Paper Title : Speech Parameter Estimation and Enhancement Using the EM Algorithm

EM 알고리즘을 이용한 음성 파라미터 추정 및 향상

Ki Yong Lee*, Young-Tae Kang*, Byung-Gook Lee**

이기용*, 강영태*, 이병국**

Abstract

In many applications of signal processing, we have to deal with densities which are highly non-Gaussian or which may have Gaussian shape in the middle but have potent deviations in the tails. To fight against these deviations, we consider a finite mixture distribution for the speech excitation. We utilize the EM algorithm for the estimation of speech parameters and their enhancement. *Robust Kalman filtering* is used in the enhancement process, and a detection/estimation technique is used for parameter estimation. Experimental results show that the proposed algorithm performs better in adverse SNR input conditions.

요약

신호처리의 많은 분야에서, 심하게 비가우시안 성질을 가지는 분포, 혹은 분포의 중간은 가우시안 특성을 가지지만 양 끝에서는 편차가 크게 나는 분포를 다루어야 하는 경우가 종종 있다. 이러한 편차에 효과적으로 대처하기 위하여 본 논문에서는 음성 신호의 여기 신호로서 혼합 분포(mixture distribution)을 고려한다. 이것은 음성 분석시 피치 주파수가 미치는 영향을 감소시키며, 배경 잡음을 제거하는 데에도 효과적이다. 음성 신호 파라미터의 추정 및 향상을 위하여 EM 알고리즘을 사용하며, 향상 과정에서는 강인 칼만 필터링 기법을, 파라미터 추정 과정에서는 검출/추정 기법을 사용한다. 실험 결과, 본 논문에서 제안하는 알고리즘이 입력 신호대잡음비가 열악한 경우에 기존의 것보다 우수한 성능을 보인다.

I. Introduction

Background noise seriously degrades the performance of speech signal processing systems. This is primarily because most systems are based upon the data obtained in a noise-free environment. In general there two ways to improve the performance of speech recognition/coding system subjected to noise. One is preprocessing that removes at the front end the noise in the speech

*Department of Electronics Engineering, Changwon National University
**Department of Electronics Engineering, Seoul National University

접수일자 : 1994년 9월 8일

signal. This enjoys the advantage that there need be no modification to the existing system structure. This technique aims to remove the noise in the speech signal, to make the preprocessed signal as close as possible to the original signal, rather than improve the quality of the speech signal [1, 2]. The other method is speech quality enhancement for listeners. This processes the speech signal with an emphasis to enhance the acoustical quality of the signal, rather than the signal itself. This method can be used under various noise conditions. Proposed by Lim *el al.* in 1978, it is a speech enhancement method based on glottal model parameter estimation assumming an all-pole model in white Gaussian noise environment [3]. And a constrained iterative speech enhancement method was proposed which enhances speech using the *maximum a posteriori* technique in order to estimate the speech parameters from the signal corrupted by noise [4]. Ephraim proposed a hidden Markov model with mixed Gaussian outputs based on statistical modeling[5, 6].

Conventional methods assume the following : The speech source assumes that pitch-periodic impulse train is used for voiced speech, and a white Gaussian noise for unvoiced speech. The least square method that is used in the analysis step employs the assumption that the input signal is Gaussian. Hence the pitch period affects the parameters during voiced signal analysis. The second assumption is that the characteristics of the degrading noise is known *a priori*. Since this assumption is not very realistic, conventional methods inevitably suffer. In order to develop an estimation method, not affected by the sound source, for obtaining parameters from speech signal ontaining unknown noise, a robust speech enhancement algorithm using the EM(estimate-maximize) algorithm is proposed.

The proposed algorithm may be divided into the M-step and the E-step. The M-step is again divided into the estimation procedure that computes the glottal model parameters and their variances, and a detection procedure for voiced signals which obtains the positions of the pulse train corresponding to the speech source. The E-step employs a robust Kalman filter that enhances parameters from corrupted speech signal. These two steps are iterated in turn to produce the speech parameter estimates.

