

음성 향상을 위한 NPHMM을 갖는 IMM 알고리즘

IMM Algorithm with NPHMM for Speech Enhancement

이 기 용*

Ki-Yong Lee

ABSTRACT

The nonlinear speech enhancement method with interactive parallel-extended Kalman filter is applied to speech contaminated by additive white noise. To represent the nonlinear and nonstationary nature of speech, we assume that speech is the output of a nonlinear prediction HMM (NPHMM) combining both neural network and HMM. The NPHMM is a nonlinear autoregressive process whose time-varying parameters are controlled by a hidden Markov chain. The simulation results shows that the proposed method offers better performance gains relative to the previous results [6] with slightly increased complexity.

Keywords : Nonlinear speech enhancement, Parallel-extended Kalman filter, Nonlinear prediction HMM, Neural network

1. 서 론

자동 회귀(AR: autoregressive) 은닉 마코프 모델(HMM: Hidden Markov models) 방법은 오랫동안 음성 향상 기술의 기반이 되어 왔다[1-4]. 이러한 방법들에서 음성 신호는 시변 선형 예측 모델로 추정되었다. 성도를 모델링하기 위해서는 성도의 모양, 열 전도에 의한 성도의 비선형성, 성도 벽에서의 점성마찰(viscous friction)과 부드러움(softness), 입술에서의 방사(radiation), 비강결합(nasal coupling) 등과 같이 시간에 따라 변화하는 특성들을 고려해야 한다. 따라서, 전형적인 선형 ARHMM에 기반한 음성 향상기법을 사용한다면 음성의 비선형성과 비정상성을 정확하게 반영하기 힘들다[5].

최근에, 이러한 문제를 해결하기 위하여, 음성 향상을 신경 예측(Neural Predictive) HMM (NPHMM)에 기반된 접근방법들이 제안되었다[6]. 이러한 방법에서 음성신호는 마코프 체인

* 숭실대학교 정보통신전자공학부

(Markov chain)의 상태에 의존하는 파라미터들을 다층 퍼셉트론(Perceptron) 신경망(Neural Network)에 적용하여 비선형 예측 모델로 구성된다. 음성 향상을 위한 재귀적 추정 은 깨끗한 음성 신호로부터 학습된 NPHMM이 주어지면, 병렬 관계인 확장 칼만 필터들의 가중합으로 구성된다. 여기서, 병렬 필터들 사이의 상호작용은 무시된다. 이러한 방법을 SMM(SMM: separate multiple model) 알고리즘이라 한다.

스위칭 요소를 사용하는 베이지안 추정은 이전 파라미터수가 기하급수적으로 증가하기 때문에 어려운 점이 있다. 왜냐하면 최적 추정은 모든 가능한 파라미터의 과거값에 맞춰진 필터 뱅크와 관계가 있기 때문이다. 결과적으로, 가설 병합(hypotheses merging) 아이디어는 실용적인 준 최적(sub-optimal) 추정 알고리즘 설계를 위하여 필요하며, 가설 병합의 최대 도수(degree)는 IMM(interacting multiple models)에 의해 얻어진다 [7,8].

본 논문에서는 음성 향상을 위해 목표 추적 문제에서 널리 사용되는 IMM 알고리즘을 적용하였으며, IMM 알고리즘이 효율적인 면에서 병렬 필터들 사이의 상호작용을 다루고 있으므로, 계산량의 커다란 증가 없이 음성 향상 성능이 개선되었다.

2. 신경 예측 은닉 마코프 모델(NPHMM)에 기반한 음성 모델

NPHMM은 마코프 연결 상태와 조합된 파라미터를 가진 비선형 예측 모델이다. L -상태를 갖는 1 계 마코프 연결과 상태전이 행렬 $A = [a_{ij}]$, $i, j = 1, \dots, L$ 가 있다고 간주하자. 그러면, 시간 t 에서 i 상태에 있는 음성 데이터는 예측기 $h_{s(t)=i}(\cdot)$ 에 기반한 신경망에 의해 다음과 같이 정의된다.

$$x(t) = h_{s(t)=i}(X(t-1)) + e_{s(t)=i}(t), \quad (1)$$

여기에서 $X(t-1) = [x(t-1), \dots, x(t-p)]^T$ 는 과거 p 개의 관측열이고, 추진열(driving sequence) $e_{s(t)=i}(t)$ 는 평균이 0이고, 분산이 σ_i^2 인 가우시안 함수이다. $h_{s(t)=i}(\cdot)$ 는 i 상태의 예측에 기반한 정방향(feedforward) 신경망이고, 다음과 같이 정의 된다.

$$h_{s(t)=i}(X(t-1)) = \sum_{k=1}^K w_{k|i} g \left(\sum_{j=1}^p w_{k,j|i} x(t-j) \right) \quad (2)$$

여기에서 $w_{k|i}$ 는 출력과 은닉층 사이의 가중벡터이고, $w_{k,j|i}$ 는 입력층과 은닉층 사이의 가중 행렬이다. $g(\cdot)$ 는 sigmoid 함수와 같은 차분 비선형 함수이다. (2)에서 상태 모델과 유사한 구조는 여러 논문 [11][12]에서 소개되었다.

