SEQUENTIAL MAP ESTIMATION BASED SPEECH FEATURE ENHANCEMENT FOR NOISE ROBUST SPEECH RECOGNITION

Chuan Jia, Peng Ding and Bo Xu

High Technology and Innovation Center
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100080
{*cjia*, *pding, xubo*}@hitic.ia.ac.cn

ABSTRACT

In this paper, the environment mismatch due to additive noise is assumed as an additive bias in power spectral domain. It is viable to introduce some constraints on the values that the bias can take due to the internal relation between bias and noise power spectrum. We propose to introduce the noise priori knowledge into bias estimation process by using maximum a *posteriori* (MAP) criterion. Moreover, the mismatch is usually non-stationary in real application and sequential algorithm can be used to track time varying environment within a test utterance. This paper proposes to use the sequential techniques to estimate the bias in the MAP framework and update the parameters of noise priori adaptively. Speech recognition experiments demonstrated that the proposed algorithm outperformed sequential ML estimation method and was obviously better than the batch mode under non-stationary noise environment.

1. INTRODUCTION

In recent years, there has been much interest in the problem of robustness of automatic speech recognition. When there is a mismatch between the training and testing environment, the performance of ASR deteriorates. These sources of mismatch include additive noise, channel and transducer mismatch, speaker mismatch, etc. [1]. Recently, many methods have been proposed to deal with the mismatch, which can be classified as signal-space, feature-space and model-space compensation.

Stochastic matching method [1] may reduce the mismatch by mapping the distorted features to an estimate of the original features or mapping the original models to the transformed models. No addition adaptation data is required for the estimation of the mismatch except the given test utterance and the given speech models. In [2], the mismatch due to additive noise is modeled by an affine transformation and a bias in the cepstral domain and applied stochastic matching method to estimate these parameters. However, it is not accurate enough because of the nonlinearity relationship between clean speech cepstrum and corrupted speech cepstrum. Contrary to the nonlinearity in cepstral domain, there exist a linear relation between clean speech power spectra and corrupted speech power spectra. Considering the linearity, we model the mismatch as an additive bias and extend the stochastic matching method in spectral domain to avoid the nonlinear function mapping and the focus is feature enhancement.

Maximum likelihood (ML) criterion is usually chosen to estimate parameters such as additive bias in classic stochastic matching algorithm [1,2]. However, maximum likelihood criterion does not introduce any constraints on the possible values of additive bias and relies only on the test utterance and the original speech models. In this paper, based on the physical consideration regarding the additive bias in spectral domain, the noise distribution is viewed as a constraint for bias estimation and this constraint is incorporated into bias estimation process acting as priori knowledge by using MAP criterion.

In implementation, the batch mode is relatively valid when dealing with stationary noise environment, which assume the bias to be constant during the whole utterance. However, the parameters of interest are sometimes subject to changes and they are time varying frequently in real noise environment. In such cases, a sequential algorithm [3,4] can be designed to adaptively track the varying parameters. The algorithm proposed in this paper applies such techniques to sequentially estimate the time varying spectral bias. The non-stationary noise environment also results in the requirement of updating the parameters of noise priori distribution sequentially. Thus in this paper, two sequential estimation processes are implemented in turn. Furthermore, sequential techniques are helpful in computational efficiency and storage requirements.

The remainder of this paper is organized as follows. In next part, we will formulate the problem. The proposed algorithm is described in the third part, which includes the sequential MAP estimation for spectral bias and the sequential estimation for parameters of noise distribution. The experiments and results are presented in the fourth part. Finally, we will draw a conclusion.

2. PROBLEM FORMULATION

In situations where there is additive noise, the corrupted speech is described by the mismatch function

$$Y_t(k) \cong X_t(k) + N_t(k) \tag{1}$$

where the additive component, $N_t(k)$, is the noise power spectrum. $X_t(k)$ and $Y_t(k)$ represent the power spectra of clean speech and corrupted speech, respectively.

In the feature space, inverse distortion function $f_b(Y)$ maps the corrupted speech features Y into original speech features X where b is the parameter of the function. From (1), the inverse distortion function in power spectral domain becomes

$$X_t = f_b(Y_t) = Y_t - b_t \tag{2}$$

Thus the mismatch is modeled by the above additive spectral bias b. Compared with that in cepstral domain, the compensation algorithm in spectral domain possesses the simpler and more accurate inverse distortion function like (2).