Ⅱ. Proposed Model

In general, the speech signal $s(t)$ can be modeled as the output from an all-pole filter with input $u(t)$:

$$s(t) = \mathbf{a}^T \mathbf{s}(t-1) + u(t),\ t = 1, 2, \ldots, N. \qquad (1)$$

where $\mathbf{a} = [\alpha_1\ \alpha_2\ \cdots\ \alpha_p]^T$, $\mathbf{s}(t-1) = [s(t-1)\ \cdots\ s(t-p)]^T$, p is the filter order, and N is the frame length. The input signal(i.e. excitation) $u(t)$ is assumed to be non-Gaussian with a mixture distribution for voiced speech. A large part of the excitations comes from a normal distribution with a very small variance, while a small share of the excitations come from one with a much larger variance. This distribution is an example of heavy-tailed non-Gaussian distribution [6, 7]. The probability density function of this distribution may be expressed as

$$P_u = (1-\lambda)N(0,\ \sigma_1^2) + \lambda N(0,\ \sigma_2^2). \qquad (2)$$

where $\sigma_2^2 \gg \sigma_1^2$ and λ $(0 \le \lambda \le 1)$ is the probability for $\mathrm{N}(0,\ \sigma_2^2)$. The input signal is then modeled as

$$u(t) = (1-q(t))u_1(t) + q(t)u_2(t), \qquad (3)$$

where $q(t)$ is an i.i.d.(independent, identically distributed) random variable sequence, and $u_1(t)$ and $u_2(t)$ are mutually independent. $u_1(t)$ and $u_2(t)$ have zero-mean, and respectively have variances σ_1^2 and σ_2^2. The random variable $q(t)$ assumes the following probability distribution

$$\Pr[q(t)] = \begin{cases} \lambda, & q(t)=1, \\ 1-\lambda, & q(t)=0. \end{cases} \qquad (4)$$

The covariance of $u(t)$ is obtained from eq.(3) as follows,

$$E[u^2(t)] = Q(t) = (1-q(t))\sigma_1^2 + q(t)\sigma_2^2. \qquad (5)$$

Then the conditional probability density function of the speech signal $s(t)$ is

$$p(s(t)\,|\,\mathbf{a},\, q(t),\, \sigma_1^2,\, \sigma_2^2) = \frac{1}{\sqrt{2\pi Q(t)}} \exp \left[-\frac{\{s(t)-\mathbf{a}^T \mathbf{s}(t-1)\}^2}{2Q(t)} \right] \qquad (6)$$

The observed speech signal $z(t)$ can be expressed as follows,

$$z(t) = s(t) + \nu(t), \qquad (7)$$

where $\nu(t)$ is assumed as a white Gaussian noise with mean zero and variance σ_ν^2.

In some speech signal uncorrupted by noise is given, the conditional probability density function of the speech $z(t)$ is expressed using the characteristics of the measurement noise,

$$p(z(t)\,|\,s(t),\, \sigma_\nu^2) = \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \cdot \exp \left[-\frac{1}{2Q(t)} (s(t) - \mathbf{a}^T \mathbf{s}(t-1))^2 \right],\ t=1,\cdots,N. \qquad (8)$$

In order to estimate the clean speech signal $s(t)$, it is necessary to know the speech signal model parameters and the observation noise variance. We set these as an unknown variable vector $\Theta = \{\mathbf{a},\, \mathbf{q},\, \sigma_1^2,\, \sigma_2^2,\, \sigma_\nu^2,\, \lambda\}$. Here $\mathbf{q} = [q(1)\ q(2)\ \cdots\ q(N)]$.

