

좌-우향 은닉 마코프 모델에서 상태결정을 이용한 음질향상

Efficient Speech Enhancement based on Left-right HMM with State Sequence Decision Using LRT

이 기 용*
(Ki-Yong Lee*)

* 숭실대학교 정보통신전자공학부

(접수일자: 2003년 6월 13일; 수정일자: 2003년 12월 9일; 채택일자: 2004년 1월 5일)

본 논문에서는 좌-우향 은닉 마코프 모델 (Left-Right Hidden Markov Model)에서 상태결정을 갖는 음성향상 방법을 제안하였다. 은닉 마코프 모델에 기초를 둔 음질향상 방법은 성능은 우수하나, 모든 상태에 대해서 음질향상 알고리즘을 계산하므로, 계산량이 많고, 메모리가 많이 필요하여 실시간 처리에 부적절하다. 좌-우향 은닉 마코프 모델은 마코프 모델은 좌측에서 우측으로의 전이만 허용하는 모델로 단순화시켜 현재 상태에서 현재 상태나 다음 상태로 전이될 수 있는 특성을 가지고 있다. 본 논문에서는, 좌-우향 은닉 마코프 모델에서 유사도비 테스트 (Log-Likelihood Ratio Test)를 이용하여 현재 음성의 상태를 결정하는 알고리즘을 제안하였다. 현재 음성의 상태를 알고 있다면, 현재 상태에 대해서만 음질향상 알고리즘을 계산하므로, 계산량이 줄어든다. 제안된 방법의 성능 평가를 위하여 음질 향상 시간과 신호 대 잡음비를 비교하였다. 제안된 방법은 기존의 방법에 비해 음질향상의 결과는 약 0.2~0.4 dB 정도 떨어졌지만, 계산량을 많이 줄일 수 있었다.

핵심용어: 음질향상, 칼만 필터, 은닉 마코프 모델, 유사도비, EM 알고리즘

투고분야: 음향 신호처리 분야 (1,2)

We propose a new speech enhancement algorithm based on left-right Hidden Markov Model (HMM) with state decision using Log-likelihood Ratio Test (LRT). Since the conventional HMM-based speech enhancement methods try to improve speech quality for all states, they introduce huge computational loads inappropriate to real-time implementation. In the left-right HMM, only the current and the next state are considered for a possible state transition so to reduce the computational complexity. In this paper, we propose a method to decide the current state by using the LRT on the previous state. Experimental results show that the proposed method improves the speed up to 60% with 0.2~0.4 dB degradation of speech quality compared to the conventional method.

Keywords: Speech enhancement, Kalman filter, HMM, Likelihood ratio test, Expectation maximization

ASK subject classification: Acoustic signal processing (1,2)

I. 서론

음성향상은 잡음에 의해 오염된 음성신호를 복구하여 음성의 인지도를 향상시킨다. 또한, 통신 시스템에서 입·출력 신호가 잡음에 의해 오염되어 있을 때 음성 인식이나 음성 부호화를 위한 전처리 과정으로서 시스템의 성능을 향상시키는 기능을 수행한다.

지금까지 제안된 음성향상을 위한 방법들은 크게 세 가지로 스펙트럼 차감법[1], 자기회귀 (AR: Autoregressive) 모델링[2,3]과 평균 최대 (EM: Expectation Maximization) 알고리즘에 기초한 방법과 은닉 마코프 모델 (HMM: Hidden Markov Model)에 기초한 방법[4-8]등으로 나눌 수 있다.

스펙트럼 차감법은 주변 잡음에 의해 손상된 음성 스펙트럼의 크기만을 제거하여 잡음을 줄이는 방법이다. 계산량이 적어서 가장 널리 사용되고 있지만, 음악잡음 (musical noise)이 발생하는 단점이 있다. AR (Autoregressive)

책임저자: 이기용 (kylee@ssu.ac.kr)
156-743 서울시 동작구 상도5동
숭실대학교 정보통신전자공학부
(전화: 02-820-0908 ; 팩스: 02-817-4591)