The bias b is usually estimated by using ML criterion because its simplicity:

$$b' = \arg\max_b p(Y \mid b, \Lambda_X) \tag{3}$$

where Y is adaptation data, Λ_X is the set of speech models.

According to (1) and (2), the bias can be interpreted as a noise power spectrum. However, (3) does not consider any priori knowledge introduced by the noise power spectrum. In order to overcome such shortcoming, we assume that the bias is random rather than fixed for each utterance. Then it can be describe by its probability density function (pdf) $p(b)$, called noise priori distribution. This pdf represents the constraint about the value that the bias can take. The priori knowledge of noise spectra is then incorporated in the estimation process using MAP criterion:

$$\begin{aligned} b &= \arg\max_b p(b \mid Y, \Lambda_X) \\ &\propto \arg\max_b p(Y \mid b, \Lambda_X)\cdot p(b) \end{aligned} \quad (4)$$

However, the batch mode described by (4) is not adequate for tracking the non-stationary noise environment. Thus, we assume the bias b to be time varying and be sequentially estimated. In this paper, the noise priori information are integrated into maximum a *posteriori* (MAP) framework in which the sequential techniques are used to estimate the spectral bias. Moreover, besides estimating the bias sequentially, the parameters of noise priori are updated upon the presentation of the latest data from test utterance. The proposed algorithm is described in detail in the next part.

3. SEQUENTIAL MAP ESTIMATION FOR SPECTRAL FEATURE ENHANCEMENT

The whole process includes two optimization processes shown in Fig. I. Firstly, given the estimated parameters of noise priori at the previous time, estimate the parameters of noise priori at the present time. Then based on the estimated parameters of noise priori at the present time and the estimated bias at the previous time, estimate the spectral bias at the present time. Repeat the steps until the end of the test utterance.

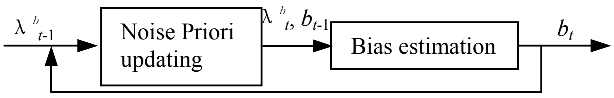

Figure I. The framework of the proposed algorithm

In what follows, due to the assumed diagonal form of covariance matrix, we present algorithms in scalar form and drop the dependence on the dimension, without loss of generality.

Next, we will firstly introduce the bias estimation algorithm and then the noise priori updating algorithm.

3.1. Sequential estimation for spectral bias in MAP criterion

The noise priori can be formally inserted into the estimation process by using MAP criterion to derive bias b sequentially as follows

$$\begin{aligned} b_{t+1} &= \arg\max_b p(b \mid Y_{t+1}, \Lambda_X, B_t, \lambda^b_{t+1}) \\ &\propto \arg\max_b p(Y_{t+1} \mid b, \Lambda_X, B_t)\cdot p(b \mid \lambda^b_{t+1}) \end{aligned} \quad (5)$$

where Y_t represents the sequences of observation $\{y_1, y_2, \ldots y_t\}$, B_t denotes the sequence of estimated bias $\{b_1, b_2, \ldots b_t\}$ and λ^b_{t+1} is the parameters of noise priori updated at time t+1. The objective function above is optimized indirectly using the EM algorithm. In the E-step, the MAP auxiliary function can be simplified as

$$\begin{aligned} Q_{MAP,t+1}(b|B_t) &= E\{\log p(Y_{t+1}, S_{t+1}, C_{t+1} \mid b, \Lambda_X) + \eta\cdot \log p(b \mid \lambda^b_{t+1}) \mid B_t, \Lambda_X, Y_{t+1}\} \\ &= Q_{ML,t+1}(b \mid B_t) + \eta \cdot \log p(b \mid \lambda^b_{t+1}) \end{aligned} \quad (6)$$

where $S_{t+1}=\{s_1, s_2, \ldots s_{t+1}\}$ be the sequence of state indices, $C_{t+1}=\{c_1, c_2, \ldots c_{t+1}\}$ be the sequence of indices of mixture components, η is a weight to control the importance of the noise priori knowledge, and

$$Q_{ML,t+1}(b \mid B_t) = \sum_{\tau=1}^{t+1} \xi^{t+1-\tau} \sum_{n=1}^{N}\sum_{m=1}^{M} \gamma_{\tau|t+1,B_{\tau-1}}(n,m)\{-\frac{(y_\tau - b - \mu^x_{n,m})^2}{2\Sigma^x_{n,m}}\} \quad (7)$$

In (7), ξ is the forgetting factor, which is to reduce the effect of past data to the new input data, and $\gamma_{\tau|t+1,B}(n,m) = p(s_\tau=n, c_\tau=m|Y_{t+1}, B_{\tau-1}, \Lambda_X)$. The maximization of $Q_{ML,t+1}(b|B_t)$ with respect to b can result in the sequential estimation of the spectral bias under ML rule.