When the observation signal $\mathbf{z} = [z(1)\ z(2)\ \cdots\ z(N)]$ is known, enhancement of the speech signal is equivalent to the estimation of the unknown parameter Θ. The maximum likelihood (ML) estimator for the estimation of Θ is then expressed as follows :

$$\Theta_{ml} = \arg\max_{\Theta}\ L(\Theta) = \arg\max_{\Theta}\ \log p(z|\Theta), \qquad (9)$$

where $L(\Theta)$ is a log likelihood function of Θ, and $p(z|\Theta)$ is the *a priori* probability density function of z given Θ. $p(z|\Theta)$ is expanded as

$$p(z|\Theta) = \prod_{t=1}^{N} p(z(t)\,|\,\Theta), \qquad (10)$$

Combining eqs.(1) and (7),

$$z(t) = \mathbf{a}^T\, \mathbf{s}(t-1) + u(t) + \nu(t). \qquad (11)$$

Since $u(t)$ and $\nu(t)$ are assumed to be mutually independent, and $p(z|\Theta)$ has zero mean, the covariance is expressed in Gaussian form as

$$C(t) = E \left[\{\mathbf{a}^T\, \mathbf{s}(t-1) + u(t) + \nu(t)\}\{\mathbf{a}^T\, \mathbf{s}(t-1) + u(t) + \nu(t)\}^T \right]$$
$$= \mathbf{a}^T E\left[\mathbf{s}(t-1)\mathbf{s}^T(t-1) \right] \mathbf{a} + (1-q(t))\sigma_\nu^2 + q(t)\sigma_2^2 + \sigma_\nu^2. \qquad (12)$$

Substituting the above result to eq. (10), the ML estimation problem is expressed as

$$\Theta_{ml} = \arg\max \left[-\frac{1}{2} \sum_{t=1}^{N} \frac{z^2(t)}{C(t)} - \frac{1}{2} \sum_{t=1}^{N} \log 2\pi C(t) \right] \qquad (13)$$

In general, eq.(13) is nonlinear and does not have simple ML solutions. In order to solve this problem we propose a method that uses the EM algorithm.

Ⅲ. Parameter Estimation Using the EM Algorithm

The EM algorithm is a general purpose iterative method that regards the observed data as incomplete and tries to solve maximum likelihood estimation problems. The EM algorithm in its general form may be found in [8]. This algorithm has guaraneteed convergence in many cases and can be adapted to suit various purposes.

For speech enhancement using the EM algorithm, the observed data z becomes the incomplete data, and $[\mathrm{z}\ \mathrm{s}]^T$ becomes the complete data. When the k-th iteration gives the parameter train $\{\Theta^0 \cdots \Theta^k\}$, the parameter estimation procedure at the next step can be carried out by repeating the E (estimation)-step and the M (maximization)-step in turn.

$$\text{E-step}: L(\Theta:\Theta^k) = E\Big[\log p(\mathrm{s},\, \mathrm{z}\,|\,\Theta)\,|\,\mathrm{z},\, \Theta^k\Big], \tag{14a}$$

$$\text{M-step}: \Theta^{k+1} = \arg\max_{\Theta}\Big[L(\Theta:\Theta^k)\Big], \tag{14b}$$

Beginning with the E-step we derive the detailed EM algorithm for speech enhancement.

– E-step

The joint probability $p(\mathrm{s},\, \mathrm{z}\,|\,\Theta)$ is written as

$$p(\mathrm{s},\, \mathrm{z}\,|\,\Theta) = p(\mathrm{z}\,|\,\mathrm{s},\, \Theta)\ p(\mathrm{s}\,|\,\Theta)$$

$$= \prod_{t=1}^{N} p(z(t)\,|\,\Theta)\ p(s(t)\,|\,\Theta). \tag{15}$$

Substituting eqs. (6) and (8) into eq.(15),

$$p(\mathrm{s},\, \mathrm{z}\,|\,\Theta) = (2\pi\sigma_v^2)^{N/2} \prod_{t=1}^{N}$$