(2)에서, 어떤 스케일의 진폭이라도 이에 대한 관측 값을 예측하기 위해서, 비선형 활성화 함수

는 출력에 전혀 개입되지 않는다.

이러한 비선형 예측기는 예측 계수의 차수나 수가 같다면, 선형 예측기에 의해 더 좋은 예측 정확도를 얻을 수 있다는 내용이 [6]에서 제시되었다.

이 경우에, NPHMM을 결정하는 파라미터 집합은 $\lambda = \{\mathbf{A}, \sigma, \mathbf{W}\}$ 로 구성되고 $\sigma = [\sigma_1 \sigma_2 \dots \sigma_L]$, $\mathbf{W} = [W_1 W_2 \dots W_L]$, $W_i = [w_{1|i}, \dots, w_{K|i}, w_{1|i}, \dots, w_{K,p|i}]$ 이다.

초기 모델 λ' 에서 시작하여, 목적 함수(objective function)는 Baum re-estimation algorithm [1]에 의해 다음과 같이 얻어진다.

$$Q(\lambda, \lambda') = \sum_s P(S|X, \lambda') \log P(S|X, \lambda) \quad (3)$$

여기에서, $S = \{s(1), s(2), \dots, s(T)\}$, $s(t) \in \{1, 2, \dots, L\}$ 는 $X = \{x(1), x(2), \dots, x(T)\}$ 에 일치하는 상태열이다. 그러면, (1)을 사용하여 목적 함수를 다음식과 같이 단순화시킬 수 있다.

$$Q(\lambda) = \sum_{i,j=1}^L \sum_{t=1}^T \gamma_{ij}(t) \left[\log a_{ij} + \log \frac{1}{\sqrt{2\pi}\sigma_i} - \frac{(x(t) - h_i(\mathbf{X}(t-1)))^2}{2\sigma_i^2} \right] \quad (4)$$

여기에서, $\gamma_{ij}(t)$ 는 관측열과 모델 λ 가 주어지면, i -상태에서 j -상태로 천이하는 사후 확률값이다.

NPHMM에서, 망의 입출력 매핑에서의 변화는 HMM의 상태(state)에 일치하는 망(network)을 선택함으로써 실행된다. 즉, 망 파라미터는 상태(state)에 따라 변화되고, 망 파라미터는 다른 예측기와는 서로 공유되지 않는다.

$Q(\lambda)$ 을 최소화하기 위해서 모델 파라미터 $\lambda = \{\mathbf{A} = \{a_{ij}\}, \mathbf{W} = \{W_i\}, \sigma^2 = \{\sigma_j^2\}, i, j = 1, \dots, L\}$ 각각을 미분함으로써, 우리는 [6]에서의 재추정식으로부터 모델 파라미터를 구한다.

음성 향상을 위하여 깨끗한 음성과 백색잡음이 부가된 음성 $x(t)$ 로부터 NPHMM의 학습 파라미터가 주어지면, 시간 t 에서 마코프 상태 $s(t) \in \{1, \dots, L\}$ 로 비선형 상태 공간(nonlinear state space) 모델을 다음과 같이 만들 수 있다.

$$\mathbf{x}(t) = f_{s(t)}(\mathbf{x}(t-1)) + G e_{s(t)}(t), \quad (5)$$

$$z(t) = H^T \mathbf{x}(t) + v(t) \quad (6)$$

여기에서 $\mathbf{x}(t) = [x(t), \dots, x(t-p)]^T$, $f_{s(t)}(\mathbf{x}(t-1)) = [h_{s(t)}(\mathbf{x}(t-1)), x(t-1), \dots, x(t-p)]^T$ 이고, H 와 G 는 $n \times n$ 행렬로 $H = [1 \ 0 \dots 0]^T$, $G = [1 \ 0 \dots 0]^T$ 이다.

음성에 부가된 잡음 신호 $v(t)$ 는 음성신호와 통계적으로 독립이고, 평균이 0이고 분산이 $R(t)$

인 백색 가우시안 잡음이라고 가정하자. 또한, $e_{s(t)}(t)$ 와 $v(t)$ 는 상관성이 없다고 가정한다.

