REGULARIZED CONSTRAINED MAXIMUM LIKELIHOOD LINEAR REGRESSION FOR SPEECH RECOGNITION

Sina Hamidi Ghalehjegh and Richard C. Rose

Department of Electrical and Computer Engineering, McGill University, Montreal, Canada

ABSTRACT

The use of a graph embedding framework is investigated as a regularization technique in the expectation-maximization (EM) algorithm applied to automatic speech recognition (ASR). The technique is motivated by the fact that graph embeddings of feature vectors have been shown to provide useful characterizations of the underlying manifolds on which these features lie. Incorporating intrinsic graphs that describe these manifolds in the optimization criteria for the EM algorithm has the effect of constraining the solution space in a way that preserves the local structure of the data. Graph embedding based regularization is applied here to estimating parameters in constrained maximum likelihood linear regression (CMLLR) speaker adaptation in continuous density hidden Markov model (CDHMM) based ASR. CMLLR adaptation has been widely used as a maximum likelihood procedure for reducing mismatch between a given HMM model and utterances from an unknown speaker through a linear feature space transformation. However, there is no guarantee that CMLLR transformations will preserve the relationships of the feature vectors along this manifold. It is argued here that graph embedding based regularization will preserve this structure. The impact of this approach on ASR performance is evaluated for unsupervised speaker adaptation on two large vocabulary speech corpora.

Index Terms— Graph embedding, Regularization, Speaker adaptation, Constrained MLLR

1. INTRODUCTION

Applying a transformation either in model-space or feature-space has been shown to be a powerful tool for speaker adaptation in CDHMM based ASR [1–4]. The most well known of these techniques include maximum likelihood linear regression (MLLR) [1–3, 5], applied to model space adaptation, and constrained MLLR (CMLLR) [6, 7], applied to feature space adaptation. The parameters of the transformation are estimated according to a maximum likelihood (ML) criterion using the expectation-maximization (EM) algorithm [8]. In the case of CMLLR, the resulting regression based transformation can be said to have the effect of transforming the feature vectors so that the adaptation data is more likely to have been generated by the CDHMM. However, there is no guarantee that the local relationships between vectors in the original feature space will be preserved among the CMLLR transformed features.

There have been two approaches applied to constraining the estimates of CMLLR/MLLR parameters. The first is to use a maximum a posteriori (MAP) criterion which introduces a prior distribution on the parameters [9, 10]. In MAP estimation, some values of parameters are more likely than others and this knowledge can be used to preserve the structure of the acoustic space [9, 10]. Another approach is to use a regularizer term in the optimization criterion in order to constrain the solution space. This can also be interpreted as an alternative way of selecting the prior distribution function [11, 12].

In this paper a new regularization approach is presented which estimates the transformation parameters while preserving the local relationships among the feature vectors in the transformed space. To do so, a graph-embedding framework is used [13]. The notion of using graph-embedding as a geometric framework for learning from labeled and unlabeled data was first proposed in [14] and was referred to as manifold regularization. Manifold regularization has also been applied to semi-supervised learning for multi-layer perceptrons [15] and deep learning [16]. In this paper, the framework is used to characterize the geometric properties of feature vectors derived from unlabeled speech spectra. In Section 3, manifold regularization is described as a two step procedure. First, a characterization of the feature space is acquired by estimating an intrinsic graph in the form of an intrinsic matrix from feature vectors. Then, the intrinsic matrix is incorporated into the auxiliary function for CMLLR parameters estimation.

The paper is organized as follows. Section 2 describes the standard CMLLR technique. Section 3 describes the new regularization formulation. In Section 4 we describe our experimental setup and the ASR results. Finally we conclude the paper in Section 5.

2. CONSTRAINED MLLR

Constrained MLLR (CMLLR) is a feature-space adaptation technique which was first introduced in [2]. Assuming $\mathbf{o}(\tau)$

This work was supported by the Nuance Foundation and the Quebec Government MDEIE.