$$\Big[2\pi\{(1-q(t))\sigma_1^2 + q(t)\sigma_2^2\}\Big]^{-1/2}$$

$$= \prod_{t=1}^{N}\ p(z(t)\,|\,s(t),\, \Theta)\ p(s(t)\,|\,\Theta). \tag{16}$$

The likelihood function can be written using eqs. (14a) and (16) with known Θ^k :

$$L(\Theta:\Theta^k) = -N\log 2\pi - \sum_{t=1}^{N} \log\Big[(1-q(t))\sigma_1^2 + q(t)\sigma_2^2\Big]$$

$$-\frac{1}{2}\sum_{t=1}^{N} \frac{1}{(1-q(t))\sigma_1^2 + q(t)\sigma_2^2}$$

$$\{E^k\Big[s^2(t))\Big] - 2\mathbf{a}^{\mathrm{T}} E^k\Big[s(t-1)s(t)\Big]$$

$$+\mathbf{a}^{\mathrm{T}} E^k\Big[\mathrm{s}(t-1)\mathrm{s}^{\mathrm{T}}(t-1)\Big]\mathbf{a}\} - \frac{N}{2}\log\sigma_v^2$$

$$-\frac{1}{2\sigma_v^2}\sum_{t=1}^{N}\{E^k\Big[z^2(t)\Big] - 2z(t)E^k\Big[s(t)\Big] + E^k\Big[s(t)\Big]\}, \tag{17}$$

where $E^k[\,\cdot\,] = E^k[\,\cdot\,|\,\mathrm{z},\ \Theta^k]$. This E-step includes a $p \times p$ matrix $E^k[\mathrm{s}(t-1)\mathrm{s}^T(t-1)]$, vectors $E^k[\mathrm{s}(t-1)s(t)]$, $E^k[s(t)]$ and $E^k[s^2(t)]$. These conditional expectations can be computed using a modified Kalman filter that utilizes the parameter Θ^k obtained at the k-th interation [8]. For Kalman filter realization, we rewrite eqs. (1) and (7) in state space form [9] as

$$\mathrm{s}_p(t) = \Phi\mathrm{s}_p(t-1) + \mathbf{G}u(t), \tag{18}$$

$$z(t) = \mathbf{H}^T\ \mathrm{s}_p(t) + v(t), \tag{19}$$

where s_p is a $(p+1)\times 1$-dim. state vector $[s(t-p) \cdots s(t)]$, Φ is a $(p+1)\times(p+1)$-dim. matrix $\Phi = \begin{bmatrix} 0 & \mathbf{I} \\ 0 & a_1 \cdots a_p \end{bmatrix}$, $\mathbf{H}^T = [0 \cdots 0\ 1]$, and $\mathbf{G}$ a $(p+1) \times 1$-dim. vector. With the parameter vectors Θ^k and $\{q^k(t) = i,\ i = 0,\ 1\}$ given, the estimation equations for the state vector are, by robust Kalman filtering [11, 12].

$$\hat{\mathrm{s}}_{\mathrm{p}}(t) = \Phi\hat{\mathrm{s}}_{\mathrm{p}}(t-1) + \{(1-i)K_0(t)$$

$$+iK_1(t)\}(z(t) - \mathbf{H}^T\,\Phi\hat{\mathrm{s}}_{\mathrm{p}}(t-1)), \tag{20}$$

$$\mathbf{K}_i(t) = \mathbf{P}_i(t\,|\,t-1)\mathbf{H}\{\mathbf{R} + \mathbf{H}^T\mathbf{P}_i(t\,|\,t-1)\mathbf{H}\}^{-1}, \tag{21}$$

$$\mathbf{P}_i(t\,|\,t-1) = \Phi\mathbf{P}_i(t-1)\Phi^T + \mathbf{G}\mathbf{Q}_i\mathbf{G}^T, \tag{22}$$

