increasing the state-specific parameters for the underfitting state. In this paper, we propose another solution and allow each sub-state to have a different subspace dimension depending on its amount of training frames so that the state-specific parameters can be robustly estimated. Experimental evaluation on the Switchboard recognition task shows that our proposed method brings improvement to the existing SGMM training procedure.

Index Terms: subspace Gaussian mixture model, phonetic dimension, sub-state, regularization.

1. INTRODUCTION AND RELATION TO PRIOR WORK

In recent years, the use of subspace Gaussian mixture model (SGMM) [1, 2, 3] to represent states in hidden Markov model (HMM) for automatic speech recognition (ASR) has received a lot of attention. It has been reported continually that SGMM produces better recognition performance over traditional Gaussian mixture model (GMM) in many speech recognition tasks. In SGMM, the state-dependent GMM parameters are generated by a projection of the globally shared parameters [4] using state-dependent subspace vectors. The compact representation of SGMM greatly reduces the number of free model parameters which can be then estimated robustly even when the amount of training data is highly limited.

In the state-of-the-art SGMM formulation, the phonetic subspace dimension S is a constant that is fixed for all phonetic states. Due to the uneven distribution of data over the different phonetic states, some states may suffer from overfitting when S is large while some others may suffer from underfitting when S is small. The later extension to the use of sub-states can ease the problem by increasing state-specific parameters appropriately for states with more training data. To see the effect of adding more sub-states on the distribution of data in the sub-states, various SGMM-HMMs with 5,000 tied states were trained using 30 hours of Switchboard data and the number of sub-states were increased from 5,000 to 30,000. The sub-states are sorted in descending order of their data occurrences, and the accumulative percentage of data covered by them is shown in Fig. 1. In the ideal case when each sub-state has the same amount of data, the plot should be a straight line with a slope of 1. From the figure, it can be seen that increasing the number of sub-states may only improve the evenness of data distribution slightly.

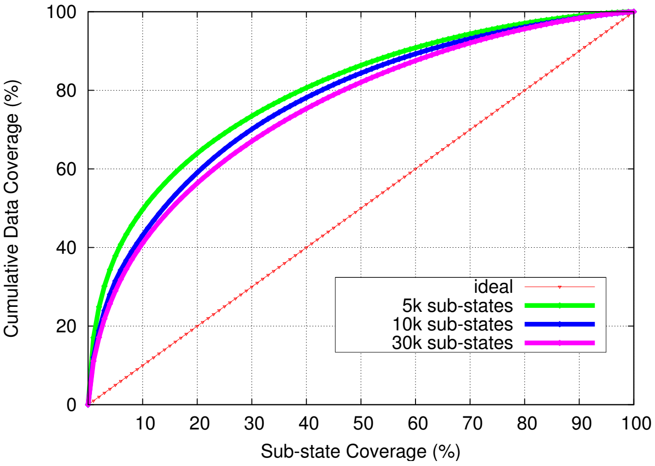

Fig. 1. Data coverage of SGMM-HMM systems with different number of sub-states.

Based on the framework of SGMM, regularized SGMM [5] is proposed to address the possible overfitting that arises when the training data is highly limited. In their work, maximum likelihood estimation of the sub-state vector $\mathbf{v}$ is regularized by adding a penalty term to the objective function. Both ℓ_1- and ℓ_2-regularization have been tried with limited success. When regularization is used, some elements of the sub-state vectors are driven towards zero.

In this paper, we would like to use another approach to address the overfitting *and* underfitting problems caused by the uneven distribution of data. We propose to use a state-dependent subspace dimension S_{jm} for each sub-state m of the tied state j. Intuitively, a sub-state with more training data should use a larger S_{jm} value, and a sub-state with less training data should use a smaller S_{jm} value.

Our proposed method is different from the regularized SGMM [5] in the following aspects:

- In [5], column vectors in the shared mean projection $\mathbf{M}_i$ are

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

not individually weighted. Thus, all elements in the sub-state vector $\mathbf{v}_{jm}$ receive the same force shrinking them towards zero from the regularization term. In contrast, our method is more similar to the regularization term used in [6, 7] where the eigenvectors are individually weighted. In our proposed method, some elements of a sub-state vector $\mathbf{v}_{jm}$ are forced to zero if it has less training data, and we do that by pushing all these zero elements to the end of the sub-state vector. As a side effect, the leading column vectors of $\mathbf{M}_i$ will be effectively shared by a larger number of sub-states, and thus are considered to be more "important".