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

as the feature vector at time τ, the adapted feature vector is:

$$\hat{\mathbf{o}}(\tau) = \mathbf{A}\mathbf{o}(\tau) + \mathbf{b} = \mathbf{W}\boldsymbol{\zeta}(\tau), \quad (1)$$

where $\mathbf{W}_{d\times(d+1)} = [\ \mathbf{A}_{d\times d} \quad \mathbf{b}_{d\times 1}\ ]$ is the transformation matrix and $\boldsymbol{\zeta}(\tau) = [\ \mathbf{o}(\tau)^T \quad 1\ ]^T$ is the extended observation vector. Considering diagonal covariance matrices, the auxiliary function for estimating $\mathbf{W}$ is given by:

$$\mathcal{Q}(\mathbf{W}) = \beta \log(\mathbf{p}_i \mathbf{w}_i^T) - \tfrac{1}{2}\textstyle\sum_i \left(\mathbf{w}_i \mathbf{G}^{(i)} \mathbf{w}_i^T - 2\mathbf{w}_i \mathbf{k}^{(i)T}\right), \quad (2)$$

where $\mathbf{w}_i$ is the i^{th} row of $\mathbf{W}$, $\mathbf{p}_i = [\ c_{i1} \quad \dots \quad c_{in} \quad 0\ ]$ is the extended cofactor vector in which $c_{ij} = \text{cof}(\mathbf{A}_{ij})$ and

$$\begin{aligned} \mathbf{G}^{(i)} &= \sum_m \frac{1}{\sigma_i^{(m)2}} \sum_\tau \gamma_m(\tau)\boldsymbol{\zeta}(\tau)\boldsymbol{\zeta}^T(\tau) \\ \mathbf{k}^{(i)} &= \sum_m \frac{1}{\sigma_i^{(m)2}} \mu_i^{(m)} \sum_\tau \gamma_m(\tau)\boldsymbol{\zeta}^T(\tau) \\ \beta &= \sum_m \sum_\tau \gamma_m(\tau). \end{aligned}$$

$\mu_i^{(m)}$ and $\sigma_i^{(m)}$ are the i^{th} and the $(i,i)^{\text{th}}$ elements of the mean and covariance for component m, respectively [17]. $\gamma_m(\tau)$ is the probability of being in component m at time τ and $\mathbf{O} = \{\mathbf{o}(1), \dots, \mathbf{o}(T)\}$ is the adaptation data sequence. Differentiating $\mathcal{Q}(\mathbf{W})$ w.r.t. $\mathbf{w}_i$ and equating it with zero, the i^{th} row of $\mathbf{W}$ can be found as:

$$\mathbf{w}_i = \alpha \mathbf{p}_i \mathbf{G}^{(i)-1} + \mathbf{k}^{(i)} \mathbf{G}^{(i)-1}, \quad (3)$$

where α is the root of equation defined as below [17]:

$$\alpha^2 \mathbf{p}_i \mathbf{G}^{(i)-1} \mathbf{p}_i^T + \alpha \mathbf{p}_i \mathbf{G}^{(i)-1} \mathbf{k}^{(i)T} - \beta = 0.$$

3. REGULARIZATION APPROACH

3.1. Defining regularized auxiliary function

The goal here is to define a cost function in order to preserve the locality of the feature vectors in the adapted space. This will follow the graph-embedding framework for characterizing geometric properties of the feature vectors [13]. We define an undirected weighted graph, called intrinsic graph $\mathcal{G} = \{\mathbf{O}, \boldsymbol{\Omega}\}$, where $\mathbf{O}$ represents the graph nodes and contains all the unadapted feature vector and $\boldsymbol{\Omega}$ is the intrinsic affinity weight matrix with elements defined as:

$$\omega_{\tau\tau'} = \begin{cases} \exp\left(-\frac{\|\mathbf{o}(\tau)-\mathbf{o}(\tau')\|_2^2}{\rho}\right) & I(\mathbf{o}(\tau),\mathbf{o}(\tau')) = 1 \\ 0 & I(\mathbf{o}(\tau),\mathbf{o}(\tau')) = 0 \end{cases}, \quad (4)$$

where ρ is the kernel scale parameter. The indicator function $I(\mathbf{o}(\tau),\mathbf{o}(\tau'))$ is equal to 1 if $\mathbf{o}(\tau)$ and $\mathbf{o}(\tau')$ lie within the same regression class[1] *and* are close in terms of Euclidean distance. In the other words, $\mathbf{o}(\tau')$ is close to $\mathbf{o}(\tau)$ if it is in the K-nearest neighbor of $\mathbf{o}(\tau)$.