$$\mathrm{P}(t) = \mathbf{P}_i(t\,|\,t-1) - \{(1-i)\mathrm{K}_0(t) - i\mathrm{K}_1(t)\}\mathrm{H}^T\,\mathbf{P}_i(t\,|\,t-1), \tag{23}$$

where $\mathbf{K}_i(t)$ is Kalman gain vector, and $\mathbf{P}(t\,|\,t-1)$ is the covariance matrix of the *a priori* error. $\mathbf{R} = \sigma_v^{2k}\mathbf{I}$ is the covariance matrix of the observed noise and $\mathbf{Q}_i = \{(1-i)\,\sigma_1^2 + i\sigma_2^2\}\ \mathbf{I}$ is the covariance matrix of the speech source.

The conditional expectations that appear in eq. (17), $E^k[s(t)]$ and $E^k[s(t-1)s^T(t-1)]$, can be obtained as follows according to eqs.(20)-(23).

$$E^k[s(t)] = \mathbf{H}^T\ \hat{\mathrm{s}}_{\mathrm{p}}(t), \tag{24}$$

$$E^k\{s(t-1)s^T(t-1)\} = \mathbf{P}_1(t|t-1)$$

$$\mathbf{H}\{\mathbf{H}^T \mathbf{P}_1(t|t-1)\mathbf{H} + \mathbf{R}_t\}^{-1}$$

$$\cdot \mathbf{H}^T \mathbf{P}_t(t|t-1) + \hat{\mathbf{s}}_p^T(t)\hat{\mathbf{s}}_p(t). \tag{25}$$

— M-step

To obtain a new parameter value Θ^{k+1}, eq.(17) must be maximized with respect to its parameters. As can be seen in eq.(17), we can divide the parameter vector into q and $\Theta_1 = \{\mathbf{a}, \mathbf{q}, \sigma_1^2, \sigma_2^2, \sigma_v^2\}$. The maximization of eq.(14) must be carried out with respect to Θ, which is in general too complex. This paper proposes an effective algorithm that repeats detection of q and estimation of Θ_1 to maximize eq.(14b).

First if Θ_1^k is given, the problem of detecting q is equal to maximizing eq.(14b) with respect to q. Second, regarding q^{k+1} as obtained from the k-th detection step, follows the estimation of Θ_1 by maximizing eq.(14b) with respect to Θ_1. Then the M-step in eq.(14b) is divided into two steps as shown below:

$$\text{detection step}: q^{k+1} = \max_q L(q : \Theta_1^k), \tag{26a}$$

$$\text{estimation step}: \Theta_1^k = \max_{\Theta_1} L(\Theta_1 : q^{k+1}, \Theta_1^k). \tag{26b}$$

These two steps are explained in detail below

— detection step

When q is unknown and Θ_1 is known, $L(q : \Theta_1^k)$ is a target function for detecting q. Fixing Θ_1 in eq.(17) to known value and removing those terms that appear constant with respect to q,

$$L(q : \Theta_1^k) = -\sum_{t=1}^{N} \log\{(1-q(t))\sigma_1^{2^k} + q(t)\sigma_2^{2^k}\}$$

$$-\frac{1}{2}\sum_{t=1}^{N} \frac{E^k[(s^2(t) - \mathbf{a}^{k'} s(t-1))^2]}{(1-q(t))\sigma_1^{2^k} + q(t)\sigma_2^{2^k}}. \tag{27}$$

Substituting eq.(27) into eq.(26a), the function for the detection of q becomes

$$\max_q \Big[-\sum_{t=1}^{N} \log\{(1-q(t))\sigma_1^{2^k} + q(t)\sigma_2^{2^k}\}$$

$$-\frac{1}{2}\sum_{t=1}^{N} \frac{1}{(1-q(t))\sigma_1^{2^k} + q(t)\sigma^{2^k}}$$

$$\cdot\{E^k[s^2(t)] - 2\mathbf{a}^{T^k}E^k[\mathbf{s}(t-1)s(t)]$$

$$+ \mathbf{a}^{T^k}E^k[\mathbf{s}(t-1)\mathbf{s}^T(t-1)]\mathbf{a}^k\}\Big]. \tag{28}$$

Using λ^k obtained at the k-th iteration, we can obtain an optimal q, among 2^N possible values of q, that maximizes eq. (28). But employing this method to all 2^N candidates in all iterations is not practical, which necessitates a technique that finds a locally optimum value of q. This may fall a little short in accuracy, but requires far less computation than looking for globally optimum value of q.

