시변가산유색잡음하의 음성 향상을 위한 효율적인 Mixture IMM 알고리즘

Efficient Mixture IMM Algorithm for Speech Enhancement under Nonstationary Additive Colored Noise

이 기 용*, 임 재 열**

(Ki Yong Lee*, Jae Yeol Rheem**)

※ 본 연구는 1998년도 숭실대학교 교내 연구비의 지원에 의하여 수행된 연구 결과입니다

요 약

본 논문에서는 시변가산유색잡음에 오염된 음성신호의 향상을 위한 MIMM(mixture interacting multiple model) 알고리즘을 제안한다. 제안된 방법에서 음성신호는 혼합 은닉필터모델(hidden filter model: HFM)로 모델링되며, 잡음신호는 하나의 은닉필터로 모델링 된다. MIMM 알고리즘은 혼합은닉필터모델에의한 다중 Kalman 필터링에 기초한 회귀계산이기 때문에 계산량이 많아, Kalman 필터링 식의 구조적 측면에서 효율적인 계산이 가능하도록 알고리즘을 구현 했다. 시뮬레이션 결과, 제안된 방법이 기존의 결과 [4,5]에 비하여 성능향상이 이루어 졌음을 보여준다.

ABSTRACT

In this paper, a mixture interacting multiple model (MIMM) algorithm is proposed to enhance speech contaminated by additive nonstationary noise. In this approach, a mixture hidden filter model (HFM) is used to model the clean speech and the noise process is modeled by a single hidden filter. The MIMM algorithm, however, needs large computation time because it is a recursive method based on multiple Kalman filters with mixture HFM. Thereby, a computationally efficient implementation of the algorithm is developed by exploiting the structure of the Kalman filtering equation. The simulation results show that the proposed method offers performance gain compared to the previous results in [4,5] with slightly increased complexity.

I. 서 론

최근에 음성향상(speech enhancement)을 위하여 은닉필터모델(hidden filter model) [1]과 IMM (interacting multiple model) 방법 [2]에 근거한 새로운 음성향상 방법이 제안 되었다 [3-5]. 여기서 음성 및 잡음 신호는 시변 AR 과정으로 모델링되며, 이 AR 모델의 파라미터는 Markov 체인의 상태에 의존한다 [4,5]. IMM 방법은 Markovian 계수를 갖는 다중 선형 시스템에 대한 Kalman 필터링 방법을 이용한 것으로, 깨끗한 음성과 잡음에 대하여 훈련된 각각의 HFM이 주어지면, 회기추정(recursive estimation) 방법을 이용하여 시변가산잡음에 의하여 오염된 잡음음성으로부터 원래의 깨끗한 음성을 추정하게 된다.

본 논문에서는 기존의 방법 [3-5]에서 음성신호를 위하여 상태당 하나의 Gaussian AR 모델로 HFM을 구성한 것을 확장하여, 음성신호에 대하여 상태당 여러 개의 Gaussian AR 모델의 혼합으로 이루어지는 mixture HFM을 제안한다. 음성신호에 대한 mixture HFM 파라미터와 잡음에 대한 하나의 HFM 파라미터가 주어지면, mixture HFM에 대한 MIMM방법을 유도할 수 있다. 기존에 제안된 방법은 본 논문에서 제안된 MIMM 방법에 대하여 mixture의 개수가 1인 특별한 경우에 해당된다. 따라서 본 논문에서 제안된 MIMM 방법은 IMM 방

*숭실대학교 정보통신 전자공학부
**한국기술교육대학교 전자공학과
접수일자 : 1999년 4월 12일

법의 일반화로 볼 수 있다. 음성신호에 대하여 HFM모델을 mixture개념으로 확장함에 따라 시스템의 복잡성 및 메모리 요구량이 크게 증가한다. 따라서 시스템의 복잡성과 메모리 요구량을 줄이기 위하여 계산량 측면에서 효율적인 알고리즘을 개발하였다. 실제의 자동차 잡음을 이용한 시뮬레이션 실험 결과, 기존의 방법 [4-5]에 비하여, 제안된 방법이 계산량이 증가하지만 약 0.4-0.7dB 향상됨을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 mixture HFM에 의한 음성신호 모델링, HFM에 의한 잡음 모델링을 설명하며, 3장에서는 음성향상을 위하여 제안된 MIMM방법을 유도 설명한다. 제안된 방법의 성능비교를 위한 실험결과는 4장에서 다루며, 5장에서 결론을 맺는다.