In the adapted space, we would like to preserve the closeness of the nearby points. It means that if two points of the graph in the original space are close, i.e. have large weight $\omega_{\tau\tau'}$, they should be close in the adapted space as well. So, we introduce a graph-preserving measure:

$$\mathcal{S}(\mathbf{W}) = \frac{1}{T} \sum_{\tau\ \neq\tau'} \|\hat{\mathbf{o}}(\tau) - \hat{\mathbf{o}}(\tau')\|_2^2\ \omega_{\tau\tau'}, \quad (5)$$

in which $\hat{\mathbf{o}}(\tau)$ and $\hat{\mathbf{o}}(\tau')$ are observation vectors in adapted space corresponding to $\mathbf{o}(\tau)$ and $\mathbf{o}(\tau')$, respectively. Following from (4), it is clear that $\omega_{\tau\tau'}$ is zero for observation vectors from different classes. So, we can re-write the graph-preserving measure for *a given class* as follows:

$$\mathcal{S}(\mathbf{W}) = \frac{1}{T} \sum_{\tau\neq\tau'} \|\mathbf{W}\boldsymbol{\zeta}(\tau) - \mathbf{W}\boldsymbol{\zeta}(\tau')\|_2^2\ \omega_{\tau\tau'}. \quad (6)$$

By defining $\mathbf{R}$ as an affinity matrix:

$$\mathbf{R} = \frac{1}{T} \sum_{\tau\neq\tau'} \omega_{\tau\tau'} \left(\boldsymbol{\zeta}(\tau) - \boldsymbol{\zeta}(\tau')\right)\left(\boldsymbol{\zeta}(\tau) - \boldsymbol{\zeta}(\tau')\right)^T, \quad (7)$$

and re-writing $\mathcal{S}$ in term of rows of $\mathbf{W}$, we have:

$$\mathcal{S}(\mathbf{W}) = \sum_i \mathbf{w}_i \mathbf{R} \mathbf{w}_i^T. \quad (8)$$

Now, we define the regularized auxiliary function as:

$$\mathcal{Q}_{\text{reg}}(\mathbf{W}) = \mathcal{Q}(\mathbf{W}) - \theta\, \mathcal{S}(\mathbf{W}), \quad (9)$$

in which $\theta > 0$ is the trade-off parameter between the ML-based auxiliary function and locality preserving measure. Substituting (2) and (8) in (9), we have:

$$\mathcal{Q}_{\text{reg}}(\mathbf{W}) = \beta \log(\mathbf{p}_i \mathbf{w}_i^T) - \tfrac{1}{2}\textstyle\sum_i \left(\mathbf{w}_i \mathbf{G}'^{(i)} \mathbf{w}_i^T - 2\mathbf{w}_i \mathbf{k}^{(i)T}\right), \quad (10)$$

in which $\mathbf{G}'^{(i)} = \mathbf{G}^{(i)} + 2\theta\mathbf{R}$. It is worth noting that this auxiliary function is very similar to that of standard CMLLR except for the additional term related to the locality preserving measure. Therefore, we can use the same optimization procedure only with substituting $\mathbf{G}^{(i)}$ with $\mathbf{G}'^{(i)}$.

3.2. Relation to other regularization techniques

It is interesting to note that in the degenerate case, when the affinity matrix $\mathbf{R}$ is set to:

$$\mathbf{R} = \begin{bmatrix} \mathbf{I}_{d\times d} & \mathbf{0}_{d\times 1} \\ \mathbf{0}_{1\times d} & 0 \end{bmatrix}, \quad (11)$$

then the graph-preserving measure will reduce to the square of Frobenius norm of transformation matrix:

$$\mathcal{S}(\mathbf{W}) = \sum_i \mathbf{w}_i \mathbf{R} \mathbf{w}_i^T = \sum_i \mathbf{a}_i \mathbf{a}_i^T = \|\mathbf{A}\|_F^2. \quad (12)$$

This case is studied in [12] under the name of "shrinkage model adaptation" in the context of regularized CMLLR.

[1] For example regression class of `speech` or `silence`

3.3. Discussion

The need for calculating on the order of N^2 nearest neighbor relationships in (4), where N is the number of feature vectors, makes the computational complexity associated with estimating the affinity matrix, $\mathbf{R}$, quite high. However, estimation of $\mathbf{R}$ is generally done in an unsupervised manner so no transcribed speech is required. While separate class dependent affinity matrices are discussed in Section 3.1, only two classes for speech and non-speech frames are defined here, which simply assumes the existence of a speech/non-speech detector. For the experimental study in Section 4, affinity matrices are estimated prior to recognition from the same training data used for estimating CDHMM models. While it could be argued that manifold based regularization could benefit from affinity matrices estimated from large amounts of speaker specific data, it is not practical to estimate these matrices during any reasonable speaker adaptation scenario. This is due to both the excessive computational complexity associated with estimating affinity matrices and insufficient amounts of available adaptation data.

