INCREMENTAL BAYES LEARNING WITH PRIOR EVOLUTION FOR TRACKING NONSTATIONARY NOISE STATISTICS FROM NOISY SPEECH DATA

Li Deng, Jasha Droppo, and Alex Acero

Microsoft Research, One Microsoft Way, Redmond WA 98052, USA

ABSTRACT

In this paper, a new approach to sequential estimation of the time-varying prior parameters of nonstationary noise is presented using the log-spectral or cepstral data of the corrupted noisy speech. Incremental Bayes learning is developed to provide a basis for noise prior evolution, recursively updating the noise prior statistics (mean and variance) using the approximate Gaussian posterior computed at the preceding time step. The algorithm for noise prior evolution is derived in detail, and is evaluated using the Aurora2 database with the root-mean-square (RMS) error measure. Experimental results show that when the time-varying variance and mean of the nonstationary noise prior are estimated and exploited, superior performance is achieved compared with using either no noise prior information or using the time-invariant, fixed mean and variance in the noise prior distribution.

1. INTRODUCTION

In a wide class of speech feature enhancement algorithms, including spectral subtraction and Bayesian estimation [4, 3, 5], the estimation accuracy of the corrupting noise or of its statistics is the most crucial factor determining the effectiveness of the enhancement algorithms. This is especially true when the corrupting noise is nonstationary, or varying over time in its statistics. In [2] and [3], we introduced the maximum-likelihood (ML) and maximum a posteriori (MAP) techniques, respectively, for sequential point estimation of nonstationary noise using an iteratively linearized nonlinear model for the acoustic environment.[1] It was demonstrated in [3] that with the use of a simple Gaussian prior model for the distribution of the noise, the MAP estimate outperformed the ML counterpart in the quality of the noise estimate. This leads to better speech feature enhancement and greater robust speech recognition accuracy.

In the MAP technique presented in [3], the mean and variance parameters associated with the Gaussian noise prior are fixed from a segment of each speech-free test utterance. When the noise is nonstationary, however, it is natural to expect that the noise prior should also change as a function of time in order to reflect realistic noise prior statistics. Just how to achieve such improved modeling is the subject of this paper.

A new approach to tracking time-varying parameters of nonstationary noise is presented in this paper using the log-spectral data of the corrupted noisy speech. The theoretical basis of the new sequential estimation algorithm is incremental Bayes learning, where a time-varying noise prior distribution is assumed and its hyperparameters (mean and variance) are updated recursively using the approximate Gaussian posterior computed at the preceding time step. In contrast to the earlier noise tracking approaches based on ML and MAP point estimates, the new approach provides additional uncertainty information regarding the time-varying variance of the noise. Experimental results using the root-mean-square (RMS) measure show that incorporating the time-varying mean and variance of the noise estimate gives superior performance compared with the earlier work which uses either no noise prior information [2] or uses the time-invariant, fixed mean and variance in the noise prior distribution [3].

This paper is organized as follows. In Section 2, we provide a general principle for incremental Bayes learning based on a recursive formulation of Bayes' rule, and its specific application to tracking noise prior evolution. An algorithm for estimating time-varying mean and variance in the noise prior distribution, under the Gaussian assumption of the prior, is derived and presented in Section 3. Experimental evaluation of the algorithm is included in Section 4, using the Aurora2 database and the RMS error measure. Finally, in Section 5, we provide a summary of the work and discuss potential use of the estimated noise statistics for speech feature enhancement.

2. INCREMENTAL BAYES LEARNING OF NONSTATIONARY NOISE

Let $y_1^t = y_1, y_2, ..., y_\tau, ..., y_t$ be a sequence of noisy speech observation data, expressed in the log domain (such as log-spectra or cepstra), and are assumed to be scalar-valued without loss of generality. Data y_1^t are used to sequentially estimate the corrupting noise sequence $n_1^t = n_1, n_2, ..., ..., n_t$ with the same data length t. Within the Bayesian learning framework, we assume that the knowledge about noise n (treated as an unknown parameter) is contained in a given a-priori distribution of $p(n)$. If the noise sequence is stationary, i.e., the statistical properties of the noise do not change over time, then the conventional Bayes inference (i.e., computing the posterior) on noise parameter n at any time can be accomplished via the "batch-mode" Bayes' rule:

$$p(n|y_1^t) = \frac{p(y_1^t|n)p(n)}{\int_\Theta p(y_1^t|n)p(n)dn},$$

where Θ is an admissible region of the noise parameter space. Given $p(n|y_1^t)$, any estimate on noise n is possible in principle. For example, the conventional MAP point estimate on noise n is computed as a global or local maximum of the posterior $p(n|y_1^t)$. The minimum mean square error (MMSE) estimate is the expectation over the posterior $p(n|y_1^t)$.