모델링과 EM (Expectation Maximization) 알고리즘에 기초한 방법은 음절이나 음성의 특성면에서 스펙트럼 차감법보다는 성능이 뛰어나지만, 깨끗한 음성과 잡음의 특성에 대한 사전 지식이 있어야 하므로, 실제 환경에서는 적용하기가 어렵다. 이를 보완하기 위하여 제안된 방법이 은닉 마코프 모델[9,10]에 기초한 음질 향상법이다. 은닉 마코프 모델을 이용한 방법은 실제 음성 신호의 확률분포와 잡음에 대한 정보를 알 수 없을 때, 음성신호로부터 생성되어 있는 훈련열로 암시적으로 주어진다. 은닉 마코프 모델을 이용한 신호 향상 방법은, 먼저 학습 신호로부터 은닉 마코프 모델의 파라미터를 추정 후, 오염된 신호를 고정된 개수의 칼만 필터 (Kalman Filter) 나 위너 필터 (Wiener Filter)를 통과시켜 각각의 출력의 가중의 합으로 추정치를 구한다. 그러나, 은닉 마코프 모델에 기초를 둔 음질향상 방법은 모든 상태에 대해 칼만 필터와 위너 필터에 대한 계산을 하므로, 계산량이 많아 실시간 처리에 부적절하다. 예를 들어, L개 상태의 M개의 혼합성분을 가지는 은닉 마코프 모델인 경우, 음성을 향상시키기 위해서 $L \times M$ 개의 칼만 필터나 위너 필터를 계산해야 한다. 이러한 계산량은 최근에 나온 펜티엄-Ⅴ, CPU 2.xGHz 에서도 음성인식이나 음성 코딩 분야에서 실시간 처리로 음성향상을 얻을 수 없다. 그러므로 실시간이 가능한 고음질의 음성향상을 위해서는 계산량이 적은 은닉 마코프 모델의 음성향상 방법을 요구하고 있다.

본 논문에서는 이러한 계산 과정을 줄이기 위한 방법으로 좌-우향 은닉 마코프 모델에서 유사도비 (Log-Likelihood Test: LRT)를 사용하여 상태열을 결정하는 알고리즘을 제안했다. 좌-우향 은닉 마코프 모델에서는 현재 음성의 상태가 결정된다면, 현재 상태와 바로 다음 상태에 대해서만 칼만 필터를 계산할 수 있으므로, $2 \times M$ 개의 칼만 필터나 위너 필터를 계산하면 된다. 그러나 잡음이 섞인 음성의 현재 상태를 결정하는 것은 쉬운 일이 아니다. 전 프레임의 개선된 음성의 상태 j 를 알고 있을 때, 현재 프레임의 음성의 상태 j 와 $j+1$ 일 때의 확률의 비를 LRT로 표현하고 LRT 값으로 현재 프레임의 상태를 결정하는 방법을 사용한다.

실시간이 가능한 음성향상을 구현하기 위해서, 제안된 방법을 사용하면 기존의 방법에 비해 SNR은 약 0.2~0.4 dB 정도 떨어지나, 계산량을 (2/L)배로 줄일 수 있었다.

본 논문은 II장에서 은닉 마코프 모델에 기초한 음성향상에 대하여 설명하였고, III장에서는 본 논문에서 제안한 좌-우향 은닉 마코프 모델에 기초한 음성향상과 상태 결정 알고리즘에 대해서 설명하였다. IV장은 실험 및

결과를 기술하고, 마지막으로 V장에서는 결론을 서술했다.

II. HMM에 기초한 음성 향상

은닉 마코프 모델은 음성 발음 모델의 파라미터를 추정하는 효과적인 방법으로, 마코프 연쇄상태와 관계되는 파라미터들을 가지는 비선형 예측 모델이다. 가우시안 부신호원은 고정된 형태를 갖는 성도로부터 발생하는 신호를 표현한다. 성도는 평탄한 스펙트럼을 갖는 여기 신호를 입력으로 하는 AR 필터로 모델링 된다. 모든 가능한 성도의 형태들로 구성된 집합은 은닉 마코프 모델에서의 같은 개수를 갖는 상태들과 대응된다.

깨끗한 음성 신호 $y(n)$ 은 마코프모델에 따라서 변하는 파라미터를 갖는 AR 모델인 HMM에 의해서 모델링 된다. L개의 상태에 해당하는 $s_t \in \{t=1, 2, \dots, L\}$, M개의 혼합 (mixture)에 해당하는 $h_t \in \{t=1, 2, \dots, M\}$ 와 상태 천이 행렬 $a_{s_{t-1}, s(t)}$ 를 갖는 일차의 마코프 모델을 생각한다. 시간 t에서의 상태 s_t 에서 혼합 h_t 를 갖는다면 깨끗한 음성 신호는 다음과 같이 표현된다.

