한국음향학회지 제23권 제7호 pp. 540~547 (2004)

HMM과 H∞필터를 이용한 강인한 음성 향상

Robust Speech Enhancement Using HMM and H∞ Filter

김 준 일*, 이 기 용**
(Jun-Il Kim, Ki-Yong Kim**)

*숭실대학교 정보통신과, **숭실대학교 정보통신전자공학부

(접수일자: 2004년 7월 9일; 수정일자: 2004년 9월 13일; 채택일자: 2004년 10월 15일)

칼만/위너 필터에 근거한 음성향상 알고리즘은 잡음의 선험적 지식을 요구하고, 음성신호와 추정신호의 오차분산을 최소화하는데 중점을 두고 있어, 잡음에 대한 통계적 추정에 오류가 있을 경우 결과에 악영향을 미칠 수 있다. 그러나 H∞ 필터는 잡음에 대한 어떠한 가정이나 선험적 지식을 요구하지 않으며, 최소상계 (Least Upper Bound)를 적용하여 추정된 모든 신호들로부터 최소에러 신호를 갖는 최상의 추정신호를 찾아내므로 칼만/위너 필터보다 잡음의 변화에 강인하다. 본 논문에서는 학습 신호로부터 은닉 마코프 모델의 파라미터를 추정한 후, 오염된 신호를 고정된 개수의 H∞필터를 통과시켜 각 출력에 가중된 합으로 향상된 음성 신호를 구하는 다중 H∞필터에 의한 강인한 음성향상 방법을 제안한다. 제안된 방법의 성능 평가를 위하여 음성 향상 시간과 신호 대 잡음비를 비교한 결과, 기존의 방법에 비해 계산량은 다소 증가하지만 신호 대 잡음비는 약 1~2dB 향상 되었다.

핵심용어: 음성향상, 칼만 필터, H∞ 필터, 은닉 마코프 모델, EM 알고리즘
투고분야: 음성처리 분야 (2.3)

Since speech enhancement algorithms based on Kalman/Wiener filter require a priori knowledge of the noise and have focused on the minimization of the variance of the estimation error between clean and estimated speech signal, small estimation error on the noise statistics may lead to large estimation error. However, H ∞ filter does not require any assumptions and a priori knowledge of the noise statistics, but searches the best estimated signal among the entire estimated signal by applying least upper bound, consequently, it is more robust to the variation of noise statistics than Kalman/Wiener filter. In this paper, we propose a speech enhancement method using HMM and multi H∞ filters. First, HMM parameters are estimated with the training data. Secondly, speech is filtered with multiple number of H∞ filters. Finally, the estimation of clean speech is obtained from the sum of the weighted filtered outputs. Experimental results shows about 1dB~2dB SNR improvement with a slight increment of computation compared with the Kalman filter method.

Keywords: Speech Enhancement, Kalman filter, H∞ filter, HMM, EM algorithm
ASK subject classification: Speech signal processing (2.3)

I. 서 론

음성 향상 (Speech Enhancement)이란, 소리를 주고받는 통신 시스템의 입출력 신호가 잡음환경에서 잡음에 의해 영향을 받았을 때, 잡음이 부가된 음성 신호로부터 잡음 신호를 최소화 하여 음성 통신 시스템의 성능을 개선하는 방법이다.

지금까지 제안된 음성 향상 방법들은 크게 스펙트럼 차감법[1], 자기회귀 (AR: Autoregressive) 모델링[2,3]과 평균 최대 (EM: Expectation Maximization) 알고리즘에 기초한 방법[4-6]과 은닉 마코프 모델 (HMM: Hidden Markov Model)에 기초한 방법[7, 8]등으로 분류할 수 있다.

스펙트럼 차감법은 배경 잡음에 의해 손상된 음성에서

책임저자: 김 준 일 (junilkim@ctsp.ssu.ac.kr)
156-743 서울시 동작구 상도5동
숭실대학교 정보통신전자공학부
(전화: 02-817-4591; 팩스: 02-817-4591)