- Our method is more like choosing the number of eigenvectors in the eigenvoice speaker adaptation method [8], and a hard decision is made on the number of vectors composing the phonetic subspace.

We evaluated our proposed method on the Switchboard corpus. Different ways of determining the variable subspace dimension were also compared. The rest of the paper is organized as follows. In Section 2, a review of the SGMM formulation is given. We then describe our proposed method in Section 3. That is followed by experimental evaluation in Section 4 and conclusions in Section 5.

Table 1. Description of various SGMM parameters.

Notation	**Description**
i	Gaussian component index
j	tied-state index
m	sub-state index
M_j	number of sub-states for tied-state j
I	number of Gaussians for each state/sub-state
$\mathbf{v}_j$/$\mathbf{v}_{jm}$	state-specific/sub-state-specific subspace vector
$\mathbf{M}_i$	mean projection matrix for the ith Gaussian
$\mathbf{w}_i$	weight projection vector for the ith Gaussian
c_{jm}	sub-state weight
$\mathbf{v}^{(s)}$	speaker-specific subspace vector
$\mathbf{N}_i$	mean projection matrix for the ith Gaussian regarding the speaker offset
$\mathbf{u}_i$	weight projection vector for the ith Gaussian regarding the speaker offset

2. REVIEW OF SGMM

The basic form of SGMM can be expressed by the following formulas:

$$p(\mathbf{x}(t)|j) = \sum_{i=1}^{I} w_{ji} \mathcal{N}(\mathbf{x}(t); \mu_{ji}, \mathbf{\Sigma}_i) \quad (1)$$

$$\mu_{ji} = \mathbf{M}_i \mathbf{v}_j \quad (2)$$

$$w_{ji} = \frac{exp(\mathbf{w}_i^T \mathbf{v}_j)}{\sum_{i'=1}^{I} exp(\mathbf{w}_{i'}^T \mathbf{v}_j)} \quad (3)$$

The description of various SGMM parameters is summarized in Table 1. The means of Gaussian i of state j is generated from the state-specific vector $\mathbf{v}_j$ through the shared mean projection $\mathbf{M}_i$. This is similar to eigenvoices [8] and cluster adaptive training [9]. Indeed, SGMM is highly novel in generating the Gaussian weights through the weight projections $\mathbf{w}_i$.

The above basic form of SGMM can be extended to the use of sub-states as follows:

$$p(\mathbf{x}(t)|j) = \sum_{m=1}^{M_j} c_{jm} \sum_{i=1}^{I} w_{jmi} \mathcal{N}(\mathbf{x}(t); \mu_{jmi}, \mathbf{\Sigma}_i) \quad (4)$$

$$\mu_{jmi} = \mathbf{M}_i \mathbf{v}_{jm} \quad (5)$$

$$w_{jmi} = \frac{exp(\mathbf{w}_i^T \mathbf{v}_{jm})}{\sum_{i'=1}^{I} exp(\mathbf{w}_{i'}^T \mathbf{v}_{jm})} \quad (6)$$

In this paper, our experiments are based on two further extensions of SGMM related to speaker adaptation. The first extension [3] is the addition of a speaker-dependent offset to the mean vector of each Gaussian. Thus, eqs. (4) and (5) become

$$p(\mathbf{x}(t)|j,s) = |det\mathbf{A}^{(s)}| \sum_{m=1}^{\mathbf{M}_j} c_{jm} \sum_{i=1}^{I} w_{jmi} \mathcal{N}(\mathbf{x}(t)'; \mu_{jmi}^{(s)}, \mathbf{\Sigma}_i) (7)$$

$$\mu_{jmi}^{(s)} = \mathbf{M}_i \mathbf{v}_{jm} + \mathbf{N}_i \mathbf{v}^{(s)} \quad (8)$$