$$y(n) = L \sum_{h_t, s_t} Y(n-1) + e_{h_t, s_t}(n), \quad (t-1)N+1 < n < tN \quad (1)$$

여기서, $L \sum_{h_t, s_t} = [l_{h_t, s_t}(1), \dots, l_{h_t, s_t}(\rho)]$ 는 상태 s_t 에서 혼합 h_t 일 때의 AR 계수 벡터이고, $Y(n-1) = [y(n-1), \dots, y(n-\rho)]^T$ 는 과거 ρ 개의 데이터로 구성된 신호의 벡터 관측열이며, $e_{h_t, s_t}(n)$ 는 평균이 0이고, 분산이 σ_{h_t, s_t}^2 인 가우시안 프로세스이다. N은 한 프레임에서의 샘플수이다.

배경 잡음 $v(n)$ 은 평균이 0이고, 분산이 σ_v^2 인 백색 가우시안 모델이고, $e_{h_t, s_t}(n)$ 과 $v(n)$ 은 상관관계가 없다고 가정하자. 통계적으로 독립인 잡음 $v(n)$ 이 부가된, 잡음에 오염된 음성은 아래의 식으로 표현된다.

$$z(t) = y(t) + v(t), \quad t=1, 2, \dots, T \quad (2)$$

여기서, $y(t) = \{y(n), (t-1)N+1 \leq n \leq tN\}$ 는 깨끗한 음성 벡터이고, $v(t) = \{v(n), (t-1)N+1 \leq n \leq tN\}$ 는 음성에 부가된 배경 잡음 벡터이다.

잡음에 오염된 신호로부터 원 신호를 추정하기 위해서 칼만 필터에 기초한 방법들을 사용한다. 이를 위하여 먼저 오염된 신호를 선형상태방정식으로 나타내어야 그 계

수들은 마코프 모델에 따라 변하게 된다. 백색 잡음에 오염된 신호는 음성에 대한 AR 모델과 잡음을 결합하면 다음과 같은 복합마코프 상태에 따른 상태방정식으로 표현할 수 있게 된다.

$$Y(n) = F(s_t, h_t)Y(n-1) + Ge_{h_t, s_t} \quad (3)$$

$$z(n) = H^T Y(n) + v(n) \quad (4)$$

위 식에서, 각 요소들은 아래와 같다.

$$Y(n) = \begin{pmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-(p-1)) \end{pmatrix}$$

$$F(s_t, h_t) =$$

$$= \begin{pmatrix} I_{h_t, s_t}(1), & I_{h_t, s_t}(2), & \dots & I_{h_t, s_t}(p-1), & I_{h_t, s_t}(p) \\ 1, & 0, & \dots & 0, & 0 \\ 0, & 1, & \dots & 0, & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0, & 0, & \dots & 1, & 0 \end{pmatrix}$$

$$H = G = [1, 0, \dots, 0]^T$$

한편, 음성의 인가 신호 $e_{h_t, s_t}(n)$ 과 잡음 $v(n)$ 는 상관관계가 존재하지 않는다(uncorrelated)고 가정 했으므로, 분산 행렬을 다음과 같이 표현할 수 있다.

$$E[e_{h_t, s_t}(n) e_{h_t, s_t}^T(n)] = Q(s_t = j, h_t = m) = [\sigma_{h_t, s_t}^2]$$

$$E[v(n) v^T(n)] = V = \sigma_v^2$$

$$E[e_{h_t, s_t}(n) v^T(n)] = 0$$

잡음 음성이 주어졌을 때, 아래 식과 같이 조건 평균 식으로 $\hat{Y}(n)$ 을 추정할 수 있다.

$$\begin{aligned} \hat{Y}(n) &= E\{Y(n) | z(n)\} \\ &= \int_{-\infty}^{\infty} Y(n) p(Y(n) | z(n)) dY(n) \end{aligned} \quad (5)$$

위 식 (5)에서 조건 분포 함수는 아래식과 같이 쓸 수 있다.

$$\begin{aligned} p(Y(n) | z(n)) &= \sum_{j=1}^L \sum_{m=1}^M (p(Y(n) | s_t = j, h_t = m, z(n)) \\ &\cdot p(s_t = j, h_t = m | z(n))) \end{aligned} \quad (6)$$

식 (6)을 식 (5)에 대입하고 적분과 합계를 바꾸면, $\hat{Y}(n)$ 추정은 식 (7)로 다시 쓸 수 있다.

$$\hat{Y}(n) = \sum_{j=1}^L \sum_{m=1}^M \hat{Y}_{h_t, s_t}(n) p(s_t = j, h_t = m | z(n)) \quad (7)$$