3. 병렬 확장 칼만 필터를 이용한 비선형 음성 향상

잡음이 부가된 신호 $\mathbf{z}(t) = \{z(1), \dots, z(t)\}$ 에서 $\mathbf{x}(t)$ 의 MMSE 추정치는 조건 평균에 의해 주어진다.

$$\hat{\mathbf{x}}(t) = E\{\mathbf{x}(t)|\mathbf{z}(t)\} = \int_{-\infty}^{\infty} \mathbf{x}(t)p(\mathbf{x}(t)|\mathbf{z}(t))d\mathbf{x}(t) \quad (7)$$

$k = (k_0, k_1, \dots, k_t)$ 일 때, $Hist(t, k) \equiv \{s(0) = k_0, s(1) = k_1, \dots, s(t) = k_t\}$ 는 모든 가능한 순서열 공간에서 얻어진 모델들의 특정 순서열이라고 가정하자. 조건 평균 $E\{\mathbf{x}(t)|\mathbf{z}(t)\}$ 는 특정 히스토리들(histories)로 주어진 전체 조건 평균들의 가중합으로써 계산될 수 있다.

$$E\{\mathbf{x}(t)|\mathbf{z}(t)\} = \sum_k E\{\mathbf{x}(t)|Hist(t, k), \mathbf{z}(t)\} \cdot p(Hist(t, k)|\mathbf{z}(t)) \quad (8)$$

식 (8)의 오른쪽 항에서 각 조건 평균은 상응되는 칼만 필터를 사용함으로써 얻을 수 있다. 그러나 이러한 히스토리들 수는 시간에 따라 급격하게 증가되므로, 최적 해를 얻기 어려워, 어떤 근사화(approximation) 방법이 요구되는데, 가장 효율적인 방법으로 IMM이 있다[5].

식 (7)의 조건 확률 분포(Conditional probability density) 함수는 다음과 같이 표현할 수 있다.

$$p(\mathbf{x}(t)|\mathbf{z}(t)) = \sum_{j=1}^L p(\mathbf{x}(t)|s(t)=j, \mathbf{z}(t))p(s(t)=j|\mathbf{z}(t)) \quad (9)$$

식 (9)를 식(7)에 대입함으로써 다음과 같이 $\hat{\mathbf{x}}(t)$ 를 추정할 수 있다.

$$\hat{\mathbf{x}}(t) = \sum_{j=1}^L \hat{\mathbf{x}}_j(t)p(s(t)=j|\mathbf{z}(t)) \quad (10)$$

여기서 $\hat{\mathbf{x}}_j(t) = E\{\mathbf{x}(t)|s(t)=j, \mathbf{z}(t)\}$ 이다.

3.1 병렬 확장 칼만 필터에 의한 $\bar{\mathbf{x}}_j(t)$ 추정

식(10)의 추정기 $\hat{\mathbf{x}}(t)$ 는 L 개의 각 추정기 $\hat{\mathbf{x}}_j(t)$ 의 가중합으로 표현된다. 가중 요소 $p(s(t)=j|\mathbf{z}(t))$ 는 잡음이 부가된 관측열에서의 상태 확률이다. 각 추정기 $\hat{\mathbf{x}}_j(t)$ 는 확장된 칼만

필터[6]와 IMM[8]으로부터 찾을 수 있다.

$$\hat{\mathbf{x}}_j(t) = f_j(\bar{\mathbf{x}}_j(t-1)) + K_j(t) \{z(t) - H^T f_j(\bar{\mathbf{x}}_j(t-1))\} \quad (11)$$

$$M_j(t) = F_j(\bar{\mathbf{x}}_j(t-1)) \bar{P}_j(t-1) F_j^T(\bar{\mathbf{x}}_j(t-1)) + GQ(s(t)=j)G^T \quad (12)$$

$$K_j(t) = M_j(t) H^T [HM_j(t)H^T + R(t)]^{-1}, \quad (13)$$

$$P_j(t) = M_j(t) - K_j(t)HM_j(t), \quad (14)$$

여기에서,

$$F_j(\bar{\mathbf{x}}_j(t-1)) = \frac{\partial}{\partial \mathbf{x}} F_{s(t)}(\mathbf{x}(t-1)) \Big|_{\mathbf{x}(t-1)=\bar{\mathbf{x}}_j(t-1)} \quad (15)$$

$$\bar{\mathbf{x}}_j(t-1) = E\{\mathbf{x}(t-1) | s(t)=j, \mathbf{z}(t-1)\}, \quad (15)$$

$$\bar{P}_j(t-1) = E\{\mathbf{x}(t-1) - \bar{\mathbf{x}}_j(t-1) [\dots]^T | s(t)=j, \mathbf{z}(t-1)\}. \quad (16)$$

이다. 먼저 $\bar{\mathbf{x}}_j(t-1)$ 과 $\bar{P}_j(t-1)$ 을 유도하기 위해서, 전체 확률의 기본이 되는 다음 식을 계산한다.

$$p(\mathbf{x}(t-1) | s(t)=j, \mathbf{z}(t-1)) = \sum_i \{p(\mathbf{x}(t-1) | s(t-1)=i, s(t)=j, \mathbf{z}(t-1)) \cdot p(s(t-1)=i | s(t)=j, \mathbf{z}(t-1))\} \quad (17)$$

이 단계에서 수식을 다음 식으로 근사화 할 수 있다.

$$p(\mathbf{x}(t-1) | s(t)=j, \mathbf{z}(t-1)) \approx N[\bar{\mathbf{x}}_j(t-1), \bar{P}_j(t-1)] \quad (18)$$

여기서 $N[\dots]$ 는 정규 분포이다.

