다중 H∞ 필터에 의한 강인한 음성향상

김준일, 이기용
숭실대학교 정보통신전자공학부

Robust Speech Enhancement By Multi H∞ Filter

Jun Il Kim, Ki Yong Lee
School of Electronic Engineering, SoongSil University
junilkim@ctsp.ssu.ac.kr, kylee@ssu.ac.kr

요약

칼만/위너 필터 같은 기존의 음성향상 알고리즘은 잡음의 선험적 지식을 요구하고, 음성신호와 추정신호의 오차분산을 최소화하는데 중점을 두었다. 따라서, 잡음에 대한 통계적 추정에 오류가 있을 경우 결과에 악영향을 미칠 수 있다. 그러나 H∞ 필터는 잡음에 대한 어떠한 가정이나 선험적 지식을 요구하지 않는다. H∞ 필터는 최소상계(Upper Bound Least)를 적용하여 추정된 모든 신호들로부터 최소 에러 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 음성신호를 구한다. 음성의 통계적 특성을 이용하여 모델 파라미터를 추정하는 은닉 마코프 모델과 잡음의 변화에 강인한 H∞ 알고리즘을 사용해서, 다중 H∞ 필터에 의한 강인한 음성향상 방법을 제안하였다.

1. 서론

음성향상이란, 입력 신호가 배경잡음에 의해 오염되었을 때, 음성통신 시스템에서 성능을 향상시키고, 잡음의 영향을 최소화 시키는 것이다. 칼만/위너 필터 같은 음성의 통계적 특성을 이용한 음성향상 방법은 음성신호의 추정에러의 분산을 최소화하는데 중점을 두고 있다. 잡음의 선험적 지식을 가지고 있을 때, 칼만/위너 필터는 최적화된 알고리즘이다. 그러나, 잡음에 대한 가정이 잘 못 되었을 경우 결과에 악영향을 미칠 수 있다. 그러나, H∞ 필터는 잡음에 대한 어떠한 가정이나 선험적 지식이 필요 없다. H∞ 필터는 최소상계를 적용하여 추정된 모든 신호들로부터 최소 에러 신호를 갖는 최상의 추정신호를 찾아내므로 칼만/위너 필터보다 잡음의 변화에 강인하다. 최근 Ephraim은 은닉 마코프 모델(HMM)과 칼만 필터를 이용한 음성향상 방법을 제안하였다.[2] 그러나, 잡음에 대한 통계적 특성을 잘 못 추정했을 경우 칼만 필터의 특성상 결과에 악영향을 가져온다. 본 논문에서는 음성의 통계적 특성을 이용하여 모델 파라미터를 추정하는 은닉 마코프 모델과 잡음의 변화에 강인한 H∞ 알고리즘을 사용해서, 다중 H∞ 필터에 의한 강인한 음성향상 방법을 제안하였다. 이 방법은 기존의 방법에 비해 계산량은 약간 증가하나, SNR이 0.5~2.5dB 향상되었다. 본 논문은 2장은 칼만과 H∞ 필터 알고리즘, 3장은 다중 H∞ 필터에 의한 음성향상, 4장은 실험 및 결과, 5장에서는 결론을 맺었다.

2. 칼만과 H∞ 필터 알고리즘

잡음신호 $s(k)$는 식(1)과 같이 표현할 수 있다.

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

여기서, $v(k)$는 배경잡음이고, $y(k)$는 식(2)로 표현할 수 있다.

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

p는 차수이고, $a(j)$는 AR 계수이다.

식(1)-(2)를 상태방정식으로 나타내면 식(3)-(4)와 같다.

$$X(k) = AX(k-1) + Bw(k) \tag{3}$$

$$s(k) = CX(k) + v(k) \tag{4}$$

여기서, $X(k) = [x(k)x(k-1)\cdots x(k-p+1)]^T$

$$A=\begin{pmatrix} a(1) & a(2) & \cdots & a(p-1) & a(p) \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{p\times p}$$

$$B^T = C = [1\ 0\ \cdots\ 0\ 0]_{1\times p}$$

2.1 칼만 필터 알고리즘

칼만 필터에서는 $w(k)$와 $v(k)$는 평균이 0이고, 둘 사이의 상관관계가 없다(uncorrelated)고 가정했으므로, 분산 행렬을 아래와 같이 표현할 수 있다.

$$E\{w(k)w(k)^T\} = W,\ E\{v(k)v(k)^T\} = V$$

$$E\{w(k)v(k)^T\} = 0$$

칼만 필터는 에러 공분산이 최소값을 갖도록, 잡음 신호에서 최적의 추정 신호를 결정한다.