4. EXPERIMENTAL STUDY

This section presents an experimental study evaluating the performance of manifold regularized CMLLR for unsupervised speaker adaptation. Performance is evaluated in terms of ASR word error rate (WER) using the Spanish CallHome and Resource Management (RM) speech corpora. All the HMM training was done using the HTK toolkit [18].

4.1. Spanish CallHome conversational speech corpus

4.1.1. Baseline System

CDHMM models were trained from 16.5 hours of conversational telephone speech. The baseline system uses three-state left-to-right HMM clustered context models, with a total of 1604 states and 16 Gaussian mixtures per state. 13 PLP features with their first and second differences were used. A trigram LM was used with a vocabulary of 45k words.

4.1.2. Adjusting regularization parameters

There are three parameters that need to be adjusted for manifold regularization. Two of these parameters, ρ and K, determine the characteristics of the local neighborhood in the affinity matrix. The third parameter, θ, determines the relative weight of the regularization term in the auxiliary function given by (9). To do so, all 122.35 minutes of test data, collected from 46 speakers, is divided into a development (Dev) set and an evaluation (Eval) set. The Dev and Eval sets contain 34.56 and 87.77 minutes of data, respectively. The best setting for these parameters is determined empirically by performing manifold regularized speaker adaptation on the Dev set. The optimum parameter settings were found to be $\rho = 60$, $\theta = 10$ and $K = 80$.

Table 1. The unsupervised ASR WER on Eval and Dev sets for Spanish CallHome corpus

Set	BL	CMLLR	reg. CMLLR	Frob. Norm
Dev.	69.48	67.92	**67.32**	67.77
Eval.	68.29	67.08	**66.19**	66.57

4.1.3. Regularized CMLLR performance

Table 1 displays WERs obtained for speaker adaptation on Dev and Eval sets. The first column displays the unadapted baseline (BL) WERs. The ASR WERs for CMLLR speaker adaptation without regularization are displayed in the second column of Table 1. CMLLR matrices are estimated for two regression classes representing speech and silence.

The third column of Table 1 displays the WERs for the manifold regularized CMLLR. Manifold regularization is performed by estimating two affinity matrices from the training corpus, one for each regression class, by using (7). The training data consists of 4.16 hours of silence and 12.34 hours of speech. Here, we use 4.16 hours of silence and 4.16 hours of speech[2]. It is clear from the table that manifold regularized CMLLR results in approximately 1% absolute reduction in WER on the Eval set relative to unregularized CMLLR. This difference in WER was found to be statistically significant at a confidence level of less than one percent according to the matched-pairs significance test [19].

In Section 3.2 it was noted that the graph-preserving measure degenerates to shrinkage model adaptation when the affinity matrix is reduced to the form shown in (11). To evaluate the performance of this regularization approach it is only necessary to adjust the value of θ because $\mathbf{R}$ is simply fixed. The optimum value of $\theta = 450$ was empirically determined on the Dev set. The last column of Table 1 displays the WERs obtained for this Frobenius norm based adaptation. The WER on the Eval set is shown to be slightly higher than that obtained for manifold regularized CMLLR. This difference in WER was found to be statistically significant at a confidence level of less than ten percent according to the matched-pairs significance test [19]. This suggests that the information obtained from the intrinsic graph for manifold based regularization has some benefit beyond that derived from constraining the norm of the adaptation parameters.

4.1.4. Speaker adaptive training

A study was performed to evaluate the impact of manifold based regularization for CMLLR adaptation when CDHMM models are trained using speaker adaptive training (SAT) [20]. SAT involves performing CMLLR based speaker normalization as part of CDHMM training, resulting in acoustic models that are inherently more robust with respect to speaker variability. Table 2 displays the ASR results on the Eval set. The first column indicates whether or not manifold regularization

[2]It was observed that using this amount of speech data is a good trade-off between robustness and complexity.