식 (7)에서 $\hat{Y}_{h_t, s_t}(n)$ 은 $s_t = j, h_t = m$ 일 때, $\hat{Y}(n)$ 의 조건 평균 추정식이며 칼만 필터식으로 얻을 수 있고, $p(s_t = j, h_t = m | z(n))$ 은 $\hat{Y}(n)$ 이 $s_t = j, h_t = m$ 일 때의 확률을 나타내며 베이시안 법칙과 확률의 성질을 이용해서 구할 수 있다. $\hat{Y}(n)$ 추정은 각각 L 상태, M 혼합 성분일 때의 조건 평균 추정 $\hat{Y}_{h_t, s_t}(n)$ 과 그 확률 $p(s_t = j, h_t = m | z(n))$ 의 곱의 합으로 구해진다. 그래서, HMM이 L 상태와 M 혼합 성분을 가지고 있는 경우, $\hat{Y}(n)$ 을 추정하는데, $L \times M$ 개의 칼만 필터를 계산해야 하므로, 많은 계산량이 요구된다. HMM에 기초를 둔 칼만 필터의 계산 과정 그림 1로 표현된다.

III. 음질 향상을 위한 좌-우향 은닉마코프모델에서 상태 (s) 결정

3.1. 좌-우향 은닉 마코프 모델에 기초한 음성 향상 방법

상태열 s_1, s_2, \dots, s_T 을 알고 있다고 가정하고, 현재 상태가 s_t^* 라고 주어진다면, 식 (7)을 좌-우향 은닉 마코프 모델식으로 아래와 같이 다시 쓸 수 있다.

$$\hat{Y}(n) = \sum_{j=1}^L \sum_{m=1}^M \hat{Y}_{h_t, s_t^*} \cdot p(s_t^* = j, h_t = m | z(n)) \quad (8)$$

위 식에서 s_t^* 를 결정하는 방법은 3.2절에서 설명할 것이다. 식 (7)과 비교하여 식 (8)은 $\hat{Y}(n)$ 을 구하기 위해서 2M개의 칼만 필터를 필요로 한다. 식 (7)에 비해 계산량이 (2/L)배 감소하는 것을 볼 수 있다. $\hat{Y}(n)$ 을 구하기 위해서 $\hat{Y}_{h_t, s_t^*}(n)$ 과 가중치 $p(s_t^*, h_t = m | z(n))$ 를 계산해야 하는데, s_t^* 를 알고 있을 때, $\hat{Y}_{h_t, s_t^*}(n)$ 을 계산하기 위한 칼만 필터 알고리즘은 아래식과 같이 쓸 수 있다.

$$\begin{aligned} \hat{Y}_{h_t, s_t^*}(n) &= F(s_t^*, h_t) \hat{Y}_{h_t, s_t^*}(n-1) \\ &+ K_{h_t, s_t^*}(n) \cdot \{z(n) - H^T F(s_t^*, h_t) \hat{Y}_{h_t, s_t^*}(n-1)\} \end{aligned} \quad (9)$$

$$\begin{aligned} \hat{M}_{h_t, s_t^*}(n) &= F(s_t^*, h_t) P_{h_t, s_t^*}(n-1) F^T(s_t^*, h_t) \\ &+ GQ(s_t^*, h_t)G^T \end{aligned} \quad (10)$$

$$K_{h_t | s_t^*}(n) = M_{h_t | s_t^*}(n) H^T [V + H M_{h_t | s_t^*}(n) H^T]^{-1} \quad (11)$$

$$P_{h_t | s_t^*}(n) = M_{h_t | s_t^*}(n) - K_{h_t | s_t^*}(n) H M_{h_t | s_t^*}(n) \quad (12)$$

가중치 요소 $p(s_t = j, h_t = m | z(t))$ 는 베이시안 법칙으로 아래와 같이 구할 수 있다.

$$p(s_t = j, h_t = m | z(t)) = \frac{p(z(t) | s_t = j, h_t = m, z(t-1)) p(s_t = j, h_t = m | z(t-1))}{p(z(t) | z(t-1))} \quad (13)$$

위 식에서 분자의 첫 번째 요소는 아래식으로 근사화된다.

$$p(z(t) | s_t = j, h_t = m, z(t-1)) = \prod_{n=1}^N p(z(n) | s_t = j, h_t = m) \quad (14)$$

