MAXIMUM LIKELIHOOD ADAPTATION OF HISTOGRAM EQUALIZATION WITH CONSTRAINT FOR ROBUST SPEECH RECOGNITION

Xiong Xiao[1]*, Jinyu Li*[2]*, Eng Siong Chng*[1]*, Haizhou Li*[3]

[1]School of Computer Engineering, Nanyang Technological University, Singapore
[2]Microsoft Corporation, USA
[3]Department of Human Language Technology, Institute of Infocomm Research, Singapore
xiaoxiong@ntu.edu.sg, jinyli@microsoft.com, aseschng@ntu.edu.sg, hli@i2r.a-star.edu.sg

ABSTRACT

In this paper, we propose a novel feature space adaptation technique to improve the robustness of speech recognition in noisy environments. Histogram equalization (HEQ) is an effective technique for improving robustness by reducing the difference between clean and noisy features. A weakness of HEQ is that it does not take into account acoustic model, resulting in possible mismatch between HEQ-processed features and the acoustic model. In this paper, we propose to adapt HEQ to maximize the likelihood of HEQ-processed features on the acoustic model, with a constraint on the parameters of HEQ. In addition, we use a Gaussian mixture model (GMM) to represent the clean feature space rather than using the acoustic model itself, and this results in both simpler implementation and better results. Experimental results show that HEQ with adaptation reduces word error rate by 7.5% and 5.7% respectively on Aurora-2 and Auroar-4 tasks over the HEQ baseline without adaptation.

Index Terms— robust speech recognition, histogram equalization, maximum likelihood adaptation, feature adaptation, feature normalization.

1. INTRODUCTION

The performance of automatic speech recognition (ASR) degrades significantly when there is mismatch between the training and testing data. For example, if the acoustic model is trained from clean speech features while the test speech is corrupted by additive noise or channel distortion, the performance will be significantly degraded. To improve the robustness of ASR against such environmental distortions, various methods have been proposed. These methods can be grouped into two categories: the feature space methods and the model space methods. The feature space methods aim to reduce the noises' effects by either estimating the clean features from the noisy features (feature compensation [1, 2]), or normalizing both clean and noisy features to make them more similar to each other (feature normalization [3, 4, 5, 6]). The model space methods adapt the clean acoustic model to fit the noisy test data ([7, 8, 9]). In this paper, we will focus on improving a popular feature normalization technique, i.e. histogram equalization (HEQ) [5, 10].

HEQ reduces noise effects by normalizing the histograms of the speech features to predefined reference histograms, e.g. histograms of clean features. When speech is corrupted by noise, the histograms of speech features are also changed. By normalizing the histograms of corrupted features to the histogram of clean features, we hope to reduce the noise effects. Despite its simplicity, HEQ is found to be very effective in improving ASR robustness [5, 10].

One weakness of HEQ is that it does not consider the information in the acoustic model. As a result, features processed by HEQ may not fit the acoustic model well. In this paper, we address this issue by adapting HEQ parameters to maximize the likelihood of the HEQ-processed test features on the acoustic model. As the HEQ adaptation studied here is a pure feature space adaptation, if there is no constraint on the HEQ parameters, the adapted HEQ will map the feature vectors towards the mean vectors of the acoustic model. To prevent this, we add a constraint to reduce the flexibility of HEQ. We also examine the use of simple Gaussian mixture models (GMM) as our target model instead of using the complex hidden Markov models (HMM) based acoustic model.

The rest of the paper is organized as follows. In section 2, the adaptation of HEQ with constraint is described. In section 3, the effectiveness of HEQ adaptation is evaluated on speech recognition tasks. Finally, conclusion is presented in section 4.

2. ADAPTATION OF HISTOGRAM EQUALIZATION

In this section, we will first represent HEQ in a parametric form to facilitate its adaptation. Then we will present the adaptation of HEQ using the maximum likelihood (ML) criterion with constraints on HEQ parameters. Finally, we will discuss some implementation issues.