II. 음성 및 잡음 모델링

잡음에 오염되기 이전의 깨끗한 음성신호가 시각 t에서 상태 $s(t)\in\{1,2,\cdots,L\}$과 각 상태에 대한 가우시안 AR 프로세스들의 mixture $m(t)\in\{1,2,\cdots,M\}$에 의해 mixture HFM로 모델링 된다고 가정하자. 이때 상태수는 L, mixture의 개수는 M이다. 또 음성신호의 상태가 1차 Markov체인으로 모델링 된다고 가정하고 상태 전이 확률, mixture 확률을 각각 $a_{s(t-1)s(t)}, c_{m(t)s(t)}$라 하면, 잡음에 오염되기 전의 음성신호는 식 (1)과 같이 모델링되어 나타낼 수 있다.

$$y(t)=B^T_{m(t)s(t)}y(t-1)+e_{m(t)s(t)}(t) \tag{1}$$

여기서 $B_{m(t)s(t)}=[b_{m(t)s(t)}(1),\cdots b_{m(t)s(t)}(p)]^T$는 AR 계수 벡터, $y(t-1)=[y(t-1),\cdots,y(t-p)]^T$는 p개의 관측열 벡터, 그리고 $e_{m(t)s(t)}(t)$는 드라이빙열(driving sequence)로 평균 0, 분산 $\sigma^2_{m(t)s(t)}$인 가우시안 프로세스이다.

음성신호에 대한 mixture HFM의 파라미터는 $\lambda_y=\{a,c,B,\sigma^2\}$로 나타낼 수 있으며, 여기서 $a=\{a_{ij}\},c=\{c_{jk}\},B=\{B_{jk}\},\sigma^2=\{\sigma^2_{jk}\}$이며, $i,j=1,2,\cdots,L$ 그리고 $k=1,2,\cdots,M$으로 Baum-Welch 알고리즘이나 segmental K-means 알고리즘을 이용하여 잡음이 섞이지 않은 음성의 학습으로 추정된다 [5].

시변유색잡음(time varying colored noise)을 모델링하기 위해, 잡음의 상태 $h(t)\in\{1,2,\cdots,M\}$로 HFM 모델에 의하여 잡음의 적절한 표현이 가능하다고 가정하면, 식 (2)와 같이 나타낼 수 있으며,

$$v(t)=C^T_{h(t)}v(t-1)+w_{h(t)}(t) \tag{2}$$

여기서 $v(t-1)=[v(t-1),\cdots,v(t-q)]^T, C_{h(t)}=[c_{h(t)}(1),\cdots,c_{h(t)}(q)]^T$, 이고 $w_{h(t)}(t)$는 평균 0 분산 $\sigma^2_{w,h(t)}$인 Gaussian 프로세스이다. 잡음에 대한 HFM 파라미터 $\lambda_v=\{\bar{a},C,\sigma^2_w\}$는 주어진 학습용 잡음을 이용하여 역시 Baum-Welch 알고리즘에 의하여 추정되며, 여기서 $\bar{a}=\{\bar{a}_{ij}\},C=\{C_j\},\sigma^2_w=\{\sigma^2_{w,j}\}$, $i,j=1,2,\cdots,N$이다.

III. MIMM을 이용한 음성향상

실제 마이크에 의해 측정된 음성신호는 주변잡음이 섞이게 되어 식 (3)과 같이 나타낼 수 있다.

$$z(t)=y(t)+v(t). \tag{3}$$

여기서 음성향상은 잡음이 섞인 음성신호 $z(t)$로부터 잡음이 섞이지 않은 깨끗한 음성신호인 $y(t)$ 열을 추정하는 문제가 되며, Kalman필터를 이용한 최소평균자승오차(MMSE: minimum mean square error) 추정을 고려하자. 음성신호에 대한 mixture HFM의 파라미터와 잡음에 대한 한 개의 HFM의 파라미터가 각각 주어졌다고 가정하면, 다음과 같은 형태로 혼성상태(composite states) $\theta(t)=\{m(t),s(t),h(t)\}$를 이용하여 state-space 모델을 구성할 수 있다.