where $\mathbf{x}(t)' = \mathbf{A}^{(s)}\mathbf{x}(t) + \mathbf{b}^{(s)}$ is the transformed feature vector after constrained MLLR (CMLLR) transformation [10], with $\mathbf{A}^{(s)}$ and $\mathbf{b}^{(s)}$ being the speaker-specific transformation matrix and bias vector respectively. The second extension is called symmetric subspace Gaussian mixture model (SSGMM) [11]. In SSGMM, besides the mean vectors, the weight vectors are also symmetrized. Thus, eq. (6) becomes

$$w_{jmi}^{(s)} = \frac{exp(\mathbf{w}_i^T \mathbf{v}_{jm} + \mathbf{u}_i^T \mathbf{v}^{(s)})}{\sum_{i'=1}^{I} exp(\mathbf{w}_{i'}^T \mathbf{v}_{jm} + \mathbf{u}_{i'}^T \mathbf{v}^{(s)})} \quad (9)$$

Similar to speaker adaptive training (SAT) [12], $\mathbf{N}_i$ and $\mathbf{u}_i$ are estimated during training, while $\mathbf{v}^{(s)}$ of the test speakers are estimated during decoding. For ease of explanation, the above extensions related to speaker adaptation will be omitted in the SGMM formulation in the following section.

3. SGMM WITH STATE-DEPENDENT SUBSPACE DIMENSIONS

In the SGMM formulation described in Section 2, a constant subspace dimension S is used for every sub-state so that $\mathbf{v}_{jm} \in \mathbb{R}^S$, $\forall(j,m)$. In this paper, we would like to further generalize the SGMM framework using a variable subspace dimension S_{jm} for each sub-state. Thus, we require $\mathbf{v}_{jm} \in \mathbb{R}^{S_{jm}}$, $\forall(j,m)$ and $S^{min} \leq S_{jm} \leq S^{max}$, where S^{max} and S^{min} are the upper bound and the lower bound of the subspace dimension respectively. Our proposed method can be simulated by assigning zeros to the last $(S^{max} - S_{jm})$ elements starting from the $(S_{jm}+1)$th position during the estimation of $\mathbf{v}_{jm}$. That is, we have

$$\mathbf{v}_{jm} = \left[\begin{array}{c} v_{jm1} \\ \vdots \\ v_{jmS_{jm}} \\ \hdashline 0 \\ \vdots \\ 0 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c} v \\ \vdots \\ v \end{array}}\right\} S_{jm} \\ \left.\vphantom{\begin{array}{c} 0 \\ \vdots \\ 0 \end{array}}\right\} S^{max} - S_{jm} \end{array} . \quad (10)$$

There are S^{max} column vectors in M_i accordingly. Due to the structure of v_{jm} in eq. (10), only the leading S_{jm} column vectors of M_i are effective in generating μ_{jmi} through eq. (5). Thus, although M_i is globally shared across all sub-states, the leading column vectors are shared by more sub-states while those trailing column vectors are shared by fewer sub-states. The same also applies to the weight projection vectors.

In our work, we follow the training procedures suggested in [13], where $\mathbf{M}_i$ and $\mathbf{w}_i$ are re-trained after the re-estimation of $\mathbf{v}_{jm}$, and maximum likelihood estimation is employed. The auxiliary function for the mean projection matrix $\mathbf{M}_i$ is:

$$\mathcal{Q}(\mathbf{M}_i) = tr(\mathbf{M}_i^T \mathbf{\Sigma}_i^{-1} \mathbf{Y}_i) - \frac{1}{2} tr(\mathbf{\Sigma}_i^{-1} \mathbf{M}_i \mathbf{Q}_i \mathbf{M}_i^T) \quad (11)$$

where

$$\mathbf{Y}_i = \sum_{t,j,m} \gamma_{jmi}(t) \mathbf{x}(t) \mathbf{v}_{jm}^T , \quad (12)$$

$$\mathbf{Q}_i = \sum_{j,m} \gamma_{jmi} \mathbf{v}_{jm} \mathbf{v}_{jm}^T , \quad (13)$$

and $tr()$ is the trace function of a square matrix; $\mathbf{x}(t)$ is the feature vector at time t; $\gamma_{jmi}(t)$ is the posterior of $\mathbf{x}(t)$ over Gaussian i in the mth sub-state of the jth tied-state; $\gamma_{jmi} = \sum_t \gamma_{jmi}(t)$ is the occupation count of the Gaussian. Provided that $\mathbf{Q}_i$ is not singular, there is a closed-form solution, which is

$$\hat{\mathbf{M}}_i = \mathbf{Y}_i \mathbf{Q}_i^{-1} . \quad (14)$$

In order to compute $\mathbf{Y}_i$ and $\mathbf{Q}_i$ in eqs. (11) and (13), technically, it is easier to keep every $\mathbf{v}_{jm}$ with the same dimension. That is why our method is simulated by the structure of v_{jm} in eq. (10).