2.1. Parametric Representation of HEQ

Let x_t^k be the input feature at frame t and dimension k with $t = 1, ..., T$ and $k = 1, ..., K$. T is the number of frames in an utterance and K is the number of feature dimensions, respectively. The HEQ-processed version of x_t^k is obtained by [5]:

$$y_t^k = C_{\text{ref},k}^{-1}(C_{x,k}(x_t^k)),\ k = 1, ..., K \tag{1}$$

where $C_{\text{ref},k}^{-1}(\cdot)$ is the inverse reference cumulative distribution function (CDF) and $C_{x,k}(\cdot)$ is the CDF of x_t^k, both for dimension k. As HEQ processes each dimension independently, we will use one dimension for illustration and drop the dimension index for simplicity.

To implement (1), $C_x(\cdot)$ can be estimated from the rank of x_t among $t = 1, ..., T$ [11]:

$$C_x(x_t) = (R(x_t) - 0.5)/T \tag{2}$$

where $R(x_t) \in [1, T]$ is the rank of x_t. For $C_{\text{ref}}^{-1}(\cdot)$, we adopt a parametric approximation similar to the polynomial regression used in [10] to facilitate HEQ adaptation. In our initial experiments, we fount that using sigmoid functions to approximate $C_{\text{ref}}^{-1}(\cdot)$ produces

slightly better results than using polynomial regression, hence we adopt sigmoid functions in this paper as follows:

$$C_{\text{ref}}^{-1}(C_x(x_t)) \approx \sum_{m=1}^{M} a_m \text{sig}_m(C_x(x_t)) + a_0 \quad (3)$$

where $\text{sig}_m(x) = [1 + \exp(-\gamma(x - \theta_m))]^{-1}$ is the m^{th} sigmoid function centered at θ_m, M is the number of sigmoid functions, γ controls the slope of all the sigmoid functions, and a_0 is an offset parameter. γ and θ_m are predefined in our study and treated as constants in HEQ. γ is chosen such that the approximated HEQ transformation is smooth and flexible, and θ_m can be evenly spaced points in the range of CDF function, i.e. [0,1].

Substitute (3) into (1), the processed feature is rewritten as

$$y_t = C_{\text{ref}}^{-1}(C_x(x_t)) \approx \mathbf{a}^T \mathbf{z}_t \quad (4)$$

where $\mathbf{a} = [a_0, a_1, ..., a_m]^T$ is a vector of HEQ parameters, $\mathbf{z}_t = [1, \text{sig}_1(C_x(x_t)), ..., \text{sig}_M(C_x(x_t))]^T$ is a vector of order statistics, and $\cdot^T$ represents matrix or vector transpose. The parametric approximation of HEQ can be seen as a linear transform of $\mathbf{z}_t$, while $\mathbf{z}_t$ is computed from the original features in a nonlinear way.

2.2. Estimation of HEQ parameters Using MMSE Criterion

Given a clean training database, we can train the HEQ parameters by minimizing the mean square error (MSE) between the clean features and their HEQ-processed versions. Let x_t denote the clean training feature of frame t. The minimum mean square error (MMSE) estimate of $\mathbf{a}$ can be approximated by the least square estimate [10]:

$$\hat{\mathbf{a}}_{\text{MMSE}} \approx \arg\min_{\mathbf{a}} \frac{1}{T} \sum_{t=1}^{T} [(x_t - \mathbf{a}^T \mathbf{z}_t)^2] \quad (5)$$

$$= \hat{E}[\mathbf{z}_t \mathbf{z}_t^T]^{-1} \hat{E}[\mathbf{z}_t x_t] \quad (6)$$

where $\hat{E}[\mathbf{z}_t \mathbf{z}_t^T]$ is the estimated auto-correlation matrix of $\mathbf{z}_t$ and $\hat{E}[\mathbf{z}_t x_t]$ is the cross correlation estimate. We denote the MMSE estimate of HEQ parameters as HEQ-MMSE.