$$x(t)=F(\theta(t))x(t-1)+Ge(\theta(t)) \tag{4}$$

$$z(t)=H^Tx(t) \tag{5}$$

여기서

$$x(t)=[y^T(t)v^T(t)]^T,$$

$$F(\theta(t))=\begin{bmatrix}\Phi(m(t),s(t)) & 0\\ 0 & F_v(h(t))\end{bmatrix}$$

$$\Phi(m(t),s(t))=\begin{bmatrix}B^T_{m(t)s(t)}\\ I\quad 0\end{bmatrix}$$

$$F_v(h(t))=\begin{bmatrix}C^T_{h(t)}\\ I\quad 0\end{bmatrix}$$

$$e(\theta(t))=\left[e_{m(t),s(t)}(t)w_{h(t)}(t)\right]^T,$$

$$G=\begin{bmatrix}g_y & 0\\ 0 & g_v\end{bmatrix},$$

$$H=\left[H_y^TH_v^T\right]^T, H_y=g_y=[10\cdots0]^T, H_v=g_v=[10\cdots0]^T$$

그리고 음성신호 및 유색잡음의 드라이빙 열 $e_{m(t)s(t)}(t)$와 $w_{h(t)}(t)$가 서로 상관관계가 없다고(uncorrelated) 가정하면 다음과 같다.

$$Q(\theta(t))=E\left\{e(\theta(t))e^T(\theta(t))\right\}=\begin{bmatrix}\sigma^2_{m(t)s(t)} & 0\\ 0 & \sigma^2_{w,h(t)}\end{bmatrix} \tag{6}$$

잡음이 섞인 음성 $z(t)=\{z(1),\cdots,z(t)\}$이 주어졌을 경우, 상태 $x(t)$에 대한 MMSE 추정은 다음과 같이 조건평

균이 된다.

$$\hat{x}(t) = E\{x(t) \mid z(t)\}. \tag{7}$$

시각 t의 복합상태 $\{m(t)=i, s(t)=j, h(t)=k\}$를 $\theta_{o_t}(t)$로 나타내면, $o_t = ijk$가 되며, 조건평균 $E\{x(t) \mid z(t)\}$ 은 다음과 같이 모든 조건평균의 가중치 합으로 계산이 가능하다.

$$E\{x(t) \mid z(t)\} = \sum_{o} E\{x(t) \mid Hist(t,o), z(t)\} p(Hist(t,o) \mid z(t)) \tag{8}$$

여기서 $Hist(t,o) = \{\theta_{o_0}(0), \theta_{o_1}(1), \cdots, \theta_{o_t}(t)\}, o = (o_0, o_1, \cdots, o_t)$는 시각 t까지의 상태의 구체적인 연역(history)을 나타내게 되며, 이것은 모든 가능한 복합상태열 공간에서 모델에 해당하는 특정 열벡터가 된다.

식 (1)에서 조건추정 $E\{x(t) \mid Hist(t,o), z(t)\}$는 식 (4)와 (5)에 근거하여 복합 상태 열 $Hist(t,o)$에 해당되는 시스템 파라미터에 대한 MMSE 추정기를 구성함으로써 얻을 수 있다. 즉, $E\{x(t) \mid Hist(t,o), z(t)\}$는 $Hist(t,o)$에 해당되는 Kalman 필터로부터 구할 수 있다. 그 결과 최적 추정기 (8)은 $(L \times M \times N)^t$ 개의 추정 $E\{x(t) \mid Hist(t,o), z(t)\}$의 가중치 합으로 나타나며, 따라서 시간에 따라서 지수적으로 증가하는 메모리와 계산량이 요구된다. 결과적으로 최적해는 매우 복잡하여 실질적으로 계산이 가능하기 위해서는 계산 및 메모리 요구량을 줄이기 위한 일련의 근사방법에 대한 연구가 필요하다.

위의 문제를 해결하기 위하여 본 논문에서는 mixture IMM 방법을 제안한다. 제안된 mixture IMM 방법은 크게 상호작용(interaction), 필터링(filtering), 모드갱신(mode-update), 결합(combination)의 4 단계로 구성되어진다. 회귀 계산의 매 싸이클마다, 이전 시각의 모든 복합상태를 지닌 필터의 상태 추정에 대한 상호작용을 이용하여 특정 모드에 해당되는 필터의 초기조건을 구하며, 이때 특정 모드는 현재 시각에 유효하다고 가정한다. 다음에는 각 모드에 대하여 병렬로 필터링 단계를 거친다. 그리고 모든 필터에 대한 갱신된 상태 추정에 대한 결합으로 상태 추정을 하게 된다. 이 과정에서 모드에 대한 확률은 상태와 공분산의 상호작용과 결합에 대한 가중치로써 실질적인 핵심 역할을 하게 된다.

제안된 알고리즘은 식 (4)와 (5)의 형태로 주어지는 $(L \times M \times N)$개의 서로 다른 모델로 구성된 $(L \times M \times N)$개의 Kalman filter에 기초하게 된다.