3.1. Determining the State-dependent Subspace Dimensions

The major guideline in determining the state-dependent subspace dimension is to assign larger subspace dimensions to the sub-states with more training data. Here, we investigate two ways of dimension assignment.

1. *absolute-count-based method*, $(S_{jm} - S^{min}) \propto \gamma_{jm}$: We let $(S_{jm} - S^{min})$ be proportional to $\gamma_{jm} = \sum_i \gamma_{jmi}$ which is the occupation count of the sub-state (j, m). S_{jm} is computed as follows:

$$S_{jm} = \left\lceil \frac{\gamma_{jm}}{k_0} \right\rceil + S^{min} \quad (15)$$

where k_0 is a constant. If the resulting S_{jm} is larger than S^{max}, $S_{jm} = S^{max}$.

2. *ordinal-count-based method*, $(S_{jm} - S^{min}) \propto R_{jm}$: We sort the sub-states according to their occupation counts in ascending order and let R_{jm} be the ordinal position of the sub-state in this sorted order where $1 \leq R_{jm} \leq \sum_j M_j$. We then let $(S_{jm} - S^{min})$ be proportional to R_{jm}. S_{jm} is computed as follows:

$$S_{jm} = \left\lceil \frac{R_{jm}}{\sum_j M_j} (S^{max} - S^{min}) \right\rceil + S^{min} . \quad (16)$$

4. EXPERIMENTAL EVALUATION

Two training sets from Switchboard I [14], a 30-hour training set and a 100-hour training set were used separately for the acoustic model estimation [1]. We first evaluated our method on the basic form of SGMM (with no sub-states) using the 30-hour training set, and then on SGMM with sub-states using the 100-hour training set. The 30-hour training set contains 24,569 utterances and the 100-hour training set contains 76,615 utterances. Recognition results are reported on the standard Hub5 2000 evaluation set. It consists of 1,831 Switchboard utterances and 2,628 Callhome utterances. There are a total of about 2 hours of conversational speech.

MFCC features with cepstral mean and variance normalization were used as the acoustic vectors. Seven consecutive feature vectors, each consisting of 13 static MFCC coefficients, were concatenated, and then they were reduced to 40-dimensional feature vectors using LDA [15]. MLLT [16], SAT [12] and fMLLR [10] were then applied on the features and acoustic models.

A trigram language model was trained on all the transcribed data of Switchboard I (about 284 hours) using the SRILM toolkit [17]. Acoustic model estimation and recognition were performed using the Kaldi toolkit [13]. The SGMM parameters were updated in 25 iterations as in the Kaldi recipe for Switchboard. For the baseline systems, S was set to 41 as recommended by Kaldi. It was then gradually increased to search for its best value which gave the best recognition on the test set. The best value of S was found to be 200. Thus, in our proposed method we set S^{min} to 41 and S^{max} to 200. No renormalization[2] was applied on the SGMM parameters.

A series of SGMM systems were implemented to answer the following questions:

- Is a larger subspace dimension better?
- Is the use of state-dependent subspace dimensions better than the use of a fixed subspace dimension?
- Which of the two methods in Section 3.1 is better to determine the state-dependent subspace dimensions?
- How is our method compared with regularized SGMM [5]?

In [5], ℓ_1-norm, ℓ_2-norm and elastic net have been reported with similar performance. For the last question, we implemented a regularized SGMM using the ℓ_2-norm regularization where an ℓ_2 penalty is added to the original ML objective function of $\mathbf{v}_{jm}$ as follows:

$$\hat{\mathbf{v}}_{jm} = \underset{\mathbf{v}_{jm}}{\operatorname{argmax}} \ \log p(O|\mathbf{v}_{jm}, \theta) - \lambda_2 \sum_r |\mathbf{v}_{jmr}|^2 \quad (17)$$

where O denotes all the acoustic observations; θ denotes the current SGMM parameters; λ_2 is the regularization parameter to weight the ℓ_2 penalty and $\mathbf{v}_{jmr}$ is the rth element of $\mathbf{v}_{jm}$.