Table 2. The unsupervised ASR WER on Eval set with SAT technique for Spanish CallHome corpus

SAT	Speaker Adaptation	WER
CMLLR	CMLLR	65.48
CMLLR	reg. CMLLR	**65.21**
reg. CMLLR	reg. CMLLR	65.26

is used for CMLLR during SAT training. The second column indicates whether manifold based CMLLR is used for speaker adaptation in decoding, and the last column displays the WER. By comparing the first and second rows, one can conclude that performing SAT without regularization in training and performing regularized CMLLR adaptation in decoding gives slightly lower WER than the case where no manifold regularization is used for CMLLR adaptation. Nevertheless, the WER reduction is very small. This could be due to the fact that using SAT during training produces a CDHMM model with less variance than the original model. This results in less perturbation of the underlying structure of the data when estimating CMLLR transformations. Therefore, the regularization technique has only a minor impact during recognition.

4.2. Resource Management read speech corpus

4.2.1. Baseline System

The RM corpus consists of 3990 utterances. The speech is parameterized using 12 MFCCs, normalized energy and the first and second differences. The baseline system uses three-state left-to-right HMM clustered context models, with a total of 1704 states and 6 Gaussian mixtures per state. No SAT was used during training the baseline models. A 991 word bi-gram language model was used.

4.2.2. Adjusting regularization parameters

All 63.98 minutes of test data, collected from 12 speakers, is divided into Dev and Eval sets. The Dev and Eval sets contain 16.52 and 47.46 minutes of data, respectively. The best setting for these parameters is determined empirically by performing manifold regularized speaker adaptation on the Dev set. The optimum parameter settings were found to be $\rho = 40$, $\theta = 65$ and $K = 70$.

4.2.3. Regularized CMLLR performance

Table 3 displays WERs obtained for speaker adaptation on Dev and Eval sets. The first column displays the BL WERs. The ASR WERs for CMLLR speaker adaptation without regularization are displayed in the second column. The third column displays the WERs for the manifold regularized CMLLR. Manifold regularization is performed by estimating two affinity matrices from the training corpus as was done in Section 4.1. A total of 25.66 minutes of silence and 3.36 hours of speech from the RM corpus are used. The last column displays the WERs obtained when Frobenius norm based regularization is applied to CMLLR adaptation as was done in Section 4.1. The results were obtained for the optimum value of $\theta = 1450$ that was determined empirically on the Dev set.

Table 3. The unsupervised ASR WER on Eval and Dev sets for RM corpus

Set	BL	CMLLR	reg. CMLLR	Frob. Norm
Dev.	5.40	3.60	3.56	**3.49**
Eval.	4.74	3.37	**3.18**	3.23

It is clear that the manifold learning based regularization has only a minor impact on ASR WER for the RM corpus. This marginal improvement may result both from the small size of the corpus and also from the fact that the corpus was collected under well controlled high SNR acoustic conditions and controlled read speech speaking conditions. So the regularization constraints may have less impact on WER. This outcome for a clean condition read speech corpus is consistent with other results reported for regularization of regression based speaker adaptation on similar speech corpora collected under controlled conditions [10, 12].

5. DISCUSSION AND CONCLUSION

Manifold learning has been applied as a regularization technique in the EM based algorithm for CMLLR parameter estimation. This was motivated for CMLLR in particular by the observation that the ML based EM algorithm in general does not necessarily preserve local relationships among feature vectors. Manifold regularization resulted in reduction in ASR WER, especially for the telephone based Spanish Call Home speech corpus.

This approach and the associated experimental study is considered preliminary in that it raises a number of questions. First, the experimental study made use of only small corpora for estimating the affinity matrices associated with the manifold constraints. Future work will address the question of whether semi-supervised learning scenarios involving the use of very large available unlabeled speech corpora for training affinity matrices will have a more substantial impact ASR performance [16]. Second, the development in Section 3 did not consider the use of discriminative manifold learning techniques as has been done for other applications involving manifold constraints [21]. The impact of these techniques will also be evaluated for the manifold regularization application considered in this paper. Finally, The EM algorithm for CMLLR parameter estimation is generally considered to be reasonably well behaved [17]. Future work will investigate the application of manifold based techniques to regularizing other less numerically stable EM based modeling formalisms like, for example, subspace Gaussian mixture model (SGMM) parameter estimation and adaptation [22–27].

6. REFERENCES

[1] A. Sankar and C. H. Lee, "Robust speech recognition based on stochastic matching," in *ICASSP*, 1995, pp. 121–124.