만약 $s(t)$ 을 알고 있다면, $s(t)$ 가 $\mathbf{x}(t-1)$ 와 독립이므로 다음 식을 쉽게 얻을 수 있다.

$$p(\mathbf{x}(t-1) | s(t-1)=i, s(t)=j, \mathbf{z}(t-1)) = p(\mathbf{x}(t-1) | s(t-1)=i, \mathbf{z}(t-1)). \quad (19)$$

다음 식에 위의 식을 대입하면

$$p(s(t-1)=i | s(t)=j, \mathbf{z}(t-1)) = a_{ij} p(s(t-1)=i | \mathbf{z}(t-1)) / p(s(t)=j | \mathbf{z}(t-1)) \quad (20)$$

가 된다. (17)식으로부터,

$$p(x(t-1)|s(t)=j, z(t-1)) = \frac{\sum_i a_{ij} p(s(t-1)=i|z(t-1)) \cdot p(x(t-1)|s(t-1)=i, z(t-1))}{p(s(t)=j|z(t-1))} \quad (21)$$

를 구하고, (21)식의 분모는 다음과 같이 정의된다.

$$p(s(t)=j|z(t-1)) = \sum_i a_{ij} p(s(t-1)=i|z(t-1)) \quad (22)$$

식 (15), (16), (21)로부터, 혼합된 초기 조건들 $\bar{x}_j(t-1)$ 와 $s(t)=j$ 일 때 해당되는 칼만 필터를 위한 $\bar{P}_j(t-1)$ 을 다음과 같이 계산한다.

$$\bar{x}_j(t-1) = \sum_i a_{ij} \hat{x}(t-1) p(s(t-1)=i|z(t-1)) / p(s(t)=j|z(t-1)) \quad (23)$$

$$\bar{P}_j(t-1) = \frac{\sum_i a_{ij} (p(s(t-1)=i|z(t-1))) \cdot [P_i(t-1) + [\hat{x}_i(t-1) - \bar{x}_j(t-1)][\dots]^T]}{p(s(t)=j|z(t-1))} \quad (24)$$

3.2 가중치 $p(s(t)=j|z(t))$ 계산

식 (10)의 가중 요소인 $p(s(t)=j|z(t-1))$ 는 베이시안 법칙에 의해 얻을 수 있다.

$$p(s(t)=j|z(t)) = \frac{p(z(t)|s(t)=j, z(t-1)) p(s(t)=j|z(t-1))}{p(z(t)|z(t-1))} \quad (25)$$

그리고, 식 (25)의 분자의 첫 번째 식은 다음과 같이 근사화된다.

$$p(z(t)|s(t)=j, z(t-1)) \approx N[H^T f_j(\bar{x}_j(t-1)), \Sigma_j(t)] \quad (26)$$

위의 식 (26)을 식 (25)에 대입함으로써 식 (22)을 사용하여 재귀적으로 $p(s(t)=j|z(t-1))$ 를 갱신할 수 있다.

$$p(s(t)=j|z(t)) = D_j N\{H^T f_j(\bar{x}_j(t-1)), \Sigma_j(t)\} \cdot p(s(t)=j|z(t-1)) \quad (27)$$

여기에서 D_j 는 $\sum p(s(t)=j|z(t))=1$ 을 보장하는 정규화 상수로 정의한다. $j=1, \dots, L$ 일 때, 초기 조건은 $p(s(0)=j|z(0))=1/L$, $\hat{x}_j(0)=\mathbf{0}$, $P_j(0)=\mathbf{0}$ 이다. 음성향상을 위한 NPHMM을 이용한 IMM 알고리즘 방법을 그림 1에 나타내었다. 첫 번째 단계는 식 (21), (23), (24)에 의해 가중되는 단계이고, 다음의 칼만 필터 단계는 식 (11)~(15)로 수행된다. 그러면 식 (27)로 확률

계산을 한 후 마지막으로 출력이 식 (10)에서 생성된다. 식 (23), (24)에 표현되었고, IMM 알고리즘의 중요한 부분인 혼합(mixing)은 SMM 알고리즘에서는 찾을 수 없다.

음성 향상률을 높이기 위해 $(t+p-1)$ 번째까지 $\hat{x}(t)$ 의 계산을 지연시켰다. 마지막으로 향상된 음성 신호는 다음 식으로 구할 수 있다.

$$\hat{x}(t) = \begin{bmatrix} 0 & \cdots & 1 \end{bmatrix} \hat{x}(t+p-1). \quad (28)$$

4. 실험 결과

제안한 방법은 입력 SNR이 각각 5,10,15,20 dB인 백색 잡음과 유색 잡음을 컴퓨터로 자동 생성시켜, 음성신호에 부가시킨 후 이 신호를 향상하는 방법으로 실험하였다. 여기서 SNR은 신호의 평균에너지와 잡음의 평균 에너지의 비율로써 정의한다. 잡음이 정적(stationary)이라고 가정하고, 잡음의 분산은 오직 잡음만 포함하고 있는 신호의 초기 부분으로부터 향상 추정한다. 유색 잡음은 [10]에서 소개한 방법을 사용하여 생성하였다.