If clean features are used to train HEQ parameters, the trained HEQ will normalize the histogram of incoming features to that of clean features. An alternative reference histogram is a predefined probability density function (p.d.f.), e.g. the Gaussian distribution [11]. To use Gaussian as reference, we need only replace the training data x_t with samples drawn from a zero-mean, unit-variance Gaussian distribution. The trained HEQ does not depend on any speech data and can be used for all feature dimensions and all databases. In this paper, Gaussian distribution is used as the reference histogram.

2.3. Adaptation of HEQ parameters Using ML Criterion

As the MMSE estimate of HEQ parameters does not consider the acoustic model, the processed features may have a poor fit with the acoustic model. This problem can be alleviated by maximizing the likelihood of the processed features on the acoustic model:

$$\hat{\mathbf{a}}_{\text{ML}}^k = \arg\max_{\mathbf{a}^k} \log p(\mathbf{Y}|\Lambda), \text{for } k = 1, ..., K \quad (7)$$

where $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_T]$, $\mathbf{y}_t = [y_t^1, ..., y_t^K]$ is the sequence of HEQ-processed feature vectors for a test utterance and Λ represents the acoustic model. As HEQ performs independently for each feature dimensions, there are K independent optimization problem in (7).

We adopt the Expectation-Maximization (EM) framework to solve the adaptation problem. As the state transition probabilities are not related to our problem at hand, we can use the following simplified auxiliary function for dimension k:

$$Q(\mathbf{a}^k, \hat{\mathbf{a}}^k) = \sum_{t=1}^{T} \sum_{sm} \gamma_{sm}(t) \log p(\mathbf{y}_t|s, m, \Lambda) \quad (8)$$

where $\gamma_{sm}(t)$ is the occupation probability of mixture m of state s at frame t, $\hat{\mathbf{a}}^k$ is the current estimate of HEQ parameters and $\mathbf{a}^k$ is the parameters to be estimated. Take the partial differentiation of $Q(\mathbf{a}^k, \hat{\mathbf{a}}^k)$ w.r.t. $\mathbf{a}^k$ and use the chain rule, we get:

$$\frac{\partial Q(\mathbf{a}^k, \hat{\mathbf{a}}^k)}{\partial \mathbf{a}^k} = \sum_{t,s,m} \gamma_{sm}(t) \frac{\partial \log p(\mathbf{y}_t|s, m, \Lambda)}{\partial y_t^k} \frac{\partial y_t^k}{\partial \mathbf{a}^k} \quad (9)$$

If $p(\mathbf{y}_t|s, m, \Lambda)$ is a multivariate Gaussian with diagonal covariance matrix, we have

$$\partial \log p(\mathbf{y}_t|s, m, \Lambda)/\partial y_t^k = (\mu_{sm}^k - y_t^k)/(\sigma_{sm}^k)^2 \quad (10)$$

From (4), it is obvious that $\partial y_t^k / \partial \mathbf{a}^k = \mathbf{z}_t^k$. Substitute this equation and (10) into (9) and make it equal to zero, we get

$$\frac{\partial Q(\mathbf{a}^k, \hat{\mathbf{a}}^k)}{\partial \mathbf{a}^k} = \sum_{t,s,m} \gamma_{sm}(t) \frac{\mu_{sm}^k - y_t^k}{(\sigma_{sm}^k)^2} \mathbf{z}_t^k = 0 \quad (11)$$