주파수 상에서 스펙트럼의 크기만을 제거하여 음성을 향상시키는 방법이다. 스펙트럼 차감법은 계산량이 적어 가장 널리 사용되는 방법 중 하나이지만, 음악 잡음 (Musical Noise)이 발생하는 단점이 있다. AR (Autoregressive) 모델링과 EM (Expectation Maximization) 알고리즘에 기초한 방법은 음질이나 특성면에서 스펙트럼 차감법보다는 성능이 좋다. 그러나, 계산량이 많고 깨끗한 음성에 대한 연결 확률 분포 (joint probability density)에 대한 사전지식이 필요 하므로, 실제 환경에 이러한 방법을 적용하는 것은 쉽지 않다. 이를 보완한 방법이 Ephraim에 의해서 제안된 은닉 마코프 모델 (HMM)을 이용한 음성 향상 방법이다[5,6]. HMM을 이용한 음성 향상 방법은, 먼저 깨끗한 학습 신호로부터 HMM의 파라미터를 추정한 후, 오염된 신호를 고정된 개수의 칼만 필터 (Kalman Filter)나 위너 필터 (Wiener Filter)를 통과시켜 각 출력에 가중치를 적용하여 추정 신호를 구하는 방법이다. 그러나, 칼만과 위너 필터는 음성신호와 잡음신호의 연결 통계 특성 (joint statistics)의 정확한 값에 대한 사전 정보가 필요하다. 그러므로, 최소 평균 제곱 에러 (MMSE: Minimum Mean Square Error)를 이용하여 구해진 AR변수를 이용하는 칼만 필터 (Kalman Filter)는 잡음의 통계적 추정에 오류가 있을 경우 결과에 악영향을 미칠 수 있다. 이러한 칼만 필터의 단점을 보완하기 위하여 H_∞ 필터 알고리즘이 제안되었다[9, 10]. H_∞ 필터는 음성신호와 잡음신호의 연결 통계 특성 (joint statistics)과 잡음에 대한 어떠한 가정이나 선험적 지식을 고려하지 않고 음성 신호를 향상 시킬 수 있다. 이러한 H_∞ 필터는 최소상계 (Least Upper Bound)를 적용하여 추정된 모든 신호들로부터 최소 에러신호를 갖는 최상의 추정신호를 찾아내므로 칼만 필터보다 잡음의 변화에 더욱 강인하다.

본 논문에서는 음성의 통계적 특성을 이용하여 모델 파라미터를 추정하는 HMM과 잡음의 변화에 강인한 H_∞ 필터 알고리즘을 사용한 음성향상 방법을 제안하였다. 제안된 음성향상 방법은 다중 H_∞ 필터와 시변 사후확률의 가중화된 합으로 이루어진다.

제안된 방법과 기존의 방법의 성능 비교는 전체 신호대 잡음비 (Global Speech to Noise Ratio), 부분 신호대 잡음비 (Segmental Speech to Noise Ratio)와 사운드 스펙트로그램으로 수행하였다. 제안된 방법이 기존의 방법과의 성능 비교를 위하여 상태별 혼합성분의 개수를 변화시키고, 또한 SNR를 다르게 하여 성능 비교를 하였다. 제안된 방법이 기존의 방법보다 음성 향상 결과를 얻는데 걸리는 소요 시간은 다소 증가하였지만 SNR에서 1dB~2dB 정도 더 향상됨을 알 수 있었다.

본 논문은 II장에서 기본적인 H_∞ 필터 알고리즘에 대하여 정의하였고, III 장에서는 신호열과 가중치 요소를 추정하여 향상된 음성 신호를 구하기 위하여 HMM에 의한 다중 H_∞ 필터를 적용한 음성향상 방법을 제시하였다. IV장에서는 제안된 방법의 성능을 검증하기 위한 실험 및 결과를 제공하고, 마지막으로 V장에서는 결론을 서술하였다.

II. H_∞ 필터 알고리즘

음성 향상을 위하여 H_∞ 필터 알고리즘은 다음과 같이 정리된다 [7,8,10]. 잡음신호 $s(k)$는 식(1)과 같이 표현할 수 있다.

$$s(k) = y(k) + v(k) \tag{1}$$

여기서, $v(k)$는 배경 잡음이고, 깨끗한 음성 $y(k)$는 식(2)로 표현할 수 있다.

$$y(k) = \sum_{j=1}^{p} a(j)y(k-j) + w(k) \tag{2}$$

식(2)에서 $a(j)$는 AR계수이고, p는 AR계수의 차수이다. 식(1)과 (2)를 상태방정식으로 나타내면 다음과 같이 나타낼 수 있다.

$$Y(k) = AY(k-1) + Bw(k) \text{ (State equation)} \tag{3}$$

$$s(k) = CY(k) + v(k) \text{ (Measurement equation)} \tag{4}$$

여기서, $Y(k) = [y(k)\, y(k-1)\, y(k-2)\, \Lambda\ y(k-p+1)]^T$는 과거 p개의 관측 열이고, A는 $p \times p$ 행렬, B^T와 C는 $1 \times p$ 벡터로 표현할 수 있다.

$$A=\begin{bmatrix} a(1) & a(2) & \Lambda & a(p-1) & a(p) \\ 1 & 0 & \Lambda & 0 & 0 \\ 0 & 1 & \Lambda & 0 & 0 \\ M & M & \Lambda & M & M \\ 0 & 0 & \Lambda & 1 & 0 \end{bmatrix}_{p\times p}$$

$$B^T = C = \begin{bmatrix} 1 & 0 & \Lambda & 0 & 0 \end{bmatrix}_{1\times p}$$

H_∞ 필터에서는 $w(k)$ 와 $v(k)$ 에 대한 어떠한 가정도 하지 않는다. 추정 신호는 임의의 $v(k), w(k) \in l_2, Y(0) \in \Re^p$ 에 대해서 선형 결합된 추정 신호열 $Y(k)$ 에서 최소 추정 에러 값을 갖는 $o(k)$ 로 정의된다[10].

$$o(k) = UY(k) \tag{5}$$

여기에서, $U = [0\ 0\ 0\ \Lambda\ 0\ 1]^T_{\infty}$, $U \in \Re^{1\times p}$ 이다.