However, when the noise sequence is nonstationary and the training data of noisy speech y_1^t is presented sequentially as in

[1] Non-iterative versions of the ML technique for noise estimation can be found in [1, 7].

most practical speech feature enhancement applications, new noise estimation techniques are needed in order to track the noise statistics that is changing over time. One common technique, which we explore in this work, is based on iterative applications of Bayes' rule [8, 6]:

$$p(n_t|y_1^t) = \frac{1}{C_t} p(y_t|y_1^{t-1}, n_t) p(n_t|y_1^{t-1}),$$

where $C_t = p(y_1^t|y_1^{t-1}) = \int_\Theta p(y_t|y_1^{t-1}, n_t)p(n_t|y_1^{t-1})dn_t$.

Assuming conditional independency between noisy speech y_t and its past y_1^{t-1} given n_t, or $p(y_t|y_1^{t-1}, n_t) = p(y_t|n_t)$, and assuming smoothness in the posterior: $p(n_t|y_1^{t-1}) \approx p(n_{t-1}|y_1^{t-1})$, we obtain

$$p(n_t|y_1^t) \approx \frac{1}{C_t} p(y_t|n_t) p(n_{t-1}|y_1^{t-1}). \quad (1)$$

Incremental learning of nonstationary noise can now be established by repeated use of Eq. 1 as follows. Initially, in absence of noisy speech data y, the posterior PDF comes from the known prior $p(n_0|y_0) = p(n_0)$. Then use of Eq. 1 for $t = 1$ produces:

$$p(n_1|y_1) \approx \frac{1}{C_1} p(y_1|n_1) p(n_0), \quad (2)$$

and for $t = 2$ it produces:

$$p(n_2|y_1, y_2) \approx \frac{1}{C_2} p(y_2|n_2) p(n_1|y_1),$$

using the $p(n_1|y_1)$ already computed from Eq. 2. For $t = 3$, Eq. 1 becomes:

$$p(n_3|y_1^3) \approx \frac{1}{C_3} p(y_3|n_3) p(n_2|y_1, y_2),$$

and so on. This process thus recursively generates a sequence of posteriors (provided that $p(y_t|n_t)$ is available):

$$p(n_1|y_1), \;\; p(n_2|y_1^2), \;\; ..., p(n_\tau|y_1^\tau), \;\; ..., p(n_t|y_1^t), ... \quad (3)$$

which provides a basis for making incremental Bayes' inference on the nonstationary noise sequence n_1^t. The general principle of incremental Bayes' inference discussed so far will now be applied to a specific acoustic distortion model, which supplies the frame-wise data PDF $p(y_t|n_t)$, and under the simplifying assumption that the noise prior be Gaussian.

3. NOISE PRIOR EVOLUTION

3.1. Prior evolution and sequential update of noise hyperparameters

The essence of incremental Bayes learning is to update the current "prior" distribution about the parameter (noise in our case) using the posterior given the observed data up to the most recent past, since this posterior is the most complete information about the parameter preceding the current time. Therefore, the posterior sequence in Eq. 3 becomes a time-varying prior sequence (i.e., *prior evolution*) for noise distributional parameters of interest (with the time shift of one frame in size).

For data likelihood $p(y_t|n_t)$, which is non-Gaussian (and will be described shortly), the posterior is necessarily non-Gaussian. A successive application of Eq. 1 would result in a fast expanding combination of the previous posteriors and lead to intractable forms. It is well known that approximations are needed to overcome the intractability [6]. The approximation that we introduce in this work is to apply the first-order Taylor series expansion to linearize the nonlinear relationship between y_t and n_t. This leads to a Gaussian form of $p(y_t|n_t)$. Therefore, the time-varying noise prior PDF $p(n_{\tau+1})$, which is inherited from the posterior for the past data history $p(n_\tau|y_1^\tau)$, can be approximated by the Gaussian:

$$\begin{aligned} p(n_\tau|y_1^\tau) &= \frac{1}{(2\pi)^{1/2}\sigma_{n_\tau}} \exp\left[-\frac{1}{2}\left(\frac{n_\tau - \mu_{n_\tau}}{\sigma_{n_\tau}}\right)^2\right] \\ &\doteq \mathcal{N}[n_\tau; \mu_{n_\tau}, \sigma_{n_\tau}^2], \end{aligned} \quad (4)$$

where μ_{n_τ} and $\sigma_{n_\tau}^2$ are called the hyperparameters (mean and variance) that characterize the prior PDF. Then the posterior sequence in Eq. 3 computed from recursive Bayes' rule Eq. 1 offers a principled way of determining the temporal evolution of the hyperparameters, which we describe below.

3.2. Acoustic-distortion and clean-speech models for computing data likelihood $p(y_t|n_t)$

We first assume a time-invariant mixture-of-Gaussian model for log-spectra of clean speech x:

$$p(x) = \sum_m p(m)\mathcal{N}[x; \mu_x(m), \sigma_x^2(m)]. \quad (5)$$

We then use a simple nonlinear acoustic-distortion model in the log-spectral domain (discussed in more detail in [2]):

$$\exp(y) = \exp(x) + \exp(n), \;\; \text{or} \;\; y = x + g(n - x), \quad (6)$$

where the nonlinear function is:

$$g(z) = \log[1 + \exp(z)].$$

In order to obtain a useful form for the data likelihood $p(y_t|n_t)$, we employ Taylor series expansion to linearize nonlinearity g in Eq. 6. This gives the linearized model of

$$y \approx x + g(n_0 - \mu_x(m_0)) + g'(n_0 - \mu_x(m_0))(n - n_0), \quad (7)$$

where n_0 is the Taylor series expansion point and the first-order series expansion coefficient can be easily computed as:

$$g'(n_0 - \mu_x(m_0)) = \frac{\exp(n_0)}{\exp[\mu_x(m_0)] + \exp(n_0)}.$$

In evaluating functions g and g' in Eq. 7, the clean speech value x is taken as the mean ($\mu_x(m_0)$) of the "optimal" mixture Gaussian component m_0.

Eq. 7 defines a linear transformation from random variables x to y (after fixing n). Based on this transformation, we obtain the PDF on y below from the PDF on x (Eq. 5):

$$\begin{aligned} p(y_t|n_t) &= \sum_m p(m)\mathcal{N}[y_t; \mu_y(m,t), \sigma_y^2(m,t)] \\ &\approx \mathcal{N}[y_t; \mu_y(m_0,t), \sigma_y^2(m_0,t)], \end{aligned} \quad (8)$$

where the optimal mixture component is determined by

$$m_0 = \arg\max_m \mathcal{N}[y_t; \mu_y(m,t), \sigma_y^2(m,t)],$$

and where the mean and variance of the approximate Gaussian are[2]

$$\begin{aligned} \mu_y(m_0,t) &= \mu_x(m_0) + g_{m_0} + g'_{m_0} \times (n_t - n_0) \\ \sigma_y^2(m_0,t) &= \sigma_x^2(m_0) + g'^2_{m_0}\sigma^2_{n_t}. \end{aligned} \quad (9)$$

3.3. Algorithm for estimating time-varying mean and variance in the noise prior

Given the approximate Gaussian form for $p(y_t|n_t)$ as in Eq. 8 and for $p(n_\tau|y_1^\tau)$ as in Eq. 4, we now derive the algorithm for determining noise prior evolution, expressed as sequential estimates of time-varying hyperparameters of mean μ_{n_τ} and variance $\sigma^2_{n_\tau}$.

Substituting Eqs. 4 and 8 into Eq. 1, we obtain

$$\begin{aligned} & \mathcal{N}(n_t;\mu_{n_t},\sigma^2_{n_t}) \\ \propto\ & \mathcal{N}[y_t;\mu_y(m_0,t),\sigma_y^2(m_0,t)]\mathcal{N}(n_{t-1};\mu_{n_{t-1}},\sigma^2_{n_{t-1}}) \\ \approx\ & \mathcal{N}[g'_{m_0}n_{t-1};\mu_1,\sigma_y^2(m_0,t)]\mathcal{N}(n_{t-1};\mu_{n_{t-1}},\sigma^2_{n_{t-1}}) \end{aligned} \quad (10)$$