상태방정식 (3)-(4)에 대해서 칼만 필터 알고리즘은 식(7)로 표현할 수 있다.

$$\hat{X}(k) = A\hat{X}(k-1) + K(k)[s(k) - CA\hat{X}(k-1)] \quad (5)$$

필터 Gain과 에러 분산식은 다음과 같이 표현할 수 있다.

$$K(k) = P(k|k-1)C[V + CP(k|k-1)C^T]^{-1} \quad (6)$$

$$P(k|k-1) = AP(k-1)A^T + BWB^T \quad (7)$$

$$P(k) = [I - K(k)C]P(k|k-1) \quad (8)$$

여기서, $\hat{X}(0) = [0]_{p\times 1}$, $P(0) = [0]_{p\times p}$이고, I는 $p\times p$ 단위행렬이다.

2.2 H∞ 필터 알고리즘

H∞ 필터는 $w(k)$와 $v(k)$에 대한 어떠한 가정도 하지 않는다. 임의의 $v(k), w(k) \in l_2, X_0 \in \mathbb{R}^n$에 대해서 선형 결합된 추정신호열 $X(k)$에서 최소 추정 에러값을 갖는 $\hat{z}(k)$를 구한다.

$$\hat{z}(k) = L\hat{X}(k) \qquad L \in \mathbb{R}^{1\times n} \quad (9)$$

성능평가 기준은 아래와 같이 표현할 수 있다.

$$J = \frac{\sum_{k=0}^{N-1} |z_k - \hat{z}_k|_Q^2}{\left|X_0 - \hat{X}_0\right|_{p_0^{-1}}^2 + \sum_{k=0}^{N-1}\left\{|w_k|_{W^{-1}}^2 + |v_k|_{V^{-1}}^2\right\}} \quad (10)$$

여기서 $Q \geq 0,\ W \geq 0,\ V \geq 0,\ p_0^{-1} \geq 0$은 설계자가 임의로 주는 가중치 행렬이다.

H∞ 필터는 모든 가능한 추정신호 $\hat{z}(k)$중에서 식(11)에 만족하고, 오차신호를 최소화하는 최적화된 $\hat{z}(k)$를 찾아낸다.

$$sup J \leq \gamma^2 \quad (11)$$

여기서, γ은 잡음 감쇄 정도를 나타낸다.

H∞ 필터는 식(11)처럼 최소상계(Upper Bound Least)를 적용하여, 균일한 오차신호를 제공한다. 따라서 칼만 필터보다 잡음의 변화에 강인하다.

$\gamma > 0$이고, 대칭 방정식 $P(k) > 0$을 만족할 때, $P(k)$를 Ricatti 방정식으로 식(12)와 같이 표현할 수 있다.

$$P(k+1) = AP(k)[I - \gamma^{-2}\overline{Q}P(k) + C^TV^{-1}CP(k)]^{-1} + BWB^T \quad (12)$$

식(9)에서 $\hat{X}(k)$는 식(13)으로 추정할 수 있다.

$$\hat{X}(k) = A\hat{X}(k-1) + H(k)(s(k) - CA\hat{X}(k-1)) \quad (13)$$

$$H(k) = AP(k)[I - \gamma^{-2}\overline{Q}P(k) + C^TV^{-1}CP(k)]^{-1}C^TV^{-1}$$

여기서, $H(k)$는 H∞ 필터의 Gain이다.