H_∞ 필터에서 성능 평가 기준은 아래와 같이 표현할 수 있다.

$$J = \frac{\sum_{k=0}^{N-1} |o(k) - \hat{o}(k)|_Q^2}{\left|Y(0) - \hat{Y}(0)\right|_{P_0^{-1}}^2 + \sum_{k=0}^{N-1} \left\{ |w(k)|_{W^{-1}}^2 + |v(k)|_{V^{-1}}^2 \right\}} \tag{6}$$

여기서, $Q \geq 0,\ p_0^{-1} > 0, W > 0, V > 0$ 은 설계자가 임의로 주는 가중치 행렬이다. 여기에서, $|o(k)|_Q^2$ 는 $o(k)$ 의 Q 행렬에 의한 L_2 norm 값, 즉, $|o(k)|_Q^2 = o(k)^T Q\, o(k)$ 이다. 여기에서, H_∞ 필터는 모든 가능한 추정신호 $\hat{o}(k)$ 중 식(11)을 만족하고, 오차신호를 최소화 하는 최적화된 $\hat{o}(k)$ 를 찾는다.

$$\sup J \leq \gamma^2 \tag{7}$$

여기서, sup는 최소상계 (Least Upper Bound)이고, .3 γ 는 잡음 감쇄 정도를 나타낸다. H∞ 필터는 식(7)처럼 최소상계 (Least Upper Bound)를 적용하여, 균일한 오차신호를 제공한다. 따라서 칼만 필터보다 잡음의 변화에 강인한 특성을 가지게 된다.

잡음감쇄 정도를 $\gamma > 0$ 라 하고, 대칭 행렬 $P(k) > 0$를 만족할 때, $P(k)$ 는 상태 방정식과 성능 기준식(6)을 이용하여 Riccati 형태 방정식[11]으로 식 (8)과 같이 표현할 수 있다.

$$P(k+1) = AP(k)(I - \gamma^{-2}\overline{Q}P(k) + C^T V^{-1} CP(k))^{-1} A^T + BWB^T \tag{8}$$

여기에서, $B^T = C = [1\ 0\ 0\ ,\Lambda\ ,\ 0]_{1\times p}$ 이다.

식(5)에서 $\hat{Y}(k)$ 는 식(8)을 이용하여 추정할 수 있다.

$$\hat{Y}(k) = A\hat{Y}(k-1) + H(k)(s(k) - CA\hat{Y}(k-1)),\ \ \hat{Y}(0) = 0 \tag{9}$$

$$H(k) = AP(k)(I - \gamma^{-2}\overline{Q}P(k) + C^T V^{-1} CP(k))^{-1} C^T V^{-1} \tag{10}$$

여기서, Equation.3 $H(k)$ 는 H∞ 필터의 이득이고, $\overline{Q} = U^T Q U$ 이다.

III. 다중 H_∞ 필터에 의한 음성 향상

본 장에서는 잡음에 대한 선험적 지식을 필요로 하지 않는 H∞ 필터를 다중으로 적용하여 음성을 향상시키는 방법에 대하여 설명하고자 한다.

깨끗한 음성 신호 $y(k)$를 각각의 L개의 상태 (state)와 M 혼합 (mixture) 성분을 갖는 ARHMM 모델로 표현하기 위해서, y 에 대응되는 상태열을 가우시안 AR모델을 사용한다. T 개의 프레임을 갖는 관측열 $y = \{y(t), t = 1, \Lambda, T\}$, $y(t) = \{y((t-1)N+1), \Lambda, y(tN)\}$에서, y 에 대응하는 상태열을 $s = \{s_t, t = 1,2,.....,T\}$, $s_t \in \{1,2...,L\}$라 두고, (s, y) 에 대응 하는 혼합열을 $h = \{h_t, t = 1,2,.....,T\}$, $h_t \in \{1,2...,M\}$라고 하면, 깨끗한 음성 모델은 아래와 같이 표현할 수 있다.

$$y(k) = B_{k,|s_t}^T Y(k-1) + e_{k,|s_t}(k)\ ,\ \ (t-1)N+1 < k < tN \tag{11}$$

여기에서, $B_{k,|s_t}^T = [b_{k,|s_t}(1), ..., b_{k,|s_t}(p)]$ 는 각 상태에서 혼합별 AR 계수들이고, $Y(k-1) = [y(k-1)..y(k-p)]^T$ 는 과거 p개의 관측열, $e_{k,|s_t}(k)$ 는 각 상태에서 혼합성분의

잔차신호이다.