We set $u^k(t)$ as $E^k[s(t) - \mathbf{a}^{T^k}\mathbf{s}(t-1)]$. Detection of q^{k+1} from this signal is regarded as a classic detection problem. The parameter q^{k+1} can easily be obtained by a threshold detector using a likelihood ratio test. Given $u^k(t)$, the likelihood function for the detection of $q(t)$ becomes

$$L(q(t)\,|\,u^k(t)) = p(u(t)\,|q(t))p(q(t)). \tag{29}$$

From eq.(6),

$$p(u^k(t)\,|q(t)) = \frac{1}{\sqrt{2\pi\{(1-q(t))\sigma_1^2 + q(t)\sigma_2^2\}}}$$

$$\cdot \exp\Big[-\frac{u^{k^2}}{2\{(1-q(t))\sigma_1^2 + q(t))\sigma_2^2\}}\Big]. \tag{30}$$

Letting $\Lambda(t)$ denote the likelihood ratio for detecting q^{k+1} given $u^k(t)$,

$$\Lambda(t) = \frac{L(q(t)=1 : u^k(t))}{L(q(t)=0 : u^k(t)}$$

$$= \frac{p(u^k(t)q(t)=1)p(q(t)=1)}{p(u^k(t)q(t)=0)p(q(t)=0)} \underset{q^{k+1}(t)=1}{\overset{q^{k+1}(t)=0}{\lessgtr}} 1. \tag{31}$$

The ML threshold detector for q^{k+1} is expressed as

$$q^k(t) \underset{q^{k+1}(t)=0}{\overset{q^{k+1}(t)=1}{\lessgtr}} \left(\ln \frac{\sigma_1^2}{\sigma_2^2} - 2\ln \frac{\lambda^k}{1-\lambda^k}\right),$$

$$t = 1, \ldots, N \tag{32}$$

-estimation step

At this step, we assume that the parameter Θ_1 is unknown and the parameter q is known. Let the function for paramenter estimation be $L(\Theta_1 ; \Theta_1^k, q^{k+1})$. When we fix q in eq.(26b) to the known value of q^{k+1}, we can get the following equation

$$L(\Theta_1 ; \Theta_1^k, q^{k+1}) = -\sum_{t=1}^{N} \log\{(1-q^{k+1}(t))\sigma_1^2 + q^{k+1}(t)\sigma_1^2\}$$
$$-\frac{1}{2}\sum_{t=1}^{N} \frac{E^k[(s(t)-\mathbf{a}^t\,\mathbf{s}(t-1)^2]}{(1-q^{k+1}(t))\sigma_1^2 + q^{k+1}(t))\sigma_2^2} - \frac{N}{2}\log\sigma_v^2$$
$$-\frac{1}{2\sigma_v^2}\sum_{t=1}^{N}\{E^k[z^2(t)] - 2z(t)E^k[s(t)] + E^k[s^2(t)]\}. \tag{33}$$