또한, 위 식에서 우변은 아래와 같이 평균 $\hat{Y}_{m|j}(n)$ 과 분산 $HP_{m|j}H^T$ 을 갖는 정규분포 ($N[\cdot, \cdot]$)식으로 표현할 수 있다.

$$p(z(n) | s_t = j, h_t = m) = N[\hat{Y}_{m|j}(n), HP_{m|j}H^T] \quad (15)$$

식 (13)에서, $p(s_t = j, h_t = m | z(t-1))$ 는 마코프 과정으로 나타낼 수 있다.

$$p(s_t = j, h_t = m | z(t-1)) = \sum_{i=1}^M \{ p(s_t = j, h_t = m | s_t = i, h_t = l, z(t-1)) \cdot p(s_{t-1} = i, h_{t-1} = l, z(t-1)) \} \quad (16)$$

식 (16)에서 우변의 첫 번째 요소는 아래식으로 다시 쓸 수 있다.

$$p(s_t = j, h_t = m | s_{t-1} = i, h_{t-1} = m, z(t-1)) = p(h_t = m | s_t = j, s_{t-1} = i, h_{t-1} = l, z(t-1)) \cdot p(s_{t-1} = i, h_{t-1} = l, z(t-1)) \quad (17)$$

h_t 와 s_t 은 통계적으로 서로 독립이므로 위 식의 첫 번째와 두 번째 요소를 아래와 같이 다시 쓸 수 있다.

$$p(h_t = m | s_t = j, s_{t-1} = i, h_{t-1} = l, z(t-1)) = c_{m|j} \quad (18)$$

$$p(s_t = j | s_{t-1} = i, h_{t-1} = l, z(t-1)) = p(s_t = j | s_{t-1} = i) = a_{ij} \quad (19)$$

식 (18), (19)를 식 (16)에 대입하면 식 (20)과 같이 나타내어진다.

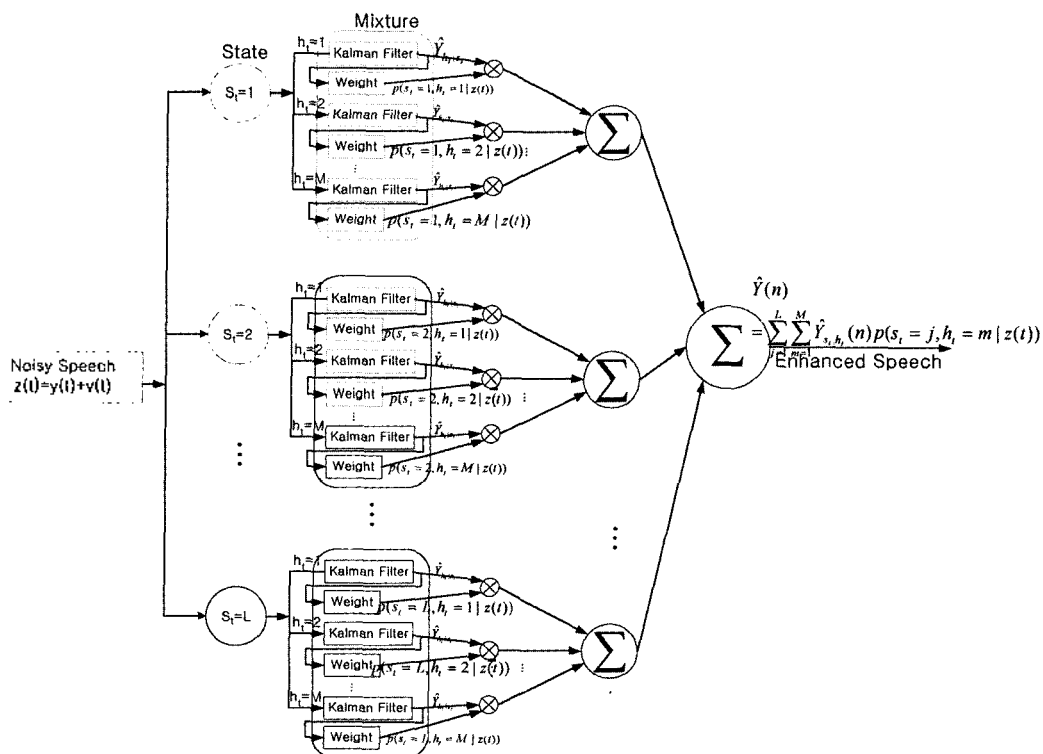


그림 1. 다중 칼만 필터 시스템
Fig. 1. Multiple Kalman filter system.