To carry out the M-step, we can use second-order Taylor series expansion and the Newton-Raphson technique [3] to sequentially estimate the bias via the following recursive form

$$b_{t+1} = b_t + I^{-1}_{t+1}(b_t)\cdot \frac{\partial Q_{MAP,t+1}(b \mid B_t)}{\partial b}\Big|_{b=b_t} \quad (8)$$

where $I_{t+1}(b_t) = -\partial^2 Q_{MAP,t+1}(b \mid B_t)/\partial b^2 \,|_{b=b_t}$.

3.2. Noise priori selection

An important element in the maximization of MAP auxiliary function in (6) is the choice of the noise priori. Care must be taken to see that the priori reflects the variability in the spectral bias. The form of priori can be properly chosen based on some physical considerations or on some mathematical attractiveness. According to (8), the different assumption of noise priori will lead to different estimation form.

If it is assumed that the noise is to be normally distributed in the log-power spectral domain with mean M and variance S^2, e.g. $\lambda^b_{t+1}=\{M, S^2\}$, the distribution of noise is modeled by log-normal distribution in power spectral domain [5], expressed as

$$p(b \mid \lambda^b_{t+1}) = \frac{1}{\sqrt{2\pi}Sb}\exp\{-\frac{(\ln b - M)^2}{2S^2}\} \quad (9)$$

Take (6) and (9) into (8) and thus the corresponding estimation form is as follows

$$b_{t+1} = \frac{S_{t+1} + \eta\cdot\{\frac{1-2(\ln b_t - M)}{S^2} - 2\}/b_t}{F_{t+1} + \eta\cdot\{\frac{1-(\ln b_t - M)}{S^2} - 1\}/b_t^2} \quad (10)$$

where

$$S_{t+1} = \sum_{\tau=1}^{t+1}\xi^{t+1-\tau}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_{\tau|t+1,B_{\tau-1}}\frac{(y_\tau - \mu^x_{n,m})}{\Sigma^x_{n,m}} \quad (11)$$

and

$$F_{t+1} = \sum_{\tau=1}^{t+1}\xi^{t+1-\tau}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_{\tau|t+1,B_{\tau-1}}\frac{1}{\Sigma^x_{n,m}} \quad (12)$$

The S_{t+1} and F_{t+1} can be obtained by recursive computation based on their previous values at time t.

Alternatively, if we assume that the Fourier expansion coefficients of noise are Gaussian distribution, the power spectrum of noise can be considered as exponential distribution [6] as follows

$$p(b) = \frac{1}{\nu^b}\exp(-\frac{b}{\nu^b}) \quad (13)$$

The corresponding estimation form is

$$b_{t+1} = \frac{S_{t+1} - \eta\cdot(1/\nu^b)}{F_{t+1}} \quad (14)$$

where the computation of S_{t+1} and F_{t+1} is the same as in (11-12).

3.3. Noise priori updating

The parameters of noise priori are commonly subject to changes especially in non-stationary noise environment. We proposed to update the mean and variance of noise power spectrum in turn by using stochastic approximations algorithm proposed in [3,4].

After obtaining the statistics of noise power spectrum at the preset time, we can update the parameters of noise priori by using these statistics. Then the updated noise priori is used to estimate the spectral bias at the present time.