Hence, the close-form solution of $\mathbf{a}^k$ is

$$\hat{\mathbf{a}}_{\text{ML}}^k = \mathbf{A}_k^{-1} \mathbf{c}_k \quad (12)$$

$$\text{where } \mathbf{c}_k = \sum_{t,s,m} \frac{\gamma_{sm}(t)}{(\sigma_{sm}^k)^2} \mu_{sm}^k \mathbf{z}_t^k \quad (13)$$

$$\mathbf{A}_k = \sum_{t,s,m} \frac{\gamma_{sm}(t)}{(\sigma_{sm}^k)^2} \mathbf{z}_t^k \mathbf{z}_t^{kT} \quad (14)$$

Note that $\mathbf{c}_k$ is similar to the cross-correlation between μ_{sm}^k and $\mathbf{z}_t^k$ and $\mathbf{A}_k$ is similar to autocorrelation matrix of $\mathbf{z}_t^k$, but both weighted by occupation probabilities and variances. It is obvious that $\mathbf{A}_k$ is positive definite and can be inverted.

2.4. ML adaptation with transformation constraints

The ML solution of HEQ tends to map the feature vectors to the mean vectors of the acoustic model specified by the occupation probabilities. This is because for a Gaussian distribution, maximum likelihood is obtained if the feature vector is equal to the mean vector. As a result, ML adaptation of HEQ will reduce the variances and discriminative power of the adapted features significantly. Hence, the ML solution is not suitable to be used alone for pure feature space adaptation like HEQ adaptation. Constraint needs to be imposed to make the ML solution suitable for speech recognition.

Constraint can be added to control the degree of difference between the initial features (i.e. HEQ-MMSE-processed features) and their adapted version. The adapted features are expected to be different from the original features such that the likelihood can be improved. However, they are not expected to be too different from the original features as this will cause a new type of mismatch. Therefore, if we use HEQ-MMSE as the starting point of HEQ adaptation, we can impose a constraint such that the search space of the HEQ parameters will be near to the MMSE solution. This will ensure that

the adapted features won't be too far away from the original HEQ-MMSE processed features.

Two types of constraints may be applied. The first is to add a regularization term $||\mathbf{a}^k - \mathbf{a}^k_{\text{MMSE}}||^2$ in the ML objective function, which directly constrains the values of HEQ parameters. The second constraint is that the adapted HEQ transformation (not the parameters themselves) should be near to the transformation of HEQ-MMSE. This means that given a $\mathbf{z}$, the processed feature using ML solution, i.e. $\mathbf{z}^T\hat{\mathbf{a}}^k_{\text{ML}}$, should be close to that using MMSE solution $\mathbf{z}^T\hat{\mathbf{a}}^k_{\text{MMSE}}$. Let $\mathbf{W} = [\mathbf{z}_1, ..., \mathbf{z}_S]$ be the matrix of S selected $\mathbf{z}$. We can add a constraint as follows:

$$\hat{\mathbf{a}}^k_{\text{ML}} = \arg\max_{\mathbf{a}^k} \log p(\mathbf{Y}|\Lambda) - \alpha T||\mathbf{W}^T\mathbf{a}^k - \mathbf{W}^T\mathbf{a}^k_{\text{MMSE}}||^2 \quad \text{for } k = 1, ..., D \tag{15}$$

where T is the number of frames in the test utterance and $\alpha > 0$ is used to control the weight of the two conflicting terms in the objective function. A bigger α will make it more difficult for the adapted HEQ transformation to deviate from the MMSE transformation. The vectors $\mathbf{z}$ can be selected to be representative, e.g. evenly from the CDF input space [0,1]. For example, if we have 5 $\mathbf{z}$ vectors, they can be chosen as $\mathbf{z}_1 = [1, \text{sig}_1(0), ..., \text{sig}_M(0)]^T$, $\mathbf{z}_2 = [1, \text{sig}_1(0.25), ..., \text{sig}_M(0.25)]^T$, ..., and $\mathbf{z}_5 = [1, \text{sig}_1(1), ..., \text{sig}_M(1)]^T$.