음성을 위한 NPHMM의 등록은 3 명의 남성과 3 명의 여성 화자로서 깨끗한 한국어 음성의 12 문장을 사용하여 수행하였다. 테스트 단에서 사용한 음성은 등록과정에서 사용하지 않은 화자의 음성을 사용하였다. 음성 향상 실험은 한 명의 남성과 다른 여성 화자들로서 발생된 4 개 테스트 문장을 사용하여 수행하였다.

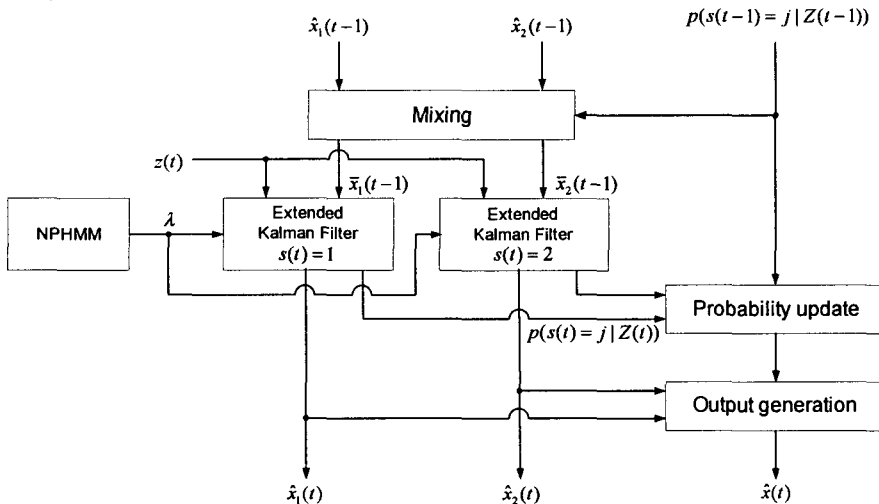


그림 1. L=2일 때, 음성 향상을 위한 NPHMM을 이용한 IMM algorithm

실험에서 음성은 11 kHz로 샘플링되었고 MLP는 출력 층에서 1 개, 은닉 층에서는 2 개, 입력계층에서는 12 개의 단위이고, NPHMM의 상태 수는 8 개를 사용하여 재귀적으로 계산하는 하나의 은닉 정방향 신경망(a single hidden-layer feedforward neural network)이다.

표 1은 선형 IMM과 제안된 방법의 성능을 비교한 결과이다. 백색 잡음과 유색잡음 모두에서 제안된 방법의 성능이 향상됨을 알 수 있다. 성능차이는 SNR비가 높아 질수록 기존의 방법과 제안한 방법의 차이는 점차 감소되었다. 표 1에서 다양한 입력 SNR이 주어진 경우, 성별에 따른 성능 차이는 비슷하였고, 출력 SNR은 약 1~2 dB 정도 개선되었다. 입력 SNR이 더 작을수록 제안한 방법의 성능이 더욱 향상됨을 알 수 있었다.

그림 2, 3, 4에서는 음성 신호와 실험결과를 나타낸 것이다. 그림 2에서 (a)는 깨끗한 음성 신호이고, (b), (c)는 백색잡음과 유색 잡음을 첨가한 음성 신호이다. 깨끗한 음성신호(a)의 스펙트로그램을 (d)에서 보여준다. 그림 3은 백색 잡음과 유색 잡음에 대한 향상된 음성 결과를 비교하였고, 그림 4는 잡음에 오염된 신호와 향상된 신호의 스펙트로그램이다. 백색 잡음의 신호를 위해 잡음이 첨가된 음성의 스펙트로그램, 제안된 방법, 전통적 방법에 의해 향상된 음성을 그림 3에서 비교하였다. 유색 잡음이 첨가된 음성도 같은 방법으로 잡음에 오염된 신호와 향상된 신호의 결과를 그림 4에서 보여주고 있다. 비록 평균 출력 SNR이 1.5 dB보다 낮지만, 상태 천이 영역 동안에는 잡음을 더욱 더 억압하기 때문에 음성의 질에서 현저한 개선이 있었다. 그러나 스펙트로그램상에서는 거의 비슷한 그림을 보이는데, SNR 1 dB 미만의 경우 스펙트로그램에서 쉽게 구분할 수 없었다.

비공식적 청취 성능 실험은 15 명 화자들과 3 문장들로서 수행한 것에서 IMM보다 제안한 방법이 더 좋은 성능을 보여주었다. 입력 SNR이 10 dB일 때, 실험 화자는 임의의 순서에서 같은 문장의 두 견해로 들었다. 45 가지의 경우 중 35 경우가 제안된 방법의 성능이 더 좋았다고 대답했다.