3. 다중 H∞ 필터에 의한 음성 향상

깨끗한 음성신호 $y(k)$를 각각의 상태에서 L개의 상태와 M개의 혼합 성분을 가진 가우시안 AR(Autoregressive) 모델로 표현할 수 있다. 여기서, y에 대응하는 상태열을 $s = \{s_t, t = 1,2,..,T\}$, $s_t \in \{1,2,..,L\}$라 두고, (s,y)에 대응하는 혼합열을 $h = \{h_t, t = 1,2..,T\}\ h_t \in \{1,2,..,M\}$라고 하면 깨끗한 음성모델은 아래와 같이 표현할 수 있다.

$$y(k) = B_{h|s_t}^T Y(k-1) + e_{h|s_t}(k),\ (t-1)N+1 < k < tN \quad (14)$$

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

식(14)는 아래식과 같이 상태방정식으로 표현된다.

$$Y(k) = F(s_t,h_t)Y(k-1) + Be_{s_t|h_t}(k) \quad (15)$$

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

$$\hat{Y}(k) = E\{Y(k)|s(k)\} = \int_{-\infty}^{\infty} Y(k)p(Y(k)|s(k))dY(k) \quad (16)$$

위 (16)식의 조건 분포함수를 아래식과 같이 쓸 수 있다.

$$p(Y(k)|s(k)) = \sum_{j=1}^{L}\sum_{m=1}^{M} p(Y(k)|s_t = j, h_t = m, s(k))p(s_t = j, h_t = m|s(k)) \quad (17)$$

(17)을 (16)식에 대입하고 적분과 합계를 바꾸면, $\hat{Y}(k)$추정은 아래식으로 표현된다

$$\hat{Y}(k) = \sum_{j=1}^{L}\sum_{m=1}^{M} \hat{Y}_{h|s}(k)p(s_t = j, h_t = m|s(k)) \quad (18)$$

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

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

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

$$\lambda=\{A,B_{h_t|s_t},W_{h_t|s_t},c_{h_t|s_t}\} \tag{19}$$

$A=[a_{ij}]$는 상태천이행렬이고, $B_{h_t|s_t}$는 각 상태에서 혼합에 대한 AR계수, $W_{h_t|s_t}$는 잔치신호의 분산, $c_{h_t|s_t}$는 가중치다.

3.1 $\hat{Y}_{h_t|s_t}(k)$의 추정

$\hat{Y}_{h_t|s_t}(k)$의 추정은 H∞ 필터로 구할 수 있고, H∞ 필터 알고리즘은 다음과 같다.

$$\begin{aligned}\hat{Y}_{h_t|s_t}(k)&=F(s_t,h_t)\hat{Y}_{h_t|s_t}(k-1)\\&\quad+H_{h_t|s_t}(k)\{s(k)-CF(s_t,h_t)\hat{Y}_{h_t|s_t}(k-1)\}\end{aligned} \tag{20}$$

$$\begin{aligned}&H_{h_t|s_t}(k)\\&=F(s_t,h_t)P_{h_t|s_t}(k)[I-\gamma^{-2}\overline{Q}P_{h_t|s_t}(k)+C^TV^{-1}C]^{-1}C^TV^{-1}\end{aligned} \tag{21}$$

$$\begin{aligned}P_{h_t|s_t}(k+1)&=F(s_t,h_t)P_{h_t|s_t}(k)[I-\gamma^{-2}\overline{Q}P_{h_t|s_t}(k)\\&\quad+C^TV^{-1}CP_{h_t|s_t}(k)]^{-1}F(s_t,h_t)^T+BW_{h_t|s_t}(k)B^T\end{aligned} \tag{22}$$

식(20)-(22)으로부터 회귀적으로 $\hat{Y}_{h_t|s_t}(k)$ 을 구한다.