식(11)과 백색 잡음에 오염된 신호는 다음의 상태방정식으로 표현할 수 있다.

$$Y(k)=F_{h_t|s_t}Y(k-1)+Be_{h_t|s_t}(k) \tag{12}$$

$$z(k)=CY(k)+v(k) \tag{13}$$

위 식에서, 각 요소들은 아래와 같다.

$$Y(k)=\begin{bmatrix} y(k) & y(k-1) & \Lambda & y(k-(p-1)) \end{bmatrix}^T$$

$$F_{h_t|s_t}=\begin{bmatrix} b_{h_t|s_t}(1) & b_{h_t|s_t}(2) & \Lambda & b_{h_t|s_t}(p-1) & b_{h_t|s_t}(p) \\ 1 & 0 & \Lambda & 0 & 0 \\ 0 & 1 & \Lambda & 0 & 0 \\ M & M & \Lambda & M & M \\ 0 & 0 & \Lambda & 1 & 0 \end{bmatrix}$$

$$B^T=C=\begin{bmatrix} 1 & 0 & \Lambda & 0 & 0 \end{bmatrix}_{1\times p}$$

잡음 음성이 주어져있을 때, $\hat{Y}(k)$을 추정하는 것은 조건 평균으로 주어진다.

$$\begin{aligned}\hat{Y}(k)&=\{E\{Y(k)\mid z(k)\}\\&=\int_{-\infty}^{\infty}Y(k)\,p(Y(k)\mid z(k))dY(k)\end{aligned} \tag{14}$$

위 (14)식의 조건 분포 함수를 아래 식과 같이 쓸 수 있다.

$$\begin{aligned}&p(Y(k)\mid Z(k))\\&=\sum_{j=1}^{L}\sum_{m=1}^{M}p(Y(k)\mid s_t=j,h_t=m,z(k))p(s_t=j,h_t=m\mid z(k))\end{aligned} \tag{15}$$

식(15)을 (14)식에 대입하고 적분을 합으로 대체하면, $\hat{Y}(k)$는 아래 식으로 추정할 수 있다.

$$\hat{Y}(k)=\sum_{j=1}^{L}\sum_{m=1}^{M}\hat{Y}_{h_t|s_t}(k)p(s_t=j,h_t=m\mid z(k)) \tag{16}$$

위 식에서 $\hat{Y}_{h_t|s_t}(k)$는 $s_t=j$, $h_t=m$ 일 때 $Y(k)$의 조건 평균 추정식이다.

$\hat{Y}(k)$을 구하기 위해서는 $\hat{Y}_{h_t|s_t}(k)$와 가중치 $p(s_t=j,h_t=m\mid z(k))$을 계산하는 두 과정으로 나눌 수 있다.

위의 두 과정에 필요한 파라미터는 학습 데이터를 은닉 마코프 모델을 통하여 추정하였다. 파라미터 집합은 다음과 같이 정의된다.

$$\lambda=\{A,\ B_{h_t|s_t},\ Q_{h_t|s_t},\ c_{h_t|s_t}\} \tag{17}$$

$A=[a_{ij}]$는 상태천이행렬이고, $B_{h_t|s_t}$는 각 상태에서 혼합에 대한 AR계수, $Q_{h_t|s_t}$는 잔차 신호의 분산 값, $c_{h_t|s_t}$는 가중치이다.

3.1. $\hat{Y}_{m|j}(k)$의 추정

$s_t=j$, $h_t=m$ 일 때, $\hat{Y}_{m|j}(k)$는 각 상태에서의 H∞ 필터로 추정할 수 있고, H∞ 필터 알고리즘은 다음 아래 식과 같이 표현된다.

$$\hat{Y}_{m|j}(k)=F_{m|j}(k)\hat{Y}_{m|j}(k-1)+H_{m|j}(k)\cdot\{z(k)-CF_{m|j}(k)\hat{Y}_{m|j}(k-1)\} \tag{18}$$

$$H_{m|j}(k)=F_{m|j}(k)P_{m|j}(k)L_{m|j}(k)C^TV^{-1} \tag{19}$$

$$P_{m|j}(k+1)=F_{m|j}(k)P_{m|j}(k)L_{m|j}(k)F_{m|j}(k)^T+BQ_{m|j}B^T \tag{20}$$

여기에서, $L_{m|j}(k)$는

$$L_{m|j}(k)=\left(I-\gamma^{-2}Q_{m|j}P_{m|j}(k)+C^TV^{-1}CP_{m|j}(k)\right)^{-1} \tag{21}$$

이고, $H_{m|j}$는 H∞ 필터의 이득, $P_{m|j}$는 에러 공분산이다. 그리고, $Q_{m|j}=\sigma_{m|j}^2$는 $e_{h_t|s_t}(k)$의 공분산 행렬이다. $\hat{Y}_{m|j}(k)$는 식(18)-(21)로부터 회귀적으로 구해진다.