5. 결 과

음성 신호의 음성 향상을 위하여 주로 시변 선형 예측 모델을 사용하였다. 그러나 음성은 비선형성과 비정상성의 특성을 가지고 있으므로, 기존의 전형적인 자동회귀 선형방법으로는 해결하기 어렵다. 이러한 문제를 해결하기 위하여 신경 예측에 기반된 접근 방법을 사용한다. 백색 잡음과 유색 잡음이 부가되어 손상된 음성 신호의 향상을 위하여 IMM 알고리즘에 기반한 효율적인 재귀알고리즘을 제안하였다. 음성 향상을 위해 목표 추적 문제에서 널리 사용되는 IMM 알고리즘을 적용하였으며, IMM 알고리즘이 효율적인 면에서 병렬 필터들 사이의 상호작용을 다루고 있으므로, 계산량의 커다란 증가 없이 음성 향상 성능이 개선되었다. 제안된 방법에서 음성의

추정은 병렬 확장 칼만 필터들의 가중된 합으로 얻어졌다. 이 필터들은 분리되어 동작하지 않고, 상호적으로 동작한다. 음성 향상 성능은 병렬 필터들 사이의 상호작용을 고려함으로써 개선된다. 선형 IMM과 제안된 방법의 성능을 비교한 결과, 다양한 입력 SNR이 주어진 경우 출력 SNR은 약 1~2 dB 정도 개선되었다.

6. 감사의 글

본 논문은 2004학년도 숭실대학교 교내학술연구비 지원에 의하여 수행되었습니다.

참 고 문 헌

- [1] Ephraim, Y. 1992. "Statistical-model-based speech enhancement systems." *Proc. IEEE*, Vol. 80, pp. 1526-1555.
- [2] Sheikhzadeh, H. & Deng, L. 1994. "Waveform-based speech recognition using hidden filter models: parameter selection and sensitivity to power normalization." *IEEE Trans. Speech and Audio Processing*, Vol. 1, no. 2, pp. 80-89.
- [3] Lee, K. Y. & Shirai, K. 1996. "Recursive estimation for speech enhancement in colored noise." *IEEE Signal Processing Letters*, Vol. 3, pp. 196-199.
- [4] Lee, K. Y. & Jung, S. 2000. "Time domain approach using multiple Kalman filters and its application to speech enhancement." *IEEE Trans. Speech and Audio Processing*, Vol. 8, pp. 282-291.
- [5] Kumar, A., Mullick, S. K. 1996. "Nonlinear dynamical analysis of speech." *Jour. the Acoustic Society of America*, Vol. 100, No 1, pp. 615-629.
- [6] Lee, K. Y., McLaughlin, S., Shirai, K. 1998. "Speech enhancement based on neural predictive hidden Markov model." *Signal Processing*, Vol. 65, pp. 373-381.
- [7] Blom, H. A. P. & Bar-Shalom, Y. 1988. "The interacting multiple model algorithm for systems with Markovian switching coefficients." *IEEE Automatic Control*, Vol. 33, pp. 780-783.
- [8] Kim, J., Lee, K. Y., Lee, C. 2000. "On the applications of IMM algorithm for enhancing noisy speech." *IEEE Trans. Speech and Audio Processing*, Vol. 8, pp. 349-352.
- [9] Jazwinski, A. H. 1970. *Stochastic Process and Filtering Theory*. New York: Academic.
- [10] Gibson, J. D., Koo, B. & Gray, S. D. 1991. "Filtering of colored noise for speech enhancement and coding." *IEEE Trans. Signal Processing*, Vol. 39, pp. 1732-1742.
- [11] Deng, L. & Ma, Z. 2000. "Spontaneous speech recognition using a statistical coarticulatory model for the hidden vocal-tract-resonance dynamics." *Jour. the Acoustic Society of America*, Vol. 108, No. 6, pp. 3036-3048.

- [12] Deng, L., Hassanein, K. & Elmasry, M. 1994. "Analysis of correlation structure for a neural predictive model with application to speech recognition", *Neural Networks*, Vol. 7, No. 2, pp. 331-339.

접수일자: 2004. 11. 01

게재결정: 2004. 11. 28

▲ 이기용

서울시 동작구 상도 5동 1-1 (우: 156-743)

숭실대학교 정보통신 전자공학부

Tel: +82-2-820-0908, Fax: +82-2-817-4591

E-mail: kylee@ssu.ac.kr

표 1. 제안된 방법과 전형적 방법의 SNR[dB] 성능

(a) 여자

SNR [dB]		0	5	10	15	20
White Noise	proposed method [dB]	8.84	12.15	15.79	19.64	23.25
	conventional method [dB]	7.79	11.17	14.92	18.70	22.79
Colored Noise	proposed method [dB]	9.53	14.45	17.84	20.75	24.06
	conventional method [dB]	8.32	13.03	16.09	19.79	23.13