3.2 가중치 $p(s_t=j,h_t=m|s(k))$의 계산

가중치 $P(s_t=j,h_t=m|s(k))$는 베이시안 법칙을 적용하여 아래식과 같이 나타낼 수 있다.

$$\begin{aligned}&p(s_t=j,h_t=m|s(k))\\&=\frac{p(s(k)|s_t=j,h_t=m,s(k-1))p(s_t=j,h_t=m|s(k-1))}{p(s(k)|s(k-1))}\end{aligned} \tag{23}$$

식(23)에서 $p(s(k)|s_t=j,h_t=m,s(k-1))$는 다음과 같이 쓸 수 있다.

$$p(s(k)|s_t=j,h_t=m,s(k-1))=N[\hat{Y}_{h_t|s_t}(k),CP_{h_t|s_t}(k)C^T] \tag{24}$$

위 식에서 $N[\cdot,\cdot]$는 정규분포이다.

식(23)에서 두 번째 요소 $p(s_t=j,h_t=m|s(k-1))$는 마코프 과정으로 나타낼 수 있다.

$$\begin{aligned}&p(s_t=j,h_t=m|s(k-1))=\\&\sum_{i=1}^{N}\sum_{l=1}^{M}\{p(s_t=j,h_t=m|s_{t-1}=i,h_{t-1}=l,s(k-1))p(s_{t-1}=i,h_{t-1}=l,s(k-1))\}\end{aligned} \tag{25}$$

위 식에서 첫 번째 요소는 아래식으로 다시 쓸 수 있다.

$$\begin{aligned}&p(s_t=j,h_t=m|s_{t-1}=i,h_{t-1}=l,s(k-1))\\&=p(h_t=m|s_t=j,s_{t-1}=i,h_{t-1}=l,s(k-1))\\&\quad\times p(s_t=j|s_{t-1}=i,h_{t-1}=l,s(k-1))\end{aligned} \tag{26}$$

h_t와 s_t는 서로 독립이므로 식(26)에서 두 요소는 아래와 같이 다시 쓸 수 있다.

$$p(h_t=m|s_t=j,s_{t-1}=i,h_{t-1}=l,s(k-1))=c_{m|j} \tag{27}$$

$$p(s_t=j|s_{t-1}=i,h_{t-1}=l,s(k-1))=p(s_t=j|s_{t-1}=i)=a_{ij} \tag{28}$$

식(27),(28)을 식(25)에 대입하면 아래식과 같이 나타내어진다.

$$p(s_t=j,h_t=m|s(k-1))=\sum_{i=1}^{N}\sum_{l=1}^{M}a_{ij}c_{m|j}p(s_{t-1}=i,h_{t-1}=l,s(k-1)) \tag{29}$$

식(23)에서 분모는 상태에 독립적이므로, 스케일 인수이므로 $p(s_t=j,h_t=m|s(k-1))$는 전의 가중치 요소를 사용해서 효과적으로 계산할 수 있다.

$$\begin{aligned}&p(s_t=j,h_t=m|s(k-1))\\&=D_tN_{m|j}\sum_{i=1}^{N}\sum_{l=1}^{M}a_{ij}c_{m|j}p(s_{t-1}=j,h_{t-1}=l|s(k-1))\end{aligned} \tag{30}$$

식(30)에서 D_t는 가중치 요소들의 모든 합이 1이 되게하는 스케일 인수이다.

$$\sum_{j=1}^{L}\sum_{m=1}^{M}p(s_t=j,h_t=m|s(k))=1 \tag{31}$$

$\hat{Y}_{h_t|s_t}(k)$과 $p(s_t=j,h_t=m|s(k))$를 계산한 후, $\hat{Y}(k)$ 추정식은 식(32)과 같다.

$$\hat{Y}(k)=\sum_{j=1}^{L}\sum_{m=1}^{M}\hat{Y}_{h_t|s_t}(k)p(s_t=j,h_t=m|s(k)) \tag{32}$$

추정된 상태열 $\hat{Y}(k)$에서 추정신호는 아래와 같이 구할 수 있다.

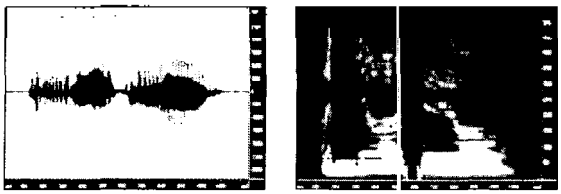
$$\hat{y}(