3.3.1. Noise mean updating

Let μ_0^b be the initial estimate of noise mean and $\Lambda_t^b=\{\mu_1^b,\mu_2^b,\ldots\mu_t^b\}$ be the sequence of the estimate of noise mean up to time t. Then μ_{t+1}^b is sought by

$$\mu_{t+1}^b = \arg\max_{\mu} Q_{t+1}(\mu^b \mid \Lambda_X, \Lambda_t^b) \tag{15}$$

where

$$Q_{t+1}(\mu^b \mid \Lambda_X, \Lambda_t^b) = E\{\log p(Y_{t+1}, S_{t+1}, C_{t+1} \mid \Lambda_X, \mu^b) \mid \Lambda_X, Y_{t+1}, \Lambda_t^b\} \tag{16}$$

In the E-step, the objective function in (16) is simplified to

$$Q_{t+1}(\mu^b \mid \Lambda_X, \Lambda_t^b) = \sum_{\tau=1}^{t+1}\sum_{n=1}^{N}\sum_{m=1}^{M} p(s_\tau = n, c_\tau = m \mid Y_{t+1}, \Lambda_X, \Lambda_{\tau-1}^b)\cdot \log p(y_\tau \mid n, m, \mu^b) \tag{17}$$

Due to the Gaussian distribution assumption of the power spectra of original speech and the corrupted speech and the relation shown in (1), the probability of corrupted speech given the state n and mixture component m in (17) is

$$p(y_\tau \mid n, m, \mu^b) = N(y_\tau; \mu_{n,m}^x + \mu_\tau^b, \Sigma_{n,m}^x + \Sigma_\tau^b) \tag{18}$$

In the M-step, using the stochastic approximate algorithm [3], sequential mean estimation can be obtained by

$$\mu_{t+1}^b = \mu_t^b + I_{t+1}^{-1}(\mu_t^b)\cdot S(\mu_t^b, y_{t+1}) \tag{19}$$

where the Fisher information item $I_{t+1}(\mu_t^b)$ and the score item $S(\mu_t^b, y_{t+1})$ are given as follows

$$I_{t+1}(\mu_t^b) = \sum_{\tau=1}^{t+1}\sum_{n=1}^{N}\sum_{m=1}^{M} p(n, m \mid y_\tau, \Lambda_{\tau-1}^b)\cdot(\Sigma_{n,m}^x + \Sigma_\tau^b) \tag{20}$$

$$S(\mu_t^b, y_{t+1}) = \sum_{n=1}^{N}\sum_{m=1}^{M} p(n, m \mid y_{t+1}, \Lambda_t^b)\cdot(\Sigma_{n,m}^x + \Sigma_{t+1}^b)^{-1}\cdot(y_{t+1} - \mu_{n,m}^x - \mu_t^b) \tag{21}$$

where Σ_{t+1}^b, which is estimated later, is replaced by Σ_t^b and I_{t+1} can be obtained by recursive computation based on its previous values at time t.

3.3.2. Noise variance updating

Applying the similar derivations to the mean updating of noise above, we can obtain the sequential estimate of noise variance at time t+1 given the estimated $\mu^b{}_{t+1}$.

$$\Sigma_{t+1}^b = \Sigma_t^b + I_{t+1}^{-1}(\Sigma_t^b)\cdot S(\Sigma_t^b, y_{t+1}) \tag{22}$$

where

$$I_{t+1}(\Sigma_t^b) = \sum_{\tau=1}^{t+1}\sum_{n=1}^{N}\sum_{m=1}^{M} p(n, m \mid y_\tau, \Sigma_{\tau-1}^b)\cdot\{\frac{-1}{2(\Sigma_t^b + \Sigma_{n,m}^x)^2} + \frac{(y_\tau - \mu_{n,m}^x - \mu_\tau^b)^2}{(\Sigma_t^b + \Sigma_{n,m}^x)^3}\} \tag{23}$$

$$S(\Sigma_t^b, y_{t+1}) = \sum_{n=1}^{N}\sum_{m=1}^{M} p(n, m \mid y_{t+1}, \Sigma_t^b)\cdot\{\frac{-1}{2(\Sigma_t^b + \Sigma_{n,m}^x)} + \frac{(y_{t+1} - \mu_{n,m}^x - \mu_{t+1}^b)^2}{2(\Sigma_t^b + \Sigma_{n,m}^x)^2}\} \tag{24}$$

It is also desirable to introduce forgetting factor into the above updating processes like in (7) to de-emphasize the contribution of the history data.

3.4. Implementation issues

3.4.1. Initial estimate of noise priori parameters

The sample mean and sample variance of noise power spectra at the beginning segment of each test utterance, which is normally assumed to be free of any speech data, is used to obtain the initial estimate of the parameters of noise priori.