We remove the terms that are constant with respect to q^{k+1} since they do not contribute to the maximization process. Substituting eq. (33) into eq. (26b) and estimating the parameters,

$$a^{k+1} = \left(\sum_{t=1}^{N}\frac{E^k[\mathbf{s}(t-1)\mathbf{s}^t(t-1)]}{W^k(t)}\right)$$
$$\cdot \sum_{t=1}^{N}\frac{E^k[\mathbf{s}(t-1)s(t)]}{W^k(t)}, \tag{34}$$

$$(\sigma_1^2)^{k+1} = \sum_{t=1}^{N}\frac{1}{1-q^{k+1}(t)}\{(1-q^{k+1}(t))E^k[s^2(t)]$$
$$+\mathbf{a}^{t^k}(1-q^{k+1}(t))E^k[\mathbf{s}(t-1)s(t)], \tag{35}$$

$$(\sigma_2^2)^{k+1} = \sum_{t=1}^{N}\frac{1}{q^{k+1}(t)}$$
$$+\mathbf{a}^{t^k}\sum_{t=1}^{N} q^{k+1}(t)E^k[\mathbf{s}(t-1)s(t)], \tag{36}$$

where $W^k(t) = (1-q^{k+1}(t))\sigma_1^{2^k} + q^{k+1}(t)\sigma_2^{2^k}$. Substituting eq. (33) into eq. (26b), the covariance of the observed noise is

$$(\sigma_v^2)^{k+1} = \max$$
$$\left\{-\frac{N}{2}\log\sigma_v^2 - \frac{1}{2\sigma_v^2}\sum_{t=1}^{N}(z(t)-E^k[s(t)])^2\right\} \tag{37}$$

Therefore σ_v^2 becomes

$$(\sigma_v^2)^{k+1} = \frac{1}{N}\sum_{t=1}^{N}\{z(t) - E^k[s(t)]\}^2. \tag{38}$$

Also, when q^{k+1} is known, the probability for $\{q(t) = 1\}$ is

$$\lambda^{k+1} = \frac{1}{N}\sum_{t=1}^{N} q^{k+1}(t). \tag{39}$$

A speech signal model has been proposed that can be used regardless of the voiced/unvoiced classification. In this regard this model differs from other conventional speech enhancement schemes. The EM algorithm is employed for ML parameter optimization, and a robust Kalman filter is used for speech enhancement.

Ⅳ. Experimental Results

In order to show the performance of the proposed speech enhancement algorithm, real and synthesized speech signals with additive white Gaussian observation noise were used. The noise was added such that 5 different input SNRs were available : 0, 5, 10, 20 and 60dB. The speech signal was collected from a male speaker with 10 kHz sampling rate.

For synthetic speech signals, parameter values in Θ and the SNR of the speech after processing are compared with those using conventional Gaussian assumption. The performance tests are conducted for variance ratios of the two mixed excitation source and occurrence probability b. The variance ratios are 5 and 10, and the values of b are 0.01, 0.02, 0.05 and 0.1. The ratio of the occurrence probability is equal to the pitch period. For real speech segment /a/ obtained in a noise-free environment, the performance comparison is made by means of post-enhancement SNR. The results

Table 1. Comparisons of SNRs of synthetic speech(C : conventional, P : proposed), (a) for variance ratio of 5.

input SNR	0dB		5dB		10dB		20dB		noise-free	
b	C	P	C	P	C	P	C	P	C	P
0.01	6.0	8.7	9.1	11.2	11.0	15.4	16.5	22	20.0	24
0.02	6.0	8.7	9.1	11.2	11.0	15.4	16.5	22	20.0	24
0.05	5.8	8.7	8.4	11.2	10.2	15.4	16.0	22	19.2	24
0.1	5.6	8.7	8.0	11.2	10.0	15.4	15.5	22	19.6	24

(b) for variance ration of 10

input SNR	0dB		5dB		10dB		20dB		noise-free	
b	C	P	C	P	C	P	C	P	C	P
0.01	5.7	8.7	8.5	11.2	10.2	15.4	15.2	22	19.0	24
0.02	5.7	8.7	8.5	11.2	10.3	15.4	15.1	22	19.0	24
0.05	5.2	8.7	8.0	11.2	9.5	15.4	14.9	22	18.5	24
0.1	5.0	8.7	7.8	11.2	9.5	15.4	14.7	22	18.2	24

Table 2. SNR improvement of real speech segment /a/.

input SNR	0	5	10	20	60 (dB)
conventional	5.8	8.2	11.2	16.5	20
proposed	8.5	10.2	14	21.5	23.8

are summarized in Tables 1 and 2. We could see that the proposed algorithm was not influenced by changing values of *b* and the variance ratios, whereas the conventional method was.