[2] V. V. Digalakis, D. Ritschev, and L.G. Neumeyer, "Speaker adaptation using constrained estimation of Gaussian mixtures," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 5, pp. 357–366, 1995.

[3] 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. 171–186, 1995.

[4] L. R. Neumeyer, A. Sankar, and V. V. Digalakis, "A comprehensive study of speaker adaptation techniques," in *Eurospeech*, 1995, pp. 1127–1130.

[5] M. J. F. Gales and P. C. Woodland, "Mean and variance adaptation within the MLLR framework," *Computer Speech and Language*, vol. 10, pp. 249–264, 1996.

[6] C. J. Leggetter, *Improved Acoustic modelling for HMMs using linear transformations*, Ph.D. thesis, Cambridge University, 1995.

[7] J. Neto et al., "Unsupervised speaker-adaptation for hybrid HMM-MLP continuous speech recognition system," 1995, pp. 187–190, Eurospeech.

[8] 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.

[9] C. H. Lee and J. L. Gauvain, "Speaker adaptation based on MAP estimation of HMM parameters," in *ICASSP*, 1993, vol. 2, pp. 558–561.

[10] C. Chesta, O. Siohan, and C. H. Lee, "Maximum a posteriori linear regression for hidden Markov model adaptation," in *Eurospeech*, 1999.

[11] M. K. Omar, "Regularized feature-based maximum likelihood linear regression for speech recognition," in *Interspeech*, 2007, pp. 1561–1564.

[12] J. Li, Y. Tsao, and C. H. Lee, "Shrinkage model adaptation in automatic speech recognition," in *INTERSPEECH*, 2010, pp. 1656–1659.

[13] Y. Shuicheng et al., "Graph embedding and extensions: a general framework for dimensionality reduction," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 1, pp. 40–51, 2007.

[14] M. Belkin et al., "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *The Journal of Machine Learning Research*, vol. 7, pp. 2399–2434, 2006.

[15] J. Malkin, A. Subramanya, and J. Bilmes, "On the semi-supervised learning of multi-layered perceptrons.," in *Interspeech*, 2009, pp. 660–663.

[16] J. Weston et al., "Deep learning via semi-supervised embedding," in *Neural Networks: Tricks of the Trade*, pp. 639–655. Springer, 2012.

[17] M. J. F. Gales, "Maximum likelihood linear transformations for HMM-based speech recognition," *Computer speech and language*, vol. 12, pp. 75–98, 1997.

[18] S. Young et al., "The HTK book (for HTK version 3.4)," 2006.

[19] L. Gillick and S. Cox, "Some statistical issues in the comparison of speech recognition algorithms," in *ICASSP*. IEEE, 1989, pp. 532–535.

[20] T. Anastasakos et al., "A compact model for speaker adaptive training," in *ICSLP*, 1996, pp. 1137–1140.

[21] V. S. Tomar and R. C. Rose, "Application of a locality preserving discriminant analysis approach to ASR," in *ISSPA*. IEEE, 2012, pp. 103–107.

[22] D. Povey et al., "The subspace Gaussian mixture model–a structured model for speech recognition," *Computer Speech & Language*, vol. 25, no. 2, pp. 404–439, 2011.

[23] S. Hamidi Ghalehjegh and R. C. Rose, "Phonetic subspace adaptation for automatic speech recognition," in *ICASSP*, 2013, pp. 7937 – 7941.

[24] A. Ghoshal et al., "A novel estimation of feature-space MLLR for full-covariance models," in *ICASSP*, 2010, pp. 4310–4313.

[25] L. Lu, A. Ghoshal, and S. Renals, "Regularized subspace Gaussian mixture models for speech recognition," *Signal Processing Letters, IEEE*, vol. 18, no. 7, pp. 419–422, 2011.

[26] A. Mohan, S. Hamidi Ghalehjegh, and R. C. Rose, "Dealing with acoustic mismatch for training multilingual subspace Gaussian mixture models for speech recognition," in *ICASSP*. IEEE, 2012, pp. 4893–4896.

[27] A. Mohan, R. C. Rose, S. Hamidi Ghalehjegh, and S. Umesh, "Acoustic modeling for speech recognition in Indian languages in an agricultural commodities task domain," *Speech Communication*, vol. 56, pp. 167–180, 2014.