시각 t에서 상호작용 단계(interaction step)에 대한 입력은 이전 시각 t-1에서의 추정기로 다음과 같다:

$$\hat{x}_{\alpha\beta\gamma}(t-1) = E\{x(t-1) \mid \theta_{\alpha\beta\gamma}(t-1), z(t-1)\}, \tag{9}$$
$$1 \le \alpha \le M, 1 \le \beta \le L, 1 \le \gamma \le N.$$

여기에 다음과 같은 근사를 하면,

$$p\{x(t-1) \mid \theta_{\alpha\beta\gamma}(t-1), z(t-1)\} \sim N(\hat{x}_{\alpha\beta\gamma}(t-1), P_{\alpha\beta\gamma}(t-1)), \tag{10}$$

여기서 $P_{\alpha\beta\gamma}(t-1)$은 $\hat{x}_{\alpha\beta\gamma}(t-1)$에 대한 에러 공분산(error covariance)이다.

상호작용 단계의 출력은 $(L \times M \times N)$ 개의 추정기로

$$\begin{aligned} x^0_{ijk}(t-1) &= E\{x(t-1) \mid \theta_{ijk}(t), z(t-1)\} \\ &= \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} E\{x(t-1) \mid \theta_{ijk}(t), \theta_{\alpha\beta\gamma}(t-1), z(t-1)\} \\ &\cdot P\{\theta_{\alpha\beta\gamma}(t-1), \mid \theta_{ijk}(t), z(t-1)\} \end{aligned} \tag{11}$$

여기서 $P\{\theta_{\alpha\beta\gamma}(t-1) \mid \theta_{ijk}(t), z(t-1)\}$는 믹싱확률(mixing probability)이다.

만일 $(\theta_{\alpha\beta\gamma}(t-1))$가 알려지면, $x(t-1)$은 $(\theta_{ijk}(t))$에 대하여 독립적이어서 다음을 얻을 수 있다.

$$\begin{aligned} x^0_{ijk}(t-1) = \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} E\{x(t-1) \mid \theta_{\alpha\beta\gamma}(t-1), z(t-1)\} \cdot \\ P\{\theta_{\alpha\beta\gamma}(t-1), \mid \theta_{ijk}(t), z(t-1)\} \end{aligned} \tag{12}$$

식 (12)로부터, 식 (13)과 같이 이전 시각에 대한 모든 필터의 믹싱 추정기(mixing estimates)에 의하여 $(\theta_{ijk}(t))$에 해당하는 필터에 대한 초기조건을 생성할 수 있다.

$$x^0_{ijk}(t-1) = \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} \hat{x}_{\alpha\beta\gamma}(t-1) P\{\theta_{\alpha\beta\gamma}(t-1) \mid \theta_{ijk}(t), z(t-1)\}. \tag{13}$$

해당 공분산은 유사한 방법으로 계산되어져 다음을 얻을 수 있다.

$$\begin{aligned} P^0_{ijk}(t-1) = \sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N} P\{\theta_{\alpha\beta\gamma}(t-1) \mid \theta_{ijk}(t), z(t-1)\} \\ \cdot [P_{\alpha\beta\gamma}(t-1) + (\hat{x}_{\alpha\beta\gamma}(t-1) - x^0_{\alpha\beta\gamma}(t-1)) \\ \cdot (\hat{x}_{\alpha\beta\gamma}(t-1) - x^0_{\alpha\beta\gamma}(t-1))] \end{aligned} \tag{14}$$

그러면 확률 $P\{\theta_{\alpha\beta\gamma}(t-1) \mid \theta_{ijk}(t), z(t-1)\}$은 다음과 같이 주어진다.

$$\begin{aligned} P\{\theta_{\alpha\beta\gamma}(t-1) \mid \theta_{ijk}(t), z(t-1)\} \\ = \frac{P\{\theta_{ijk}(t) \mid \theta_{\alpha\beta\gamma}(t-1), z(t-1)\} P\{\theta_{\alpha\beta\gamma}(t-1) \mid z(t-1)\}}{P\{\theta_{ijk}(t) \mid z(t-1)\}} \end{aligned} \tag{15}$$

시각 t에서 혼성 상태 $\theta_{ijk}(t)$에서 음성 신호 상태 $\{m(t)=i, s(t)=j\}$와 잡음 상태 $h(t)=k$는 상호 독립적이므로, $P\{\theta_{ijk}(t) \mid \theta_{\alpha\beta\gamma}(t-1), z(t-1)\}$를 다음과 같이 구성할 수 있다.