3.2. $p(s_t=j,h_t=m\mid z(k))$의 계산

가중치 요소 $p(s_t=j,h_t=m\mid z(k))$는 베이시안(Bayesian) 법칙을 적용하여 아래 식과 같이 나타낼 수 있다.

$$p(s_t = j, h_t = m \mid z(k)) = \frac{p(z(k) \mid s_t = j, h_t = m, z(k-1))\, p(s_t = j, h_t = m \mid z(k-1))}{p(z(k) \mid z(k-1))} \quad (22)$$

식(22)에서 분자의 $p(z(k) \mid s_t = j, h_t = m, z(k-1))$는 다음과 같이 쓸 수 있다.

$$p(z(k) \mid s_t = j, h_t = m, z(k-1)) = \prod_{k=1}^{N} N\left[\hat{Y}_{m|j}(k), C^T P_{m|j}(k) C\right] \quad (23)$$

여기서, $N[\cdot,\cdot]$는 정규분포이다.

식(22)에서 두 번째 요소 $p(s_t = j, h_t = m \mid z(k-1))$는 마코프 과정으로 나타낼 수 있으므로 식(24)로 나타낼 수 있다.

$$p(s_t = j, h_t = m \mid z(k-1)) = \sum_{i=1}^{L}\sum_{l=1}^{M} \{p(s_t = j, h_t = m \mid s_{t-1} = i, h_{t-1} = l, z(k-1))\, p(s_{t-1} = i, h_{t-1} = l, z(k-1))\} \quad (24)$$

위 식에서 첫 번째 요소는 아래 식으로 다시 쓸 수 있다.

$$p(s_t = j, h_t = m \mid s_{t-1} = i, h_{t-1} = l, z(k-1)) = p(h_t = m \mid s_t = j, s_{t-1} = i, h_{t-1} = l, z(k-1)) \times p(s_t = j \mid s_{t-1} = i, h_{t-1} = l, z(k-1)) \quad (25)$$

h_t 와 s_t 는 서로 독립이므로 식(25)에서 두 요소는 다음과 같이 다시 표현할 수 있다.

$$p(h_t = m \mid s_t = j, s_{t-1} = i, h_{t-1} = l, z(k-1)) = c_{m|j} \quad (26)$$

$$p(s_t = j \mid s_{t-1} = i, h_{t-1} = l, z(k-1)) = p(s_t = j \mid s_{t-1} = i) = a_{ij} \quad (27)$$

식(26), (27)를 식(24)에 대입하면 아래 식과 같이 나타내어진다.

$$p(s_t = j, h_t = m \mid z(k-1)) = \sum_{i=1}^{L}\sum_{l=1}^{M} a_{ij}\, c_{m|j}\, p(s_{t-1} = i, h_{t-1} = l, z(k-1)) \quad (28)$$

식(22)에서 분모는 상태에 독립적이므로, 이 요소는 스케일 인수가 된다.

$p(s_t = j, h_t = m \mid z(k))$는 이전 가중치 요소 (previous weighting factor)를 사용해서 순환적으로 계산할 수 있다.

$$p(s_t = j, h_t = m \mid z(k)) = D_t N_{m|j} \sum_{i=1}^{L}\sum_{l=1}^{M} a_{ij}\, c_{m|j}\, p(s_{t-1} = i, h_{t-1} = l \mid z(k-1)) \quad (29)$$

식(29)에서 D_t 는 가중치 요소들의 모든 합이 1이 되게 하는 스케일 인수이고, $N_{m|j}$ 는 $s_t = j$, $h_t = m$ 일 때의 정규분포이다.

$$\sum_{j=1}^{L}\sum_{m=1}^{M} p(s_t = j, h_t = m \mid z(k)) = 1 \quad (30)$$

$\hat{Y}_{m|j}(k)$와 $p(s_t = j, h_t = m \mid z(k))$를 계산한 후, $\hat{Y}(k)$를 구할 수 있다..

$$\hat{Y}(k) = \sum_{j=1}^{L}\sum_{m=1}^{M} \hat{Y}_{m|j}(k)\, p(s_t = j, h_t = m \mid z(k)) \quad (31)$$

(31)식에서 추정된 상태열 $\hat{Y}(k)$에서 음성 향상된 추정 신호는 아래와 같이 구할 수 있다.