3.4.2. Representation of clean speech distribution

The computation of $\gamma_{\tau|t+1,B}(n,m)$ in (7) is difficult and need making some approximations. Forward-backward algorithm and Viterbi approximation can be used to compute the probability. However, it is time consuming and not suitable for sequential computation. In this paper, we assume that the power spectral space of clean speech is represented by N Gaussian mixture models, e.g. Λ_X, and each model has M components with mixture coefficients, means and diagonal covariance matrices $\{w_{n,m}, \mu_{n,m}, \Sigma_{n,m} | 1\le n\le N, 1\le m\le M\}$. It is assumed the speech frame to be independent and identically distributed. Then the probability is approximated by

$$\gamma_{\tau|t+1,B}(n,m) = p(s_\tau = n, c_\tau = m \mid y_\tau, B_{\tau-1}, \Lambda_X)$$
$$= \frac{w_{n,m} N(f_{b_{\tau-1}}(y_\tau); \mu_{n,m}^x, \Sigma_{n,m}^x)}{\sum_{i=1}^{N}\sum_{j=1}^{M} w_{i,j} N(f_{b_{\tau-1}}(y_\tau); \mu_{i,j}^x, \Sigma_{i,j}^x)} \tag{25}$$

The other probabilities in (20), (21), (23) and (24) can also be obtained easily under the representation of speech space.

3.4.3. Post-processing

To ensure that the power spectra are non-negative, same techniques in spectral subtraction are applied as postprocessing [7]. The obtained estimate of original speech spectrum is as follows

$$x_t = \begin{cases} y_t - \alpha\cdot b_t, & if \quad y_t - \alpha\cdot b_t > \beta\cdot b_t \\ \beta\cdot b_t, & otherwise \end{cases} \tag{26}$$

where α represents over-subtraction coefficient, β represents floor coefficient.

4. EXPERIMENTS AND RESULTS

The sequential MAP estimation for spectral feature enhancement described in this paper has been evaluated in large vocabulary continuous speech recognition. One triphone model set for recognition was trained by clean speech. A trigram language model was used in all the tests with a 40000 words vocabulary. Other settings, including acoustic front-end, HMM topology, were the same as described in [8].

The output energies of Mel-scaled filterbank were used to replace the power spectrum due to its lower dimension and the still kept linear relationship denoted in (1).

The clean-speech model set described in Section 3.4.2 includes 184 G corresponding to 61 Mandarin base phones with three outputs plus a silence unit. In the process to generate the models, the classification information of cepstral features is used to guild the clustering process of corresponding spectral features in order to keep consistency with recognition system and to take advantage of reduced correlation within a cepstral feature.

The test sets are generated by adding noise to a clean speech set which includes 400 sentences spoken by 10 male and 10 female speakers. The lengths of utterances range from 5 seconds to 9 seconds. The recognition correction rate for the clean speech set is 76.9%. In the following experiments about sequential MAP algorithm, the Log-normal distribution was chosen as the noise priori.

4.1. Evaluation on stationary noise

The main objective of these experiments is to compare the proposed algorithm with methods in ML criterion. Moreover, the batch method [1] is implemented in order to compare it with sequential estimation techniques. In this set of experiments, the proposed algorithm is evaluated on white noise source from NOISEX92 database. The white noise source is added to clean speech by varying the signal-to-noise ratio (SNR) from 0db to 20db.

Recognition results are shown in Table I. Within this table, *No_Comp* represents no noise compensation. *Batch_ML*, *Seq_ML* denote batch and sequential estimation techniques in ML rule. *Seq_MAP* denotes the proposed sequential estimation method in MAP rule, respectively. In *Seq_MAP*, no noise priori updating process was adopted due to the stationarity of white noise. All the methods are implemented in Mel-power spectral domain.

SNR(db)	**0**	**5**	**10**	**15**	**20**
No_Comp	3.56	14.37	33.43	49.18	61.46
Batch_ML	10.34	28.52	45.97	60.20	65.28
Seq_ML	10.42	27.81	43.93	58.25	66.07
Seq_MAP	10.45	28.30	45.45	60.32	66.72

Table I. Recognition correction rates (%) with additive white noise (The forgetting factor ξ=0.95 and weight η=1.0)

From Table I, it has shown the validity of the compensation methods in power spectral domain. The sequential MAP method performed better than sequential ML method especially at middle SNR. Compared with batch method, sequential methods cannot guarantee to be better when the background noise is rather stationary. However, the performance of sequential MAP method can be comparable to batch method over all the SNR ranges for the sake of the introduction of noise priori knowledge.