where $\mu_1 = y_t - \mu_x(m_0) - g_{m_0} + g'_{m_0}n_0$, and the assumption of noise smoothness was used. We match the means and variances, respectively, of the left and right hand sides in Eq. 10 to obtain the prior evolution formulas:

$$\begin{aligned} \mu_{n_t} &= \frac{g'_{m_0}\ \bar{\mu}_1\ \sigma^2_{n_{t-1}} + \mu_{n_{t-1}}\sigma_y^2(m_0,t-1)}{g'^2_{m_0}\sigma^2_{n_{t-1}} + \sigma_y^2(m_0,t-1)}, \quad (11) \\ \sigma^2_{n_t} &= \frac{\sigma_y^2(m_0,t-1)\sigma^2_{n_{t-1}}}{g'^2_{m_0}\sigma^2_{n_{t-1}} + \sigma_y^2(m_0,t-1)}, \end{aligned}$$

where $\bar{\mu}_1 = y_t - \mu_x(m_0) - g_{m_0} + g'_{m_0}\mu_{n_{t-1}}$. In establishing Eq. 11, we used the previous time' prior mean as the Taylor series expansion point for noise; i.e., $n_0 = \mu_{n_{t-1}}$. We also used the well established result in Gaussian computation (setting $a = g'_{m_0}$):

$$\mathcal{N}(ax;\mu_1,\sigma_1^2)\mathcal{N}(x;\mu_2,\sigma_2^2) = \frac{1}{2\pi\sigma_1\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 + K\right],$$

where

$$\mu = \frac{a\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{a^2\sigma_2^2 + \sigma_1^2};\quad \sigma^2 = \frac{\sigma_1^2\sigma_2^2}{a^2\sigma_2^2 + \sigma_1^2}.$$

4. EVALUATION EXPERIMENTS

The algorithm for noise prior evolution described in the preceding section has been evaluated on the Aurora2 database. Since the true noise in the form of log-spectrum is available from the database, it is possible to quantitatively evaluate exactly how accurate the estimated noise is.

4.1. Baseline estimates of nonstationary noise

In order to evaluate the effectiveness of the noise prior evolution algorithm presented in Section 3.3, in addition to baseline estimates, we also need to establish a corresponding point estimate of noise based on the estimated prior evolution sequence. Since the mode of a Gaussian is its mean, the MAP (point) estimate for the time-varying noise $\hat{n}_t$ is in the same form as Eq. 11 for prior evolution of the noise mean:

$$\hat{n}_t^{MAP} = \frac{g'_{m_0}\mu_1^{MAP}\sigma^2_{n_{t-1}} + \hat{n}_{t-1}^{MAP}\sigma_y^2(m_0,t-1)}{g'^2_{m_0}\sigma^2_{n_{t-1}} + \sigma_y^2(m_0,t-1)}, \quad (12)$$

where $\mu_1^{MAP} = y_t - \mu_x(m_0) - g_{m_0} + g'_{m_0}\hat{n}_{t-1}^{MAP}$. Using such a MAP estimate of noise and the true noise in the same form of log-spectra, we can then compute the RMS error to evaluate the effectiveness of the noise prior evolution algorithm.

In our experiments, we compare three point estimates of non-stationary noise in the Aurora2 database:

A. The MAP estimate ($\hat{n}_t^{MAP}$) of Eq. 12, incorporating time-varying noise prior mean and variance hyperparameters determined by the noise prior evolution algorithm;

B. The MAP estimate (denoted by $\hat{n}_t^{map}$) of Eq. 12, except replacing the estimated hyperparameters $\mu_{n_{t-1}}$ (which is $\hat{n}_{t-1}^{MAP}$) and $\sigma^2_{n_{t-1}}$ by fixed, time-invariant values of $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ computed from the first 15 speech-free frames in each of the Aurora2 utterances. This estimate becomes[3]

$$\begin{aligned} \hat{n}_t^{map} &= \frac{g'_{m_0}\ \mu_1^{map}\ \hat{\sigma}_n^2 + \hat{\mu}_n\sigma_y^2(m_0)}{g'^2_{m_0}\hat{\sigma}_n^2 + \sigma_y^2(m_0)} \\ &= \frac{2g'_{m_0}\hat{\sigma}_n^2\hat{\mu}_n + \sigma_x^2(m_0)\hat{\mu}_n + d_t}{2g'^2_{m_0}\hat{\sigma}_n^2 + \sigma_x^2(m_0)}, \end{aligned} \quad (13)$$

where $\mu_1^{map} = y_t - \mu_x(m_0) - g_{m_0} + g'_{m_0}\hat{\mu}_n$, and $d_t = g'_{m_0}\hat{\sigma}_n^2 \times [y_t - \mu_x(m_0) - g_{m_0}]$.