(b) 남자

SNR [dB]		0	5	10	15	20
White Noise	proposed method [dB]	9.76	13.32	16.50	20.25	23.90
	conventional method [dB]	8.13	12.16	14.95	18.36	22.94
Colored Noise	proposed method [dB]	10.77	13.85	17.26	21.26	24.67
	conventional method [dB]	8.68	11.82	15.62	19.40	23.58

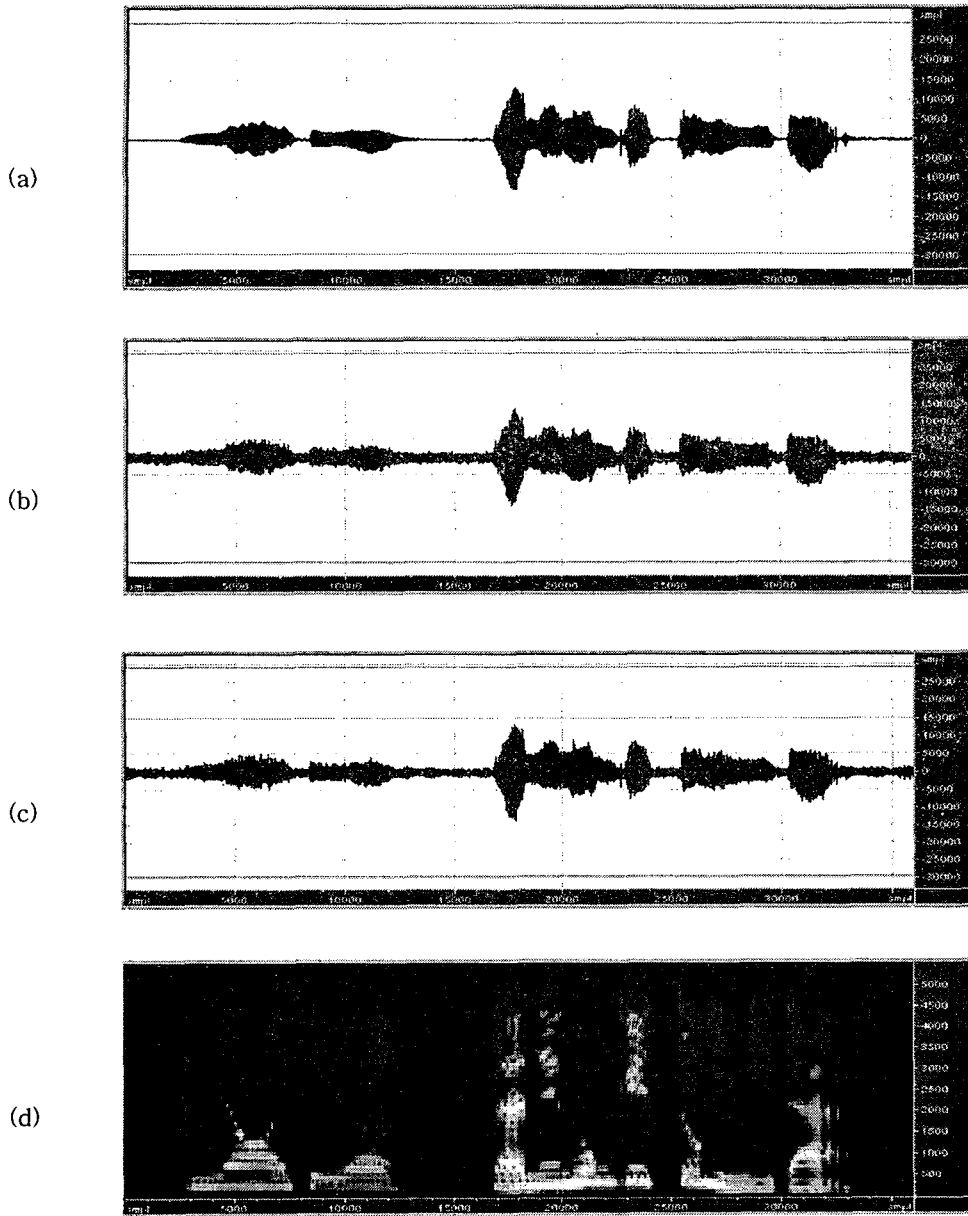


그림. 2 (a) clean speech waveform, (b) noisy speech waveform with 10dB white noise, (c) noisy speech waveform with 10 dB colored noise, (d) spectrogram of clean speech