$$p(s_t = j, h_t = m | z(t-1)) = \sum_{l=1}^M a_{jl} c_{ml} p(s_{t-1} = i, h_{t-1} = l | z(t-1)) \quad (20)$$

식 (13)에서, 분모 $p(z(t) | z(t-1))$ 는 상태 j 에 독립적이므로, 이 요소는 스케일 인수가 된다. 그래서, 식 (13)은 최종적으로 아래식과 같이 구할 수 있다.

$$p(s_t = j, h_t = m | z(t)) = K_t N_{m|j} \sum_{l=1}^M a_{jl} c_{ml} p(s_{t-1} = i, h_{t-1} = l | z(t-1)) \quad (21)$$

식 (21)에서 K_t 는 가중치 요소들의 모든 합이 1이 되게 하는 스케일 인수이다.

$$\sum_{j=1}^M \sum_{m=1}^M p(s_t = j, h_t = m | z(t)) = 1$$

3.2. LRT에 의한 상태 결정

좌-우향 은닉 마코프 모델에 기반한 음질 향상 방법에서, 현재 상태 $s_t^* = j$ 를 알고 있다고 가정했다. 그러나 실제로 우리는 그 상태를 알 수 없다. 본 논문에서는 음성의 상태열 s_1, s_2, \dots, s_T 를 구하기 위해 확률값 $p(s_t = j | z(t))$ 를 사용하였다. 이때, $p(s_t = j | z(t))$ 는 식 (22)로 구할 수 있다.

$$p(s_t = j | z(t)) = \sum_{m=1}^M p(s_t = j, h_t = m | z(t)) \quad (22)$$

좌-우향 은닉 마코프 모델에서 과거의 음성 프레임 상태가 $s_{t-1} = j$ 라고 주어져 있을 때, 현재 음성 프레임의 상태 s_t 는 j 또는 $j+1$ 이 된다. 현재 프레임의 음성 상태 s_t 는 j 또는 $j+1$ 이라는 가정 하에 두 상태의 확률 값을 비교해서 구할 수 있다. 상태 j 와 $j+1$ 일 때의 확률 비에 로그를 취하여 유사도 (Likelihood Ratio) $\Lambda(t)$ 를 아래와 같이 정의한다.

$$\Lambda(t) = \log \left(\frac{p(s_t = j | z(t))}{p(s_t = j+1 | z(t))} \right) \begin{matrix} s_t = j \\ \geq 0 \\ s_t = j+1 \end{matrix} < 0 \quad (23)$$

위 식에서 $\Lambda(t) \geq 0$ 이면 현재 프레임의 상태가 j 일 때의 확률 $p(s_t = j | z(t))$ 이 $j+1$ 일 때의 확률 $p(s_t = j+1 | z(t))$ 보다 크다는 의미이므로, 현재 프레임의 상태는 $s_t = j$ 라고 결정되고, 반대로, $\Lambda(t) < 0$ 이면 현재 프레임의 상태가 $j+1$ 일 때의 확률이 j 일 때의 확률보다 크다는

의미가 되므로, $s_t = j+1$ 라고 결정된다.

좌-우향 은닉 마코프 모델에 기초를 둔 칼만 필터 계산과 상태 결정 알고리즘의 순서는 그림 2에 나타내었다.

IV. 실험 및 결과

제안된 방법은 입력신호의 신호 대 잡음비가 각각 0 dB, 5 dB, 15 dB일 때 백색 잡음이 부가되었을 때 음성향상 결과를 보여주고 있다.

학습은 5명의 남성과 5명의 여성이 각각 다른 3개의 문장을 발성하여 구성된 잡음이 없는 30개의 문장을 사용하였고, 테스트는 학습에 참여하지 않은 남녀 각 2명이 발성한 문장을 사용하였다. 이 실험에서, 샘플링 주파수는 11,025 Hz이고, 깨끗한 음성의 AR 모델은 15차이다. 마코프 모델에서 5 상태, 2 혼합성분을 사용하였고, 펜티엄-IV, CPU 2.0 GHz PC에서 실험하였다. 실험은 기존의 방법과 제안된 방법에 대해서, 입력신호 대 잡음비와 잡음 감소된 출력 신호 대 잡음비의 결과를 비교하고, 음