$$
\begin{aligned}
&P\left\{\theta_{ijk}(t)\mid\theta_{\alpha\beta\gamma}(t-1),z(t-1)\right\}\\
&=P\left\{m_i(t),s_j(t)\mid\theta_{\alpha\beta\gamma}(t-1),z(t-1)\right\}P\left\{h_k(t)\mid\theta_{\alpha\beta\gamma}(t-1),z(t-1)\right\}\\
&=P\left\{m_i(t),s_j(t)\mid m_\alpha(t-1),s_\beta(t-1),z(t-1)\right\}P\left\{h_k(t)\mid h_\gamma(t-1),z(t-1)\right\}\\
&=P\left\{s_j(t)\mid s_\beta(t-1),z(t-1)\right\}P\left\{m_i(t)\mid s_j(t),z(t-1)\right\}P\left\{h_k(t)\mid h_\gamma(t-1),z(t-1)\right\}\\
&=a_{\beta j}c_{ij}\bar{a}_{\gamma k}.
\end{aligned}
$$

그러면, 식 (15)의 믹싱 확률은 다음과 같이 다시 쓸 수가 있다.

$$
P\left\{\theta_{\alpha\beta\gamma}(t-1)\mid\theta_{ijk}(t),z(t-1)\right\}
=\frac{a_{\beta j}c_{ij}\bar{a}_{\gamma k}P\left\{\theta_{\alpha\beta\gamma}(t-1)\mid z(t-1)\right\}}{\sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N}a_{\beta j}c_{ij}\bar{a}_{\gamma k}P\left\{\theta_{\alpha\beta\gamma}(t-1)\mid z(t-1)\right\}}. \tag{16}
$$

이러한 추정기는 혼성상태하에서 Kalman 필터에의한 예측(prediction) 및 측정(measurement)의 기본이 된다.

필터링 단계(filtering step)에서는 각 $\left(x_{ijk}^0(t-1),P_{ijk}^0(t-1)\right)$ 쌍이 혼성 상태 $\theta_{ijk}(t)$에 해당되는 Kalman 필터의 입력으로 사용되며, 출력은 $\left(\hat{x}_{ijk}(t),P_{ijk}(t)\right)$으로 표현된다. $\hat{x}_{ijk}(t)$와 $P_{ijk}(t)$은 $(L\times M\times N)$개의 병렬 Kalman 필터에 의하여 다음과 같이 얻어진다:

$$
\hat{x}_{ijk}(t)=F\left(\theta_{ijk}(t)\right)x_{ijk}^0(t-1)+K_{ijk}(t)\left\{z(t)-H^TF\left(\theta_{ijk}(t)\right)x_{ijk}^0(t-1)\right\}, \tag{17}
$$

$$
P_{ijk}^0(t)=M_{ijk}(t)-K_{ijk}(t)H^TM_{ijk}(t), \tag{18}
$$

$$
M_{ijk}(t)=F\left(\theta_{ijk}(t)\right)P_{ijk}^0(t-1)F^T\left(\theta_{ijk}(t)\right)+GQ\left(\theta_{ijk}(t)\right)G^T, \tag{19}
$$

$$
K_{ijk}(t)=M_{ijk}(t)H[H^TM_{ijk}(t)H]^{-1}. \tag{20}
$$

모드 갱신 단계(mode-update step)시, 확률은 필터의 이노베이션(innovations)으로부터 계산되어진다.

$$
\begin{aligned}
p\left\{\theta_{ijk}(t)\mid z(t)\right\}&=\frac{p\left\{z(t)\mid\theta_{ijk}(t),z(t-1)\right\}p\left\{\theta_{ijk}(t)\mid z(t)\right\}}{p\left\{z(t)\mid z(t-1)\right\}}\\
&=\frac{p\left\{z(t)\mid\theta_{ijk}(t),z(t-1)\right\}}{p\left\{z(t)\mid z(t-1)\right\}}\\
&\quad\cdot\sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N}a_{\alpha j}c_{ij}\bar{a}_{\gamma k}P\left\{\theta_{\alpha\beta\gamma}(t-1)\mid z(t-1)\right\}.
\end{aligned} \tag{21}
$$

식 (21)에서, $p\left\{z(t)\mid\theta_{ijk}(t-1),z(t-1)\right\}$은 Kalman 필터의 이노베이션으로부터 얻은 유사도(likelihood) 이다.

$$
P\{z(t)\mid\theta_{ijk}(t),z(t-1)\}
=\frac{1}{\sqrt{2\pi\sum_{ijk}}}\exp\left\{-\frac{1}{2}(\tilde{z}(t))^T\sum_{ijk}^{-1}(z(t))\right\}=N_{ijk} \tag{22}
$$

여기서 $\tilde{z}(t)=z(t)-H^TF\left(\theta_{ijk}(t)\right)x_{ijk}^0(t-1)$은 $\left(\theta_{ijk}(t)\right)$-차 Kalman 필터의 이노베이션 열이다.

따라서, 가중인자(weighting factor)는 식 (23)에서와 같이 이전 가중인자를 이용하여 회귀적으로 계산할 수 있다.

$$
p\left\{\theta_{ijk}(t)\mid z(t)\right\}=D_t\cdot N_{ijk}\cdot c_{ij}\sum_{\alpha=1}^{M}\sum_{\beta=1}^{L}\sum_{\gamma=1}^{N}a_{\alpha j}\bar{a}_{\gamma k}p\left\{\theta_{\alpha\beta\gamma}(t-1)\mid z(t-1)\right\} \tag{23}
$$

여기서 D_t는 시각 t에서 결정된 스케일 인자(scale factor)이다.

최종적으로 식 (23)의 계산에 필요한 항은 시각 0에서의 음성 및 잡음 모델의 각 상태에 있을 초기 확률이다. 참고로 제안된 회귀 방법의 실험결과 혼성상태의 초기 확률의 선택에 대하여 상대적으로 민감하지 않다.

결합 단계(combination step)에서 최종 추정기 $\hat{x}(t)$를 다음과 같이 얻을 수 있다.

$$
\hat{x}(t)=\sum_{i=1}^{M}\sum_{j=1}^{L}\sum_{k=1}^{N}\hat{x}_{ijk}(t)p\left(\theta_{ijk}(t)\mid z(t)\right). \tag{24}
$$

해당 공분산은 식 (25)로 주어지나, 음성향상에는 사용되지 않으므로, 실제의 계산에서 생략한다.

$$
P(t)=\sum_{i=1}^{M}\sum_{j=1}^{L}\sum_{k=1}^{N}p\left(\theta_{ijk}(t)\mid z(t)\right)\left[P_{ijk}(t)+\left(\hat{x}_{ijk}(t)-\hat{x}(t)\right)\left(\hat{x}_{ijk}(t)-\hat{x}(t)\right)^T\right] \tag{25}
$$

실제 계산에 있어서 문제는 L, M, K가 클 경우, 식 (23)의 계산에 많은 계산량과 메모리가 필요하여, 개선된 알고리즘의 개발이 필요하다는 점이다.

식 (23)을 살펴보면 N_{ijk}와 $p\left\{\theta_{\alpha\beta\gamma}(t-1)\mid z(t-1)\right\}$이 극단적으로 작은 값을 갖기 때문에, 결과적으로 가중인자 $\left\{p\left(\theta_{ijk}(t)\mid z(t)\right)\right\}$ 값의 계산에서 대부분이 무시될 수 있다는 사실을 이용할 수 있다. 즉 식 (23)을 계산하는데 있어서 경험적으로 정한 문턱값 이상의 N_{ijk} 값과 $p\left\{\theta_{\alpha\beta\gamma}(t-1)\mid z(t-1)\right\}$ 값만을 이용하여 계산을 한다. 그럴경우 계산량과 메모리 요구량이 현격히 감소하게 되며, 식 (23)은 다음과 같이 쓸 수 있다.

$$
p\left\{\theta_{i^\#j^\#k^\#}(t)\mid z(t)\right\}=D_t\cdot N_{i^\#j^\#k^\#}c_{i^\#j^\#}\sum_{\alpha^*}\sum_{\beta^*}\sum_{\gamma^*}a_{\beta^*j^\#}\bar{a}_{\gamma^*k^\#}\cdot
p\left\{\theta_{\alpha^*\beta^*\gamma^*}(t-1)\mid z(t-1)\right\} \tag{26}
$$

여기서 #과 *는 다음과 같이 문턱값을 이용하여 선택된 상태를 나타낸다.

$$
i^\#j^\#k^\#\Leftarrow\text{if } N_{ijk}\geq threshold1
$$

$$
\alpha^*\beta^*\gamma^*\Leftarrow\text{if } p\left\{\theta_{\alpha\beta\gamma}(t-1)\mid z(t-1)\right\}\geq threshold2.
$$

여기서 다음과 같이 모든 가중 인자의 합이 1에 해당

된다는 것은 보장이 된다.

$$\sum_{i_1}\sum_{i_2}\sum_{i_3} p\left\{\theta_{i_1,i_2,i_3}(t)\,|\,z(t)\right\}=1$$

따라서 계산량 측면에서 개선된 결합 단계의 최종 추정기 는 다음과 같이 나타낼 수 있다.

$$\hat{x}(t)=\sum_{i_1}\sum_{i_2}\sum_{i_3}\hat{x}_{i_1,i_2,i_3}(t)p\left(\theta_{i_1,i_2,i_3}(t)\,|\,z(t)\right). \tag{27}$$

이 경우 향상과정 처리에 필요한 계산량 부담이 음성 및 잡음 모델의 상태수에 무관하게 된다. 또 음성신호를 좀더 잘 나타내기 위해서 $(t+p-1)$ 만큼 $\hat{y}(t)$의 계산을 지연시킬 수 있다. 이럴 경우 최종 향상된 음성신호는 식 (28)이 된다.

$$\hat{y}(t)=[\underbrace{0\ldots0}_{p-1}1\underbrace{0\ldots0}_{q}]\hat{x}(t+p-1). \tag{28}$$

IV. 실험 결과

제안된 알고리즘의 성능평가를 위해서 실제의 시변성 자동차 잡음에 의해 오염된 음성 신호에 대한 컴퓨터 시뮬레이션을 수행하였다. 잡음은 시속 10, 50, 80 Km/h로 자동차 주행시 획득한 잡음을 사용하였으며, 고려된 입력 신호대잡음비(signal-to-noise ratio: SNR)는 5,10,15, 20dB이었다. 여기서 신호대잡음비는 음성신호와 잡음의 평균전력비를 사용했다.

잡음 신호의 모델링에는 하나의 mixture를 갖는 단일 HFM을 사용했으며, 상태수는 4로 설정했고 AR 모델의 차수는 15로 설정하여 시변성 및 유색잡음 모델링이 가능하도록 하였다. 사용한 음성신호는 3명의 남성화자로부터 얻은 9개의 국어 문장을 이용하였으며, 표본화 주파수는 12 kHz 였다. 음성의 mixture HFM 모델을 위하여 AR 모델의 차수는 12, mixture수와 상태수는 각각 5로 설정하였다. 모델 파라미터의 추정에는 Baum-Welch의 재추정 알고리즘이 계산시간과 메모리 요구량이 많아서 segmental k-means 알고리즘을 이용하였다 [1]. 향상실험은, 학습과 테스트의 조건이 동일하다고 가정하에, 학습에 사용된 음성 중 3개의 문장에 대하여 수행하였다.

그림 1은 잡음이 섞인 음성에 대한 이전의 IMM 방법 [4-5]과 본 논문에서 제안된 mixture IMM방법에 대한 segmental_SNR 결과를 보여준다. 입력 SNR은 10dB이었으며 프레임 사이즈는 200 이었다. 그림 1을 보면, 거의 모든 구간에서 제안된 mixture IMM방법이 기존의 IMM방법에 비하여 향상된 성능을 보여주며, 그 차이는 상태 전이 영역에서 두드러짐을 관찰할 수 있다.

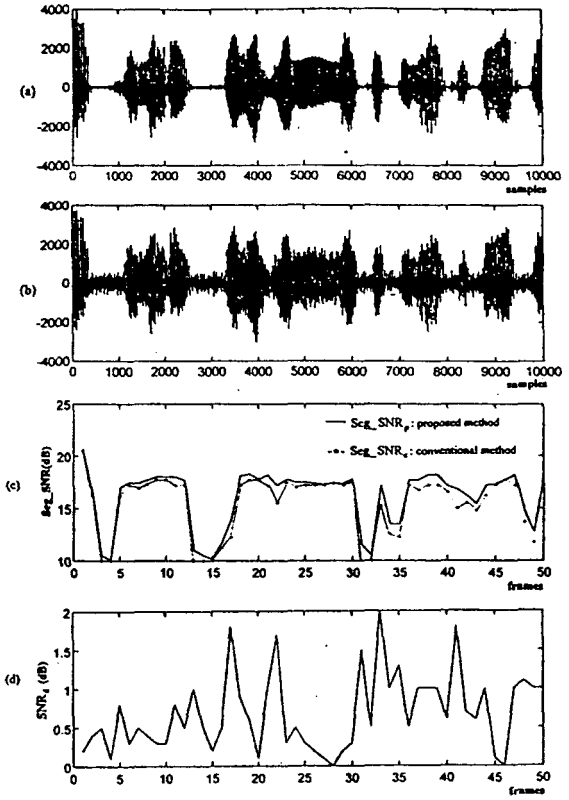

그림 1. (a) 잡음이 없는 음성신호
(b) SNR 10dB의 잡음에 오염된 음성신호
(c) 제안된 방법과 기존 방법으로부터 얻은 Segmental_SNR의 곡선
(d) 차분 Segmental_SNR의 곡선

Fig. 1. (a) clean speech singal
(b) noisy speech signal with input SNR 10 dB
(c) Segmental_SNR contour of proposed and conventional method
(d) difference Segmental_SNR contour (SNRd=Seg_SNRp-Seg_SNRc).

표 1. 자동차 잡음에 대한 여러 입력 SNR에 따른 출력 SNR 비교

Table 1. Comparisons of output SNR at various input SNRs for car noise.

INPUT SNR (dB)	5	10	15	20
proposed MIMM (dB)	11.27	14.95	18.53	22.25
IMM (dB)	10.45	14.24	17.92	21.87

표 1은 여러 입력 SNR에 따른 IMM과 제안된 mixture IMM 두 방법의 성능차이를 보여준다. 출력 SNR을 살펴보면 제안된 방법이 기존의 IMM 방법에 대하여 0.4 - 0.7 dB 향상되었음을 알 수 있다. 그리고 수치상의 작은 차이에 대한 사람들의 주관적인 반응을 살펴보기 위하여 비공식적인 청취실험을 실시하였다. 입력 SNR 10 dB에 대한 두 방법의 음성향상 결과인 각각의 3개 문장을 2개씩 임의의 방법 순서로 15명에게 들려주어 선호도를 조사한 결과 전체 45개중 35개가 선택됨으로써 기

존의 IMM방법에 비하여 본 논문에서 제안한 mixture IMM방법에 대한 강한 선호가 있음을 확인하였다. 제안된 mixture IMM 알고리즘에서의 음질 향상의 주된 요인은 상태 전이 영역에서의 잡음 억제 능력에 기인하며, 이는 mixture를 이용함으로써 음성의 시변특성 모델링이 향상됨에 따라 상대적으로 잡음억제능력이 강화되었기 때문이다.

V. 결 론

본 논문에서는 시변가산유색잡음에 오염된 음성신호의 향상을 위하여 mixture IMM에 근거한 효율적인 회귀추정알고리즘을 제안했다. 제안된 방법은 계수가 Markovian적으로 스위칭되는 선형시스템에 근거한 다중 Kalman 필터 모델을 이용했으며, 그 결과 MMSE 추정기는 음성 및 잡음에 대한 모델의 복합상태에 대한 조건평균추정기의 가중치합으로 구성된다. 이때, 가중치는 주어진 잡음이 섞인 신호에 대한 복합상태의 사후확률이 된다. 시뮬레이션 실험 결과 제안된 방법이 이전에 제안된 IMM 방법 [4-5]에 비하여 성능이 향상되었음을 확인하였다.

참 고 문 헌

1. H. Sheikhzadeh and L. Deng, "Waveform-based speech recognition using hidden filter models:parameter selection and sensitivity to power normalization," *IEEE Trans. Speech and Audio Processing*, vol.2, pp.80-89, Jan. 1994.
2. H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Automatic Control*, vol.33, pp.780-783, August 1988.
3. K. Y. Lee and K. Shirai, "Recursive estimation for speech enhancement in colored noise," *IEEE Signal Processing Letters*, vol.3, pp.196-199, July 1996.
4. 김재범, 이기용, 이충웅, "HFM에 기초한 음성신호의 향상을 위한 효율적인 순환 추정," 한국음향학회지, vol. 16, no. 7, pp. 75-79, 1997.
5. J.B. Kim, K.Y. Lee, and C.W. Lee, "On the application of the interacting multiple model algorithm for enhancing noisy speech," *IEEE Trans. Speech and Audio Processing*, accepted for publication.

▲이 기 용(Lee, Ki Yong)

1983년 2월 : 숭실대 전자공학과 졸업
1985년 2월 : 서울대대학원 전자공학과 졸업(공학석사)
1991년 2월 : 서울대대학원 전자공학과 졸업(공학박사)
1994년8월 ~ 1995년 8월 : 일본 와세다대학/영국 에딘버러 대학 Post-Doc.
1996년 : 일본 와세다대학 방문연구원(JSPS 초청)
1997년 : 독일 뮌헨공대 방문연구원 (DAAD 초청)
1991년9월 ~ 1997년 8월 : 창원대학교 전자공학과 조교수
1997년9월 ~ 현재 : 숭실대 정보통신 전자공학부 부교수

▲임 재 열(JaeYeol Rheem)

1986년 2월 : 서울대학교 전자공 학과 (공학사)
1988년 2월 : 서울대학교 전자공학과 (공학석사)
1995년 2월 : 서울대학교 전자공학과 (공학박사)
1995년 9월: ~ 현재 : 한국기술교육대학교 전자공학과(조교수)

※주관심분야 : 음성신호처리, DSP, 통신신호처리