4.1. Experiments on Basic SGMM with No Sub-states

The following 5 basic SGMM systems with no sub-states were trained for comparison:

- basic SGMM with a fixed subspace dimension, S = 41
- basic SGMM with a fixed subspace dimension, S = 200

[1] The partition of these two training sets were suggested by the Switchboard recipe in the Kaldi toolkit.

[2] We empirically find that using no renormalization does not affect the accuracy in our task.

- regularized SGMM with a fixed subspace dimension, S = 200; ℓ_2-norm regularization[3] with λ_2 = 10
- basic SGMM with state-dependent subspace dimension[4], $(S_{jm} - S^{min}) \propto \gamma_{jm}$, S^{max} = 200, S^{min} = 41, k_0 = 100
- basic SGMM with state-dependent subspace dimension, $(S_{jm} - S^{min}) \propto R_{jm}$, S^{max} = 200, S^{min} = 41

The above systems were trained using the 30-hour training set. All the systems have 5,000 tied-states with 700 Gaussians in each state. The values of these parameters were determined in a preliminary experiment. Word recognition results of these systems are shown in Table 2.

Table 2. Word recognition accuracy (%) of various SGMM systems with no sub-states on HUB5 Eval2000 using the 30-hour training set. Recognition was done with a trigram language model.

Model Description	**Acc.**
Basic SGMM with fixed subspace dimension, S = 41	66.9
Basic SGMM with fixed subspace dimension, S = 200	67.3
Regularized SGMM [5] with fixed subspace dimension, S = 200; ℓ_2-norm regularization with λ_2 = 10	67.2
Basic SGMM with variable subspace dimension, $(S_{jm} - S^{min}) \propto \gamma_{jm}$, S^{max} = 200, S^{min} = 41, k_0 = 100	67.4
Basic SGMM with variable subspace dimension, $(S_{jm} - S^{min}) \propto R_{jm}$, S^{max} = 200, S^{min} = 41	**67.9**

4.1.1. Results and Discussions

From the experimental results of the basic form of SGMM (with no sub-states) using a smaller training set in Table 2, we find that

- When the subspace dimension is fixed, using a larger dimension S = 200 gives an absolute 0.4% improvement over using the Kaldi default S value of 41. We believe that the use of a larger subspace dimension avoids the underfitting problem in the estimation of sub-state probability density functions.
- Using state-dependent subspace dimensions may further give a 0.1% and 0.6% absolute improvement over using a fixed subspace dimension S = 200. The ordinal-count-based method is better in the determination of the state-dependent subspace dimension. We attribute the improvement to the resolve of overfitting of sub-states which may have occurred when S is large. In fact, we further analyze the number of zero elements in the two subspace dimension determination methods and find that the absolute-count-based method results in reducing 69% of $\mathbf{v}_{jm}$ elements to zero, while the corresponding figure for the ordinal-count-based method is only 40%. The absolute-count-based method is probably too agressive.
- Our implementation of the regularized SGMM using ℓ_2-norm regularization does not suggest any improvement.

Table 3. Word recognition accuracy (%) of various SGMM systems with sub-states on HUB5 Eval2000 using the 100-hour training set. Recognition was done with a trigram language model.

Model Description	**Acc.**
SGMM with fixed subspace dimension, S = 41	70.4
SGMM with fixed subspace dimension, S = 200	70.8
SGMM with variable subspace dimension, $(S_{jm} - S^{min}) \propto R_{jm}$, S^{max} = 200, S^{min} = 41	**71.1**

4.2. Experiments on SGMM with Sub-states

The following 3 SGMM systems with sub-states were trained for comparison:

- SGMM with fixed subspace dimension, S = 41
- SGMM with fixed subspace dimension, S = 200
- SGMM with state-dependent subspace dimension, $(S_{jm} - S^{min}) \propto R_{jm}$, S^{max} = 200, S^{min} = 41

The above systems were trained using the 100-hour training set. All the systems have 9,000 tied-states and a total of 30,000 sub-states with 700 Gaussians in each sub-state. These figures are suggested by the recipe in the Kaldi toolkit. Word recognition results of these systems are shown in Table 3.