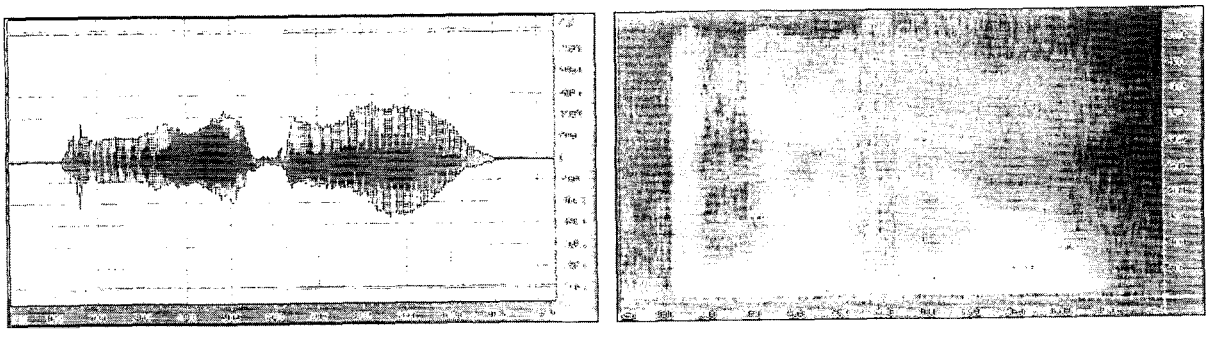
$$\hat{o}(k) = U\hat{Y}(k) \quad (32)$$

여기에서, $U = [0\ 0\ 0 \Lambda\ 0\ 1]^T_{pxl}$, $U \in \Re^{lxp}$ 이다.

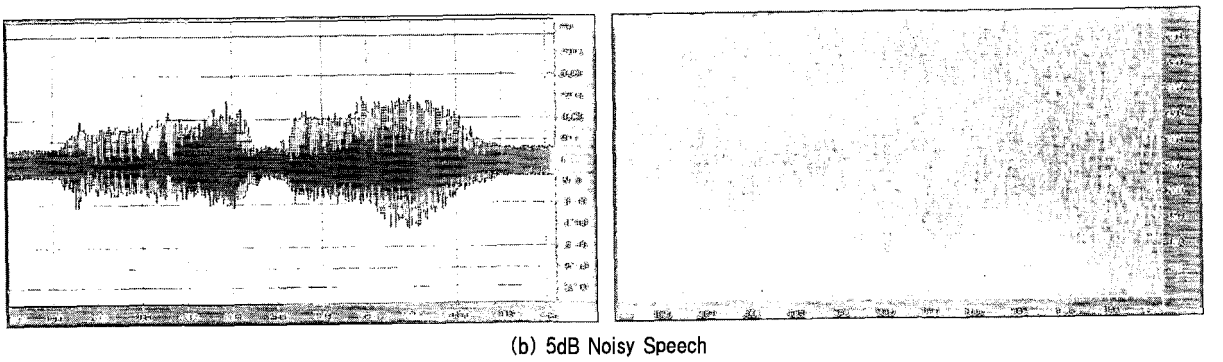
IV. 실험 및 결과

제안된 방법의 성능을 확인하기 위해서 입력신호의 신호 대 잡음비가 각각 0dB, 5dB, 10dB에 해당하는 백색 잡음이 부가되었을 때 제안된 방법과 기존의 방법의 성능을 비교하였다.

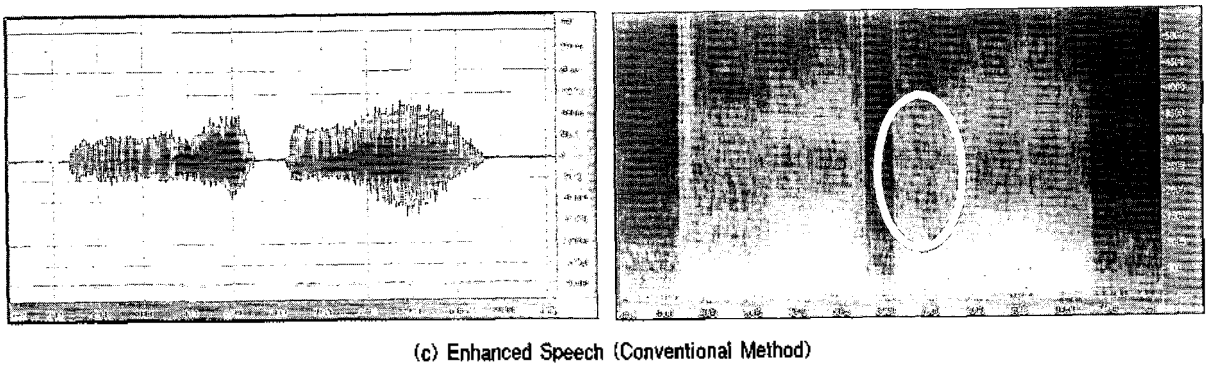
실험에 사용한 문장은 "안녕하세요" 로 학습은 발성된 6개의 문장을 사용 하였고, 테스트는 동일 문장으로 학습에 사용하지 않은 문장을 사용하였다. 샘플링 주파수