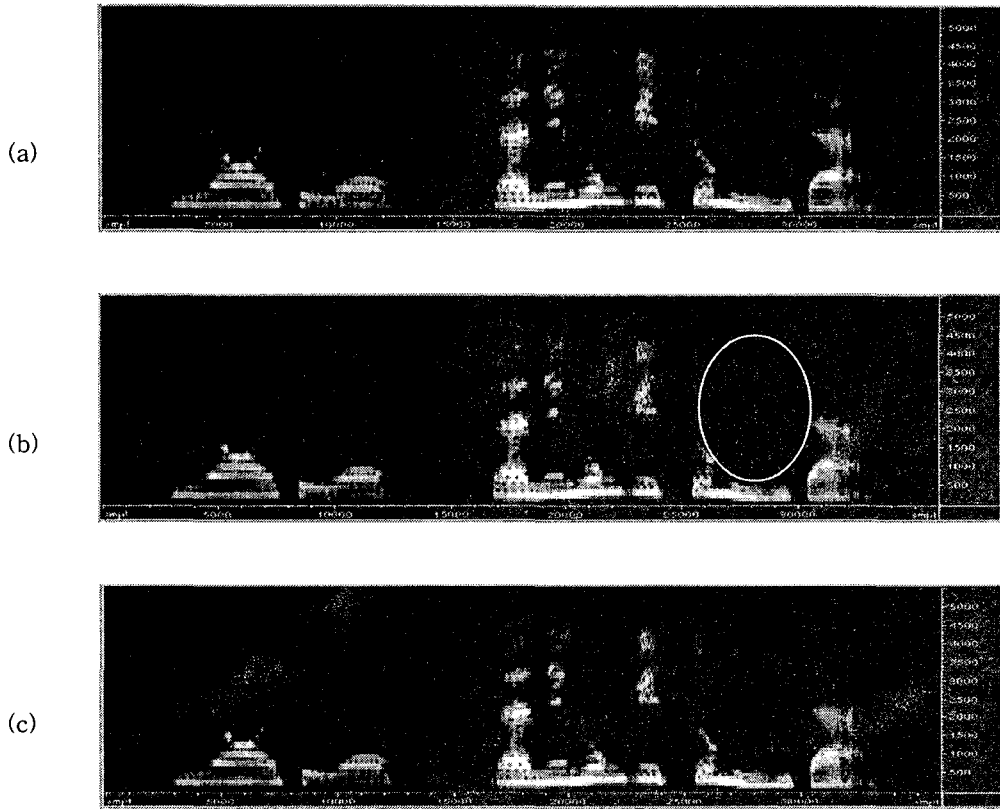


그림 3. Spectrograms of (a) white noisy speech, (b) enhanced speech by proposed method, (c) enhanced speech by conventional method

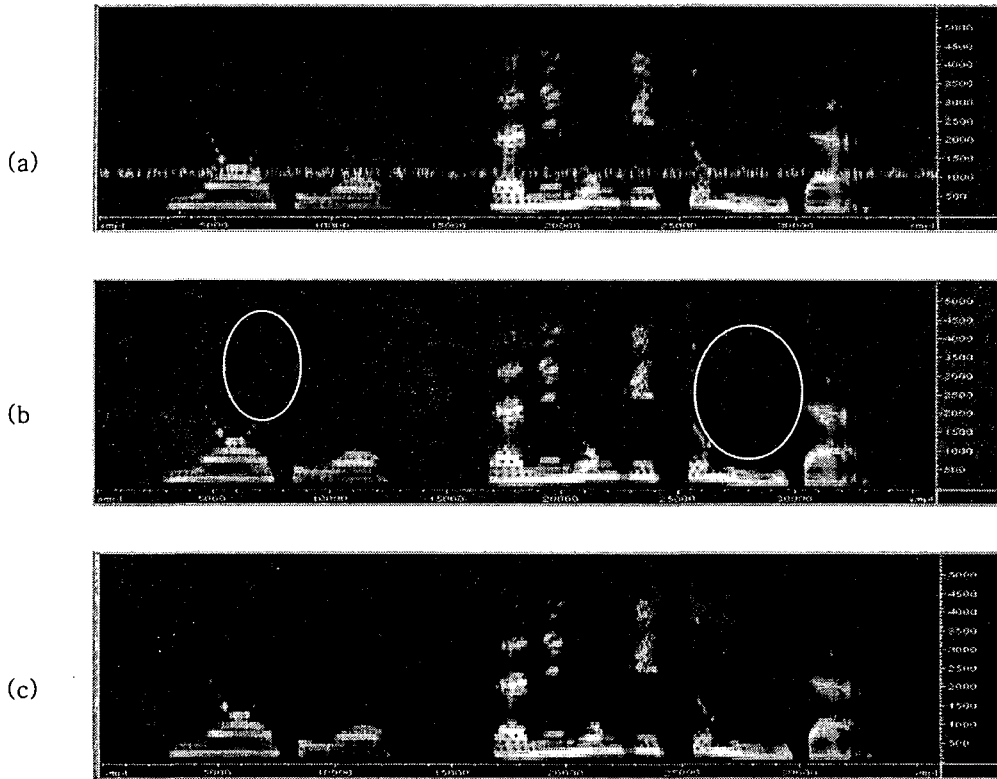


그림 4. Spectrograms of (a) colored noisy speech, (b) enhanced speech by proposed method (c) enhanced speech by conventional method