4.2.1. Results and Discussions

The performances of the various SGMM systems with sub-states that were trained using the 100-hour training set are shown in Table 3. Similar findings are observed: increasing S from 41 to 200 gives a 0.4% absolute recognition improvement, and the use of state-dependent subspace dimensions on top of a large S gives another 0.3% improvement.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we investigate the use of state-dependent subspace dimensions in SGMM to address the possible overfitting problem caused by uneven distribution of training data over the states/sub-states. In our proposed method, a sub-state with more training data will be assigned a larger subspace dimension to increase its model complexity. According to the structure of $\mathbf{v}_{jm}$ in our method, the leading column vectors of the mean projection matrix $\mathbf{M}_i$ are considered to be more "important" as they are shared by a larger number of sub-states whereas its trailing column vectors will be shared by sub-states with relatively less data.

In the future, we would like to extend our method to use variable dimensions on the speaker-subspace vectors $\mathbf{v}^{(s)}$. Also, we would like to investigate the use of variable subspace dimension on discriminative training of SGMM.

6. ACKNOWLEDGEMENTS

This work was supported by the Research Grants Council of the Hong Kong SAR under the grant number HKUST616513.

[3] The λ_2 value is the same as the one used in [5].

[4] k_0 was chosen empirically so that the mapping of S_{jm} to γ_{jm} is one-to-many.

7. REFERENCES

[1] D. Povey et al., "Subspace Gaussian mixture models for speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2010, pp. 4330–4333.

[2] D. Povey et al., "Multilingual acoustic modeling for speech recognition based on subspace Gaussian mixture model," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2010, pp. 4334–4337.

[3] D. Povey and L. Burget et al., "The subspace Gaussian mixture model — A structured model for speech recognition," *Computer Speech and Language*, vol. 25, no. 2, pp. 404–439, 2011.

[4] M. J. F. Gales and K. Yu, "Canonical state models for automatic speech recognition," in *Proceedings of Interspeech*, 2010, pp. 58–61.

[5] L. Lu, A. Ghoshal, and S. Renals, "Regularized subspace Gaussian mixture models for speech recognition," *IEEE Signal Processing Letters*, vol. 18, pp. 419–422, 2011.

[6] T. Ko and B. Mak, "Eigentriphones for context-dependent acoustic modeling," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 6, pp. 1285–1294, 2013.

[7] T. Ko and B. Mak, "A fully automated derivation of state-based eigentriphones for triphone modeling with no tied states using regularization," in *Proceedings of Interspeech*, 2011, pp. 781–784.

[8] 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. 4, pp. 695–707, 2000.

[9] M. J. F. Gales, "Multiple-cluster adaptive training schemes," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 361–364.

[10] M. J. F. Gales, "Maximum likelihood linear transformations for HMM-based speech recognition," *Computer Speech and Language*, vol. 25, no. 2, pp. 75–78, 1997.

[11] D. Povey, M. Karafiat, A. Ghoshal, and P. Schwarz, "A symmetrization of the subspace Gaussian mixture model," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2011, pp. 4504–4507.

[12] T. Anastasakos, J. McDonough, R. Schwartz, and J. Makhoul, "A compact model for speaker-adaptive training," in *Proceedings of the International Conference on Spoken Language Processing*, 1996, pp. 1137–1140.

[13] D. Povey et al., "The Kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*, 2011.

[14] J. J. Godfrey, E. C. Holliman, and J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992, pp. 517–520.

[15] R. Haeb-Umbach and H. Ney, "Linear discriminant analysis for improved large vocabulary continuous speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992, pp. 13–16.

[16] M. J. F. Gales, "Semi-tied covariance matrices for hidden Markov models," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 272–281, 1999.

[17] A. Stolcke, "SRILM — An extensible language modeling toolkit," in *Proceedings of the International Conference on Spoken Language Processing*, 2002, pp. 901–904.