The close-form ML solution with constraint is still $\hat{\mathbf{a}}^k_{\text{ML}} = \mathbf{A}_k^{-1}\mathbf{c}_k$, with $\mathbf{c}_k$ and $\mathbf{A}_k$ changed to:

$$\mathbf{c}_k = \sum_{t,s,m} \frac{\gamma_{sm}(t)}{(\sigma^k_{sm})^2}\mu^k_{sm}\mathbf{z}^k_t + 2\alpha T\mathbf{W}\mathbf{W}^T\mathbf{a}^k_{\text{MMSE}}$$

$$\mathbf{A}_k = \sum_{t,s,m} \frac{\gamma_{sm}(t)}{(\sigma^k_{sm})^2}\mathbf{z}^k_t\mathbf{z}^{k\,T}_t + 2\alpha T\mathbf{W}\mathbf{W}^T \tag{16}$$

In our study, we find that the constraint on transformation in (15) produces better performance than adding the regularization term $||\mathbf{a}^k - \mathbf{a}^k_{\text{MMSE}}||^2$. Hence, in this study, we will only use the solution in (16) in experimental studies and denote it as HEQ-ML.

2.5. Implementation Issues

A two-pass decoding strategy is necessary to implement HEQ adaptation. In the first pass, the most likely state sequences are obtained by decoding the HEQ-MMSE-processed test features. The mixture occupation probabilities $\gamma_{sm}(t)$ can be obtained from the state sequences. Then HEQ-ML can be computed using the close-form solution in (16). Only one iteration of EM is used in our study and HEQ-MMSE is used as the initial estimate of HEQ-ML. In the second pass, the final recognition output is obtained by decoding the HEQ-ML-processed test features with the acoustic model.

To avoid the first pass decoding, we use a simpler model as the target model for HEQ adaptation, e.g. Gaussian mixture models (GMM). If the clean acoustic space is represented by a GMM, the posterior probability of the GMM mixtures can be computed without using Viterbi decoding. Hence, the HEQ adaptation becomes a pure feature space technique and easy to be used in most speech recognition systems. In [12], similar approach was used in the scenario of constrained maximum likelihood linear regression (CMLLR).

3. EXPERIMENTS

3.1. Experimental Settings

The HEQ adaptation is evaluated on the Aurora-2 [13] and Aurora-4 tasks [14]. The acoustic model of Aurora-2 task follows the standard

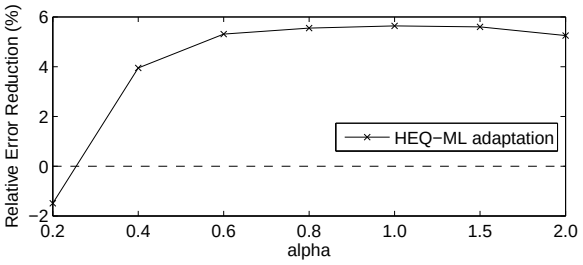

Fig. 1. Relative WER reduction achieved by HEQ-ML over HEQ-MMSE with different α in (15).

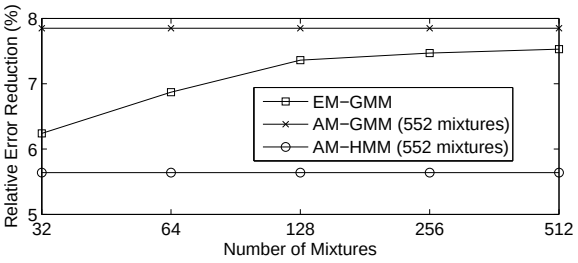

Fig. 2. Relative WER reduction achieved by HEQ-ML over HEQ-MMSE with different number of mixtures in target GMM.

configuration in [13]. For Aurora-4, a triphone-based acoustic model is used, with 2800 shared states and 8 mixtures per state. A decision tree is used for generating the shared states. Bigram langauge model is used for recognition. Both Aurora-2 and Aurora-4 acoustic models are trained from clean speech data.