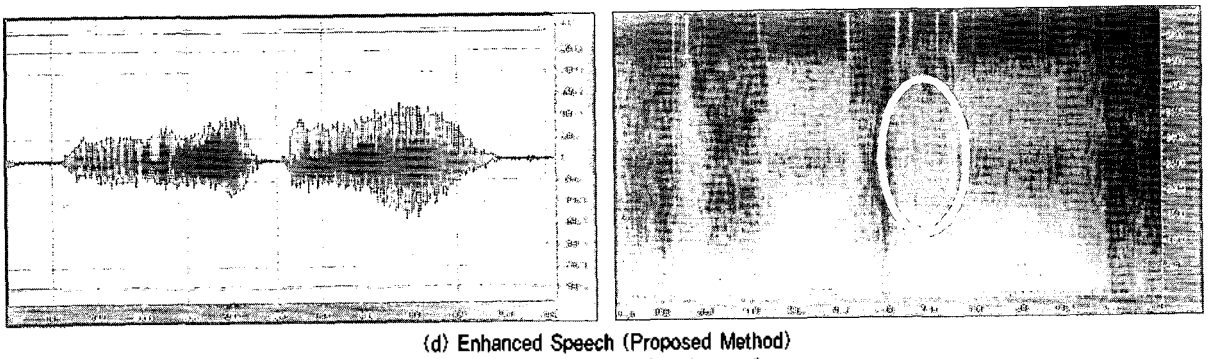
(a) Clean Speech



(b) 5dB Noisy Speech



(c) Enhanced Speech (Conventional Method)



(d) Enhanced Speech (Proposed Method)

그림 1. 음성 향상 파형과 스펙트로그램

Fig 1. Speech enhancement results.

는 11,025Hz, 깨끗한 음성의 AR 모델은 15차, HMM은 7개 상태 및 2개 혼합성분을 사용하였다. 전체실험은 Pentium IV 2.4GHz PC에서 이루어졌다.

실험은 기존의 방법과 제안된 방법에 대해서, 입력신호 대 잡음비와 잡음이 감소된 출력 신호 대 잡음비의 결과를 비교하고 음성 향상을 위한 소요 시간을 비교하였다. 여기서 기존의 방법은 $\hat{Y}_{mlj}(k)$를 추정하는데 H_∞ 필터대신에 칼만 필터를 사용한 방법이다.

표 1은 입력 신호 대 잡음비가 10 dB환경에서 잡음에 대한 사전 정보를 모르는 경우에 음성 향상의 결과를

표 1. State 수와 Mixture 수에 따른 음성향상 결과 비교
Table 1. Performance comparison between conventional and proposed method when states and mixtures are changed

State	Mixture	Conventional Method	Proposed Method
5	2	11.44	12.90
	3	11.49	13.15
	4	11.77	12.77
6	2	11.88	13.16
	3	11.44	13.33
	4	11.41	13.24
7	2	10.80	13.59
	3	10.34	13.46
	4	11.17	13.35
8	2	10.24	13.31
	3	11.91	13.09
	4	10.59	13.30

State수와 Mixture수 별로 나타낸 것이다. 각각의 방법에 대해서 제안된 방법이 기존의 방법보다 성능이 더 좋음을 알 수 있는데, 제안된 방법에서는 State 7, Mixture 2를 사용했을 경우에 가장 성능이 좋음을 알 수 있다.

표 2는 입력 신호 대 잡음비가 0, 5, 10 dB인 경우에 음성 향상된 후 출력 신호 대 잡음비 및 수행시간 결과를 나타낸 것이다. 잡음에 대한 정보를 알고 있을 경우와 모를 경우에 대해서 비교 분석하였다. 잡음에 대한 정보를 알고 있는 경우는 잡음 신호를 생성할 때 신호 대 잡음비와 깨끗한 음성 신호의 분산 값을 사전에 미리 알고 있으므로, 다시 말해서 부가된 임의의 잡음의 분산 값을 알 수 있으므로, 이 값을 사용하여 음성을 향상시켰다. 그러나, 잡음에 대한 정보를 모르는 경우에는 잡음 환경에 상관없이 고정된 임의의 잡음 분산 값을 적용하여 음성 향상 결과를 확인하였다.

기존의 방법은 잡음에 대한 정보 유/무에 따라 성능 변화가 크지만 제안된 방법은 잡음에 대한 정보에 상관없이 상대적으로 성능이 일정하게 유지함을 알 수 있다. 잡음에 대한 정보를 모를 경우에는 기존의 방법에서는 임의의 잡음을 가정하고, 제안된 방법에서는 잡음에 대한 성분을 직접 계산하므로 계산량은 다소 증가하나, SNR은 약 1~2 dB정도 향상된 것을 확인할 수 있다. 따라서, 제안된 방법이 기존의 방법보다 잡음에 훨씬 강인한 것을 알 수 있다.

그림 1은 입력 신호 대 잡음비가 5 dB환경에서 향상된 음성의 스펙트로그램을 보여주고 있다. (a), (b)는 각각 깨끗한 음성과 잡음이 부가된 음성의 파형 및 스펙트로그램이고, (c), (d)는 기존의 방법과 제안된 방법으로 음성을 향상 시킨 결과이다. 기존의 방법과 제안된 방법 모두에 대해서 잡음이 제거되어 음성이 향상됨을 볼 수 있지만, 동그라미로 표시된 부분과 같이 (d)의 스펙트로그램이 상대적으로 (c)에 비해 원래의 음성신호에서 존재하는 낮은 에너지 성분에 대하여 잘 복원이 됨을 알 수 있다. 즉, (c)에서는 낮은 에너지를 갖는 성분들이 거의 제거되어 향상된 음성 신호에 왜곡이 발생하게 된다.