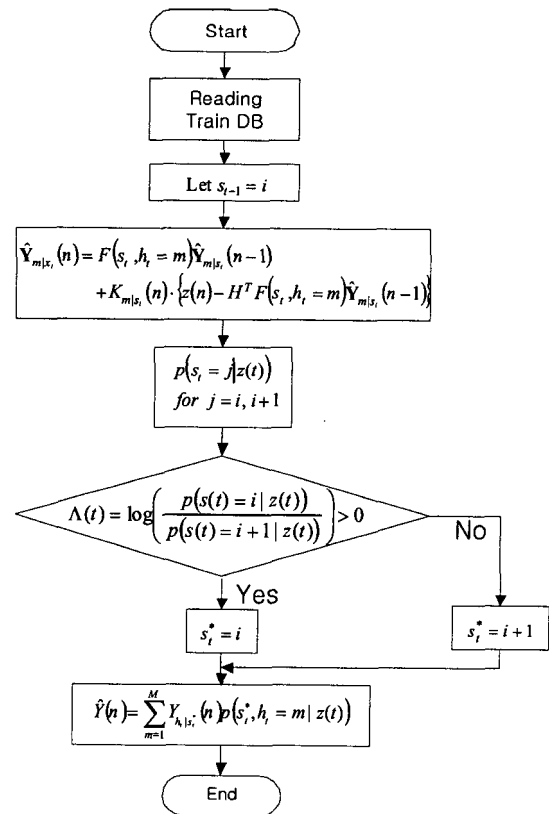


그림 2. LRT에 의한 상태 결정 알고리즘
Fig. 2. State decision algorithm based on LRT.

표 1. 기존 방법과 제안된 방법의 음질 향상 결과 비교 (상태: 5, 혼합성분: 2)

Table 1. The comparison between conventional method and proposed method (state: 5, mixture: 2).

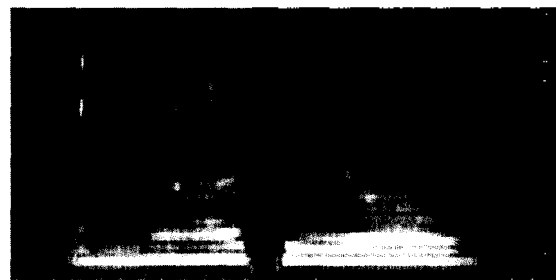
입력 신호 대 잡음비 [dB]	기존의 음질 향상 방법			상태 결정을 갖는 좌 우향 은닉 마이크 모델		
	출력 신호 대 잡음비 [dB]	필터 수	시간 [Sec]	출력 신호 대 잡음비 [dB]	필터 수	시간 [Sec]
0	8.50	L×M	7	8.30	2M	3
5	11.35			11.01		
10	15.00			14.60		
20	24.20			23.95		



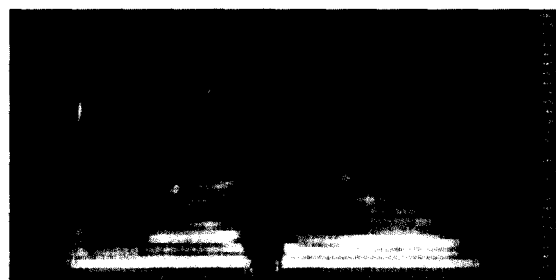
(a) Clean Speech



(b) 10 dB Noisy Speech



(c) Enhanced Speech (Conventional Method)



(d) Enhanced Speech (Proposed Method)

그림 3. 음질 향상 파형과 스펙트럼

Fig. 3. Speech enhancement results.

질 향상에 걸리는 시간을 비교하였다.

표 1은 각 입력 신호 대 잡음별로 음질이 향상된 후 출력 신호 대 잡음비 결과를 비교하고 있다. 전체적으로 기존의 방법에 비해 좌-우향 은닉 마코프 모델에 기초를 둔 칼만 필터 방법의 성능이 약 0.2~0.4 dB 정도 떨어지는 것을 볼 수 있으나, 계산 과정은 $(2/L)$ 배로 줄어드는 것을 볼 수 있다. 기존의 음질향상 방법에서의 실험 측정 시간은 약 7초 가량 소요되었지만, 제안된 방법에서는 3초의 시간이 소요된 것으로 보아 계산 과정이 줄어들음을 확인할 수 있다. 상태 결정 방법이 아직은 불안정하여 기존의 방법에 비해 성능이 저하되었지만 알고리즘의 안정화를 통하여 조금 더 개선될 수 있을 것으로 판단된다.