C. The ML estimate (denoted by $\hat{n}_t^{ML}$), as a special case of Eq. 13 by setting $\sigma_n^2 \to \infty$ and taking the limit of the ratio.[4]

4.2. Experimental results

The new MAP noise estimate (A: $\hat{n}_t^{MAP}$) computed from Eq. 12, which incorporates the time-varying noise prior mean and variance according to Eq. 11, is evaluated by comparing its RMS error with the RMS errors from two baseline noise estimates (B: $\hat{n}_t^{map}$ or MAP estimate with a fixed noise prior mean and variance; and C: $\hat{n}_t^{ML}$ or ML estimate). The estimates are computed in the Mel log-spectral domain first, and then converted to the cepstral domain (13 MFCCs: C0-C12) via the cosine transform. The RMS error is computed using the MFCCs of true noise via the same cosine transform on the Mel log-spectra.

The comparative RMS values for the full range of MFCCs are listed in Table 1, averaged over the utterances from the SNR=0dB portion of the Exhibition environment condition (N4) in the Aurora2's Set-A. While the new noise estimate A is uniformly better (i.e., lower RMS errors) than the two baseline estimates, the greatest gain is on C0, resulting in approximately 10% RMS error reduction. The higher-order MFCCs tend to have lower gains. Also, as consistent with the robust speech recognition results published in [3], the MAP estimate (B) with a fixed noise prior is slightly but consistently better than the ML estimate (C).

The same type of RMS-error comparison for the SNR=10dB portion of the N4 condition in Set-A is shown in Table 2. Similar advantages of the new estimate A over that of B or C are demonstrated. We note that the overall errors of the estimates are slightly higher for the SNR=10dB than the SNR=0dB conditions.

For a different type of noise, N1 (Subway) of Set-A, in the Aurora2 database, we also computed the RMS errors for the three point estimates of the noise. The results for the SNR=5dB portion

[2]As a notational shorthand, we use g_{m_0} to denote $g(n_0 - \mu_x(m_0))$, and g'_{m_0} to denote $g'(n_0 - \mu_x(m_0))$.

[3]Note this is a special case of the MAP estimation algorithm in [3], which also used the same first 15 speech-free frames in each of the Aurora2 test utterances to estimate the time-invariant hyperparameters.

[4]Note this is a special case of the ML estimation algorithm in [2].

	C0	C1	C2	C3	C6	C9	C12
A	5.4	3.2	2.7	2.4	2.3	1.9	1.5
B	5.9	3.5	2.8	2.5	2.4	2.0	1.5
C	6.1	3.7	2.9	2.5	2.5	2.0	1.5

Table 1. RMS errors for the new MAP noise estimate (A: $\hat{n}_t^{MAP}$) and the two baseline MAP (B: $\hat{n}_t^{map}$) and ML (C: $\hat{n}_t^{ML}$) noise estimates. A and B differ in whether the noise prior's mean and variance are fixed, or are allowed to change over time according to Eq. 11. Noise and noisy speech data are from the SNR=0dB portion of the N4 condition (Exhibition) in the Aurora2's Set-A.

	C0	C1	C2	C3	C6	C9	C12
A	6.2	3.4	3.0	2.6	2.1	1.8	1.6
B	6.7	3.7	3.2	2.7	2.2	1.9	1.7
C	6.9	3.8	3.3	2.7	2.4	1.9	1.7

Table 2. Data: SNR=10dB portion of the N4 condition of Set-A

of the utterances are listed in Table 3. Since the "subway" noise in N1 is significantly more nonstationary than the "exhibition" noise in N4, the RMS errors (in C0) are much larger also. This reflects the well-known difficulty of estimating highly nonstationary noise. However, with the use of the time-varying noise prior as proposed in this paper, again the RMS error in C0 has been reduced by about 10% for the more difficult type of noise of N1. The RMS-error results for the same type of noise but with SNR=15dB are presented in Table 4. Again, the overall RMS errors are somewhat higher compared with the condition having a lower SNR. But again, the RMS error reduction with a similar magnitude to Table 3 is observed due to the use of the time-varying noise prior.