Mel-frequency cepstral coefficients (MFCC) are extracted by the feature extraction program WI007 delivered with the Aurroa-2 [13]. MFCC, together with their delta and accelerations, are used as the features for acoustic modeling. $c0$ energy is used and log energy is not. When applied, HEQ-MMSE and HEQ-ML are applied to all the 39 feature dimensions independently.

For parametric representation of HEQ, we use 11 sigmoid functions evenly spaced in the interval [0,1], i.e. $\theta_1 = 0$, $\theta_2 = 0.1$, $\theta_3 = 0.2$,..., and $\theta_{11} = 1$. The γ in (3) is chosen to 30 to ensure smoothness of the approximated HEQ transformation function. When HMMs are used as the target model for HEQ adaptation, only the best state sequence is used to compute the occupation probabilities of mixtures. There are totally 11 constraint vectors used in (15), and their locations are the same as the values of θ described above.

3.2. Tuning of α

The parameter α controls the weight of constraint in HEQ adaptation. When $\alpha = 0$, pure ML solution is obtained, and when $\alpha = \infty$, pure MMSE solution is used. The relative word error rate (WER) reduction achieved by HEQ-ML over HEQ-MMSE on Aurora-2 with various selection of α is plotted in Fig. 1. HMMs are used as the target models for HEQ adaptation. From the figure, it is observed that the performance of HEQ-ML is quite stable near $\alpha = 1$, and more than 5% improvement can be achieved. Therefore, in the following experiments, we fix α to be 1 for all cases.

Although the best state sequence generated from the first-pass decoding contains a lot of errors, especially for low SNR levels, HEQ-ML is still able to improve the performance. This shows that

Table 2. Recognition WER (%) on AURORA-4 task. Avg. represents the average results over all test cases.

Test Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Avg.
HEQ-MMSE	13.33	21.33	34.33	35.84	36.24	34.77	37.57	20.77	31.23	42.43	46.74	51.16	44.27	46.45	35.46
HEQ-ML	12.60	19.82	32.23	34.22	34.11	32.74	36.28	19.82	29.94	40.59	44.49	47.18	40.99	43.13	33.44
R.R.	5.5	7.1	6.1	4.5	5.9	5.8	3.4	4.6	4.1	4.3	4.8	7.8	7.4	7.1	**5.7**

Table 1. Recognition WER on Aurora-2 task in different SNR levels. ∞ represents clean test cases and 0-20 represents average WER from 0dB to 20dB. R.R represents relative WER reduction achieved by HEQ-ML over HEQ-MMSE.

SNR (dB)	∞	20	15	10	5	0	-5	0-20
HEQ-MMSE	0.99	2.30	4.38	9.50	22.97	51.05	80.68	18.04
HEQ-ML	0.96	2.12	4.03	8.86	21.26	47.14	77.61	16.68
R.R.	3.5	7.7	8.2	6.7	7.5	7.6	3.8	**7.5**

instead of being guided by the errors in the state sequence, HEQ-ML captures and compensates the environmental differences between the test features and the acoustic model. This should be due to the constraint in (15) which significantly limits the flexibility of HEQ.

3.3. Results with GMM-based Target Model

Rather than using HMMs as the target model, we can also use a simple GMM as the target model. In Fig. 2, the relative WER reduction of HEQ-ML over HEQ-MMSE is shown with different kinds of target model. The EM-GMM curve in the figure is obtained by using EM-trained GMMs with different number of mixtures as the target model. The AM-HMM curve is obtained when HMMs are used as the target model, i.e. the results presented in the previous section. There are totally 552 mixtures in the HMMs. The AM-GMM curve is the result with a GMM created by pooling the mixtures of the HMMs. From the figure, it is observed that, despite their simplicity, EM-GMM and AM-GMM both performs better than AM-HMM. This may be due to the fact that we only use one best state path in AM-HMM. It is also observed that when there are 128 mixtures in the GMM, EM-GMM's performance becomes stable and is near to that of AM-GMM. The best results with EM-GMM is about 7.5% relative WER reduction over HMM-MMSE. The detailed results of EM-GMM (with 512 mixtures) are compared with that of HEQ-MMSE for every signal-to-noise ratio (SNR) in Table 1. It is observed that the HEQ-ML reduces WER in all SNR levels.