The excitation model we proposed is a generalized one, and Gaussian form can be derived with ease. This algorithm can be extended to the case where more than two excitation source components considered.

We developed a speech signal model on a two-mixture excitation source, and used the EM algorithm to estimate and enhance the speech parameters. Through computer experiments we showed the performance of the proposed algorithm. We could see that the proposed technique is better than the conventional method based on Gaussian assumption.

References

1. S. F. Ball and D. C. Pulsipher, "Suppression of acoustic noise in speech using two microphone adaptive noise cancellation," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. ASSP-28, pp. 725-753, Dec. 1980.
2. W. A. Harrison, J. S. Lim and E. Singer, "A new application of adaptive noise cancellation," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-34, pp. 21-27, Feb. 1986.
3. J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-26, pp. 197-210, June, 1978.
4. J. Hasen and M. A. Clements, "Constrained iterative speech enhancement with application to speech recognition," *IEEE Trans. Signal Proc.*, vol. 39, pp. 795-805, Apr. 1991.
5. Y. Ephraim, D. Malah and B. -H. Juang, "On the application of hidden Markov models for enhancing noisy speech," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-27, pp. 1846-1856, Dec. 1989.
6. Y. Ephraim, "Statistical-model based speech enhancement systems," *Proc IEEE*, vol. 80, pp. 1526-1555, Oct. 1992.

7. C. H. Lee, "On robust linear prediction of speech," *IEEE Trans. Acoust. Speech and Signal Proc.*, vol. ASSP-36, pp. 642-650, May, 1988.

8. A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B*, vol. 39, pp. 1-38, 1977.

9. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.

10. M. Feder and E. Weinstein, "Parameter estimation of superimposed signal using the EM algorithm," *IEEE Trans. Acoust. Speech, and Signal Proc.*, vol. ASSP-36, pp. 477-489, Apr. 1988.

11. B.-G. Lee, K. Y. Lee and S. Ann, "A sequential algorithm for robust parameter estimation and enhancement of noisy speech," *Proc. ISCAS*, pp. I. 243-246, Chicago, 1993.

12. B.-G. Lee, K. Y. Lee and S. Ann, "An EM-based approach for speech parameter estimation and enhancement," *Signal Processing*, (revised)

• This paper was supported (in part) by NON-DIRECTED RESEARCH FUND, Korea Research Foundation, 1993.

▲Ki Yong Lee

1983년 B. Sc. : (Electronics Engineering), Soongsil University, Seoul

1985년 M. S. : (Electronics Engineering), Seoul National University, Seoul

1991년 Ph. D. : (Electronics Engineering), Seoul National University, Seoul

Sep. 1991년 ~ : Assistant Professor, Department of Electronics Engineering, Changwon National University, Changwon, Kyungbook

research interests : speech analysis and enhancement, statistical signal processing, and nonlinear adaptive signal processing.

▲Young-Tae Kang

1991년 B. Sc. : (Electronics Engineering), Changwon National University

1994년 ~ : pursuing M. S. degree in Electronics Engineering at the Dept. of Electronics Engineering, Changwon National University.

research interests : statistical signal processing and digital communication

▲Byung-Gook Lee

1988년 B. Sc. : (Electronics Engineering), Seoul National University

1990년 M. S. : (Electronics Engineering), Seoul National University

1990년 ~ : pursuing Ph. D. degree at the Department of Electronics Engineering, Seoul National University

research interests : speech enhancement, nonlinear signal processing