4.3. Evaluation on non-stationary noise

In order to study further the performance of the proposed algorithm under non-stationary environment, the experiments were evaluated on artificial time-varying noise. White and babble noise sources that have equal energy were linearly mixed by making the mixing weight from one to zero for white noise and conversely for babble noise within each utterance. The noise priori updating was adopted in order to track the time varying environment. The recognition results are shown in Table II.

From Table II, we can find out that sequential methods outperformed the batch method almost over all the SNR ranges under the non-stationary noise environment. The sequential MAP method with noise priori updating process was the best among these compensation methods.

SNR(db)	**0**	**5**	**10**	**15**	**20**
No_Comp	2.47	11.75	34.45	54.82	66.11
Batch_ML	5.26	19.58	41.46	59.10	65.84
Seq_ML	4.87	21.43	43.08	60.46	66.50
Seq_MAP	6.11	21.81	44.10	62.21	67.02

Table II. Recognition correction rates (%) with additive time varying noise (The forgetting factor ξ=0.6~0.9 for low SNR to high SNR and weight η=0.8)

The above experimental results also demonstrated the effectivity of the chosen log-normal distribution.

5. CONCLUSIONS

The classic stochastic matching algorithm did not consider any priori information and usually implemented in batch mode in cepstral domain. The proposed algorithm accurately models the environment mismatch as an additive bias in power spectral domain. A new framework based on MAP criterion is proposed, which incorporates the noise priori knowledge. Sequential techniques are adopted into the MAP framework to adaptively estimate the spectral bias. In order to track time varying noise environment, the parameters of noise priori are also sequentially updated. The preliminary results reported on continuous speech recognition demonstrated the validity of the proposed algorithm not only under stationary but also non-stationary environment.

To further improve the proposed algorithm, the choice of noise priori, how to select the forgetting factor and weight and the application of other Bayesian methods are under investigation.

6. REFERENCES

[1] Ananth Sankar, Chin-Hui Lee, "A maximum-likelihood approach to stochastic matching for robust speech recognition", *IEEE Trans. on Speech and Audio Processing*, pp.190-202, vol. 4, No. 3, 1996

[2] Char-Shih Huang, Hsiao-Chuan Wang, Chin-Hui Lee, "An SNR-incremental stochastic matching algorithm for noisy speech recognition", *IEEE Trans. on Speech and Audio Processing*, pp.866-873, vol. 9, No. 8, 2001

[3] Vikram Krishnamurthy, John B. Morre, "On-line estimation of Hidden Markov Model parameters based on the Kullback-Leibler information measure", *IEEE Trans. on Signal Processing*, pp.2557-2573, vol. 41, no. 8, 1993

[4] D. M. Titterington, "Recursive parameter estimation using incomplete data", *Journal of the Roy. Stat. Soc. B*, pp.257-267, vol. 46, Issue 2, 1984

[5] M. J. Gales, S. J. Young, "Robust speech recognition in additive and convolutional noise using parallel model combination", *Computer Speech and Language*, pp. 289-307, 1995

[6] Yariv Ephraim, David Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator", *IEEE Trans. on Acoustic, Speech and Signal Processing*, pp.1109-1121, vol. 32, no. 6, 1984

[7] M. Berouti, R.Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise", *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol.28, pp.137-145, Apr. 1980.

[8] Gao sheng, Tan Lee, Y.W. Wong, Bo Xu, P.C. Ching and Taiyi Huang, "Acoustic modeling for Chinese speech recognition: a comparative study of Mandarin and Cantonese", *Proc. ICASSP*,Vol.3,pp.1261-1264, June 2000.

[9] Li deng, Jasha Droppo, Alex Acero, "Log-domain speech feature enhancement using sequential MAP noise estimation and a phase-sensitive model of the acoustic environment", *Proc. ICSLP*, pp.1813-1816, 2002

[10] Mohamed Afify, "Sequential bias compensation for robust speech recognition", *Proc. Eurospeech*, 1999

[11] Lionel Delphin-Poulat, Chafic Mokbel, Jérôme Idier, "Frame-synchronous stochastic matching based on the Kullback-Leibler information", *Proc. ICASSP*, 1998