We also evaluate HEQ-ML on the Aurora-4 task. A GMM with 512 mixtures are used as the target model for HEQ-ML. The recognition WER is shown in Table 2. From the table, it is observed that HEQ-ML reduces WER consistently for all the 14 test cases. The average reduction of WER is 5.7%.

4. CONCLUSIONS

In this paper, we propose to estimate HEQ parameters by maximizing the likelihood of the test features on the acoustic model. Constraint is imposed to prevent HEQ transformation to deviate too much from the initial transformation. Experimental results on Aurora-2 and Aurora-4 tasks show that the adapted HEQ consistently outperforms the original HEQ in all test cases when the acoustic model is trained from clean features. The proposed adaptation scheme may be extended to multi-class HEQ in the future, where one HEQ transformation is used for each acoustic class.

5. REFERENCES

[1] L. Buera, E. Lleida, A. Miguel, A. Ortega, and O. Saz, "Cepstral vector normalization based on stereo data for robust speech recognition," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1098–1113, March 2007.

[2] L. Deng, J. Droppo, and A. Acero, "Enhancement of log mel power spectra of speech using a phase-sensitive model of the acoustic environment and sequential estimation of the corrupting noise," *IEEE Trans. Speech and Audio Processing*, vol. 12, no. 2, pp. 133–143, Mar. 2004.

[3] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.

[4] O. Viikki and K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition," *Speech Communication*, vol. 25, pp. 133–147, 1998.

[5] A. de la Torre, A. M. Peinado, J. C. Segura, J. L. Perez-Cordoba, M. C. Benitez, and A. J. Rubio, "Histogram equalization of speech representation for robust speech recognition," *IEEE Trans. Speech and Audio Processing*, vol. 13, no. 3, pp. 355–366, 2005.

[6] X. Xiao, E. S. Chng, and H. Li, "Normalization of the speech modulation spectra for robust speech recognition," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1662–1674, Nov. 2008.

[7] 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, Apr. 1995.

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

[9] J. Li, L. Deng, D. Yu, Y. Gong, and A. Acero, "High-performance HMM adaptation with joint compensation of additive and convolutive distortions via vector taylor series," in *Proc. ASRU '07*, Kyoto, Japan, Dec. 2007, pp. 65–70.

[10] S.-H. Lin, B. Chen, and Y.-M. Yeh, "Exploring the use of speech features and their corresponding distribution characteristics for robust speech recognition," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 84 –94, jan. 2009.

[11] J. C. Segura, Carmen Benítez, A. de la Torre, A. J. Rubio, and Javier Ramírez, "Cepstral domain segmental nonlinear feature transformations for robust speech recognition," *IEEE Signal Processing letters*, vol. 11, no. 5, pp. 517–520, 2004.

[12] G. Stemmer, F. Brugnara, and D. Giuliani, "Adaptive training using simple target models," in *Proc. ICASSP '05*, Philadelphia, USA, Mar. 2005, vol. I, pp. 997–1000.

[13] D. Pearce and H.-G. Hirsch, "The AURORA experimental framework for the performance evaluation of speech recogntion systems under noisy conditions," in *Proc. ICSLP '00*, Beijing, China, Oct. 2000, vol. 4, pp. 29–32.

[14] N. Parihar and J. Picone, "Aurora working group: DSR front end LVCSR evaluation AU/384/02," Tech. Rep., Institute for Signal and Infomation Processing, Mississippi State Univ., MS, Dec. 2002.