	C0	C1	C2	C3	C6	C9	C12
A	10.0	3.6	3.1	2.5	2.3	1.6	1.4
B	11.2	3.8	3.2	2.6	2.4	1.7	1.4
C	11.5	3.9	3.3	2.6	2.4	1.7	1.4

Table 3. Data: SNR=5dB portion of the N1 (Subway) condition

5. SUMMARY AND DISCUSSIONS

In this paper, we propose and evaluate an incremental Bayes learning approach to sequential estimating or tracking the time-varying mean and variance of nonstationary noise, using the log-spectral or cepstral data of the corrupted noisy speech. This approach generalizes the earlier noise tracking approaches based on the ML and MAP point estimates, where either no prior information about the noise was exploited [2], or such information was assumed to be fixed over the entire length of the utterance [3].

The main contribution of this paper is to apply incremental Bayes learning, which has been successfully used for on-line adaptation of HMM parameters [6], to the problem of nonstationary noise tracking. Key differences, however, exist between these two types of applications. First, the parameter update interval is significantly shorter in noise tracking (frame as the interval) than in HMM adaptation (utterance as the interval). Second, the data likelihood for noise tracking is derived from a linearized acoustic distortion model, while that for HMM adaptation comes from the HMM likelihood computation. Third, the form of the prior PDF for noise tracking is assumed with a much simpler form to facilitate the tracking algorithm development. Based on a set of simplified yet effective assumptions, we used approximate recursive Bayes' rule and quadratic term matching to successfully derive the noise prior evolution formulas as summarized in Eq. 11.

Experimental results show moderate improvement on the noise tracking accuracy, measured by RMS error reduction, compared with the two baseline noise tracking algorithms developed from the previous work [2, 3]. Our future work will focus on utilizing both the estimated mean and variance in the noise prior to improve the effectiveness of speech feature enhancement within the framework of phase-sensitive modeling of acoustic environments established in [3, 5]. The enhancement algorithm presented in [3] was only able to use the point (mean) estimate of noise. Extension of the algorithm to make use of the variance information provided by the prior evolution algorithm presented in this paper is expected to improve the enhancement performance, thus demonstrating the practical value of noise prior evolution.

	C0	C1	C2	C3	C6	C9	C12
A	11.5	4.0	3.1	2.6	2.2	1.7	1.5
B	12.4	4.2	3.2	2.7	2.3	1.8	1.5
C	12.5	4.3	3.4	2.8	2.3	1.8	1.5

Table 4. Data: SNR=15dB portion of the N1 (Subway) condition

6. REFERENCES

[1] M. Afify and O. Siohan. "Sequential noise estimation with optimal forgetting for robust speech recognition," *Proc. ICASSP*, Vol.1, 2001, pp. 229-232.

[2] L. Deng, J. Droppo, and A. Acero. "Recursive estimation of nonstationary noise with iterative stochastic approximation," *Proc. ASRU Workshop*, Trento, Italy, Dec. 2001.

[3] L. Deng, J. Droppo, and A. Acero. "Log-domain speech feature enhancement using sequential MAP noise estimation and a phase-sensitive model of the acoustic environment," *Proc. ICSLP*, Denver CO, Sept. 2002, pp. 192-195.

[4] L. Deng, J. Droppo, and A. Acero. "A Bayesian approach to speech feature enhancement using the dynamic cepstral prior," *Proc. ICASSP*, Orlando, Florida, May 2002, pp. 829-832.

[5] J. Droppo, A. Acero, and L. Deng. "A nonlinear observation model for removing noise from corrupted speech log Mel-spectral energies," *Proc. ICSLP*, 2002, pp. 182-185.

[6] Q. Huo and C. Lee. "On-line adaptive learning of the continuous density HMM based on approximate recursive Bayes estimate," *IEEE Trans. Speech Audio Proc.*, Vol. 5, No. 2, March 1997, pp. 161-172.

[7] N.S. Kim. "Nonstationary environment compensation based on sequential estimation," *IEEE Sig. Proc. Letters*, Vol.5, 1998, pp. 57-60.

[8] J. Spragins. "A note on the iterative application of Bayes' rule," *IEEE Trans. Inform. Theory*, Vol.11, No. 4, pp. 544-549.