그림 3은 입력 신호 대 잡음비 10 dB 환경에서 음질 향상 결과 스펙트럼을 나타내고 있다. (a), (b)는 각각 깨끗한 음성과 잡음이 부가된 음성이고, (c), (d)는 각각 기존의 방법과 제안된 방법으로 음질을 향상시킨 결과를 나타내고 있다. (c), (d)에서 기존의 방법과 제안된 방법 모두에 대해서 (b)에서 보이는 잡음이 제거되어 음질이 향상됨을 볼 수 있다. (c)의 스펙트럼과 (d)의 스펙트럼을 비교해보면, 기존의 방법인 (c)가 출력 대 잡음비에서 0.2 dB 정도 더 향상되었다. 그러나, 향상된 파형을 청취한 결과 많은 차이가 나지 않는다.

V. 결론

본 논문에서는 기존의 은닉 마코프 모델을 이용한 음성 신호 향상 방법의 계산량을 줄이기 위한 방법으로 좌-우향 은닉 마코프 모델에서 유사도비 (Log-Likelihood Test)를 사용하여 상태열을 결정하는 알고리즘을 제안하였다. 좌-우향 은닉 마코프 모델에서 음성의 상태 j 를 알고 있을 때, 현재 프레임의 음성의 상태가 j 와 $j+1$ 일 때와 확률의 비를 LRT로 표현하고 LRT 값으로 현재 프레임의 상태를 결정하는 방법을 사용하여, 현재 음성의 상태를 결정한다. 현재 음성의 상태를 알고 있다면, 현재 상태와 바로 다음 상태에 대해서만 칼만 필터를 계산하므로, $2 \times M$ 개의 칼만 필터만 계산하면 된다. 제안된 방법은 기존의 HMM 모델을 이용한 방법에 비해 신호 대 잡음비는 0.2 dB~0.4 dB정도 떨어지나, $(2/L)$ 배로 계산량을 절감할 수 있었다. 상태 결정 방법이 아직은 불안정하므로, 안정화를 위한 연구가 좀 더 필요하며, 일반적인 HMM에 대한 적용 연구도 필요하다. 제안한 방법을 분장 종속 화자 인식 분야에서 개인별 음성향상 방법으로 발전시킨다

면, 화자 인식 분야에서 좀 더 좋은 성능을 얻을 수 있게 될 것이라 생각한다.

감사의 글

본 논문은 2003학년도 숭실대학교 교내 학술연구비 지원에 의하여 수행되었습니다.

참고 문헌

1. S. F. Boll, "Suppression of acoustics noise in speech using spectral subtraction," *IEEE Trans. Acoustic, Speech Signal Processing*, 27, 113-120, Apr., 1979.
2. J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoustic, Speech Signal Processing*, 26, 197-210, June, 1978.
3. K. Y. Lee and K. Shirai, "Efficient recursive estimation for speech enhancement in colored noise," *IEEE Signal Processing Letters*, 3 (7), 196-199, July, 1996.
4. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time amplitude estimator," *IEEE Trans. Acoustic, Speech Signal Processing*, 32 (6), 1109-1121, Dec., 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 Signal Processing*, 37 (12), 1846-1856, Dec., 1989.
6. Y. Ephraim, "A Bayesian approach for speech enhancement using hidden markov models," *IEEE Trans. Signal Processing*, 41, 725-735, Apr., 1992.
7. B. T. Logan and A. J. Robinson, "Noise estimation for enhancement and recognition within an autoregressive Hidden-Markov-Model framework," *Proceedings, 6th Australian International Conference on Speech Science and Technology, Adelaide, Australia, Dec., 1996*.
8. K. Y. Lee and J. Y. Rheem, "Smoothing approach using forward-backward Kalman filter with Markov Switching Parameters for Speech Enhancement," *Signal Processing*, 80, 2579-2588, Apr., 2000.
9. B. H. Juang, "Maximum-likelihood estimation for mixture multivariate stochastic observation of markov chains," *AT&T Technical Journal*, 64 (6), 1235-1248, July-August, 1984.
10. B. H. Juang and L. R. Rabiner, "Mixture autoregressive Hidden Markov Models for speech signals," *IEEE Trans. Acoustic, Speech Signal Processing*, 33 (6), 1404-1413, Dec., 1985.

저자 약력

• 이 기 용 (Ki-Yong Lee)

한국음향학회지 제15권 제3E호 참조
1997년 9월~현재 숭실대학교 정보통신전자공학부 부교수
※ 주관심분야: 음성신호 향상, 화자인식, 음성인식