V. 결 론

실제의 잡음환경에서는 배경잡음에 대한 사전지식을 일반적으로 알 수가 없고, 비정상적인 통계적 특성을 가지므로, 배경잡음에 대한 사전지식과 비정상 상태라는 가정을 하는 기존의 방법에 의해서는 좋음 음질 개선을 기대할 수 없다. 이를 극복하기 위하여 잡음에 대한 사전지식과 가정이 필요 없는 H_∞ 알고리즘 및 은닉 마코프 모델에 근거한 음성향상 방법을 제안하였다.

잡음에 대한 사전지식이 없는 경우에 대하여 제안된 방법이 기존 방법보다 계산량은 다소 증가하나 SNR 성능이 좋아짐을 알 수 있다. 또한 잡음에 대한 사전지식을 요구하는 기존의 방법에서는 잡음에 대한 사전지식 유무에 따라 SNR 성능이 차이가 있는 반면, 제안된 방법은 잡음에 대한 사전지식 유무에 관계 없이 SNR 성능의 차이가 거의 없음을 알 수 있었다. 따라서 제안된 방법이 기존의 방법보다 잡음에 대해서 강인한 특성을 가짐을 알 수 있다.

표 2. 기존 방법과 제안된 방법의 음성향상 결과 비교 (상태: 7, 혼합성분: 2)
Table 2. Performance comparison between conventional and proposed method when the number of states is 7 and the number of mixtures is 2.

Input SNR (dB)	Conventional method			Proposed method		
	Output SNR(dB)		시간(초)	Output SNR(dB)		시간(초)
	Known	Unknown		Known	Unknown	
0	7.68	7.05	6.0	7.71	7.52	8.8
5	11.16	8.84		10.60	10.44	
10	14.72	10.80		14.25	13.59	

감사의 글

본 논문은 2004학년도 숭실대학교 교내학술연구비 지원에 의하여 수행되었습니다.

참 고 문 헌

1. S. F. Boll, "Suppression of acoustics noise in speech using spectral subtraction", IEEE Trans. Acoustic. Speech Signal Processing, **27**, pp113 120, 1979.
2. J. S. Lim and A. V. Oppenheim, "All pole modeling of degraded speech," IEEE Trans. Acoust., Speech, Signal Processing, ASSP **26**, 197~210, 1978.
3. Ki Yong Lee, Katsuhiko Shirai, "Efficient Recursive Estimation for Speech Enhancement in Colored Noise," IEEE Signal Processing Letters, 3(7), 196~199, 1996.
4. Y. Ephraim, D. Malah, "Speech enhancement using a minimum mean square error short time amplitude estimator", IEEE Trans. Acoustic. Speech Signal Processing, **32**(6), 1109~1121, 1984.
5. Y. Ephraim, D. Malah, and B. H. Juang, "On the application of hidden Markov models for enhancing noisy speech," IEEE Trans. Acoustic. Speech Process. 37(12), 1846~1856, 1989.
6. Y. Ephraim, " A Bayesian approach for speech enhancement using hidden Markov models," IEEE Trans. Signal Processing, **41**, 725~735, 1992.
7. 이기용, "좌 우향 은닉 마코프 모델에서 상태결정을 이용한 음질향상", 한국음향학회지, 23(1), 47~53, 2004
8. Ki Yong Lee, JaeYeal Rheem, "Smoothing approach using forward backward Kalman filter with Markov Switching Parameters for Speech Enhancement," Signal Processing, **80**, 2579~2588, 2000.
9. C. E. de Souza, U. Shaked, and M. Fu, "Robust H∞ filtering with parametric uncertainty and deterministic signal", in Proc. IEEE CDC'92, 2305~2310, 1992.
10. X. Shen and L. Deng, "A Dynamic System Approach to Speech Enhancement Using the H∞ Filtering Algorithm", IEEE Trans. Speech and Audio Processing, 7(4), 391~399, 1999.
11. C.D. Souza, M. Gevers, and G. Goodwin, "Riccati equations in optimal filtering of nonstabilizable systems having singular state transition matrices,"IEEE Trans. On Automatic control, **31**(9), pp.831 838, 1986.

저자 이력

○ 김 준 일 (Jun-Il Kim)

2002년 2월: 숭실대학교 정보통신과(공학사)
2004년 8월: 숭실대학교 정보통신과(석사)
2004년 9월~현재: ㈜텔텍스 분당연구소 연구원
*주관심분야: 음성신호처리, 음성신호 향상, 모바일

○ 이 기 용 (Ki-Yong Lee)

한국음향학회지 제 15권 제3E 참조
1997년 9월~현재: 숭실대학교 정보통신전자공학부 부교수
※주관심분야: 음성신호 향상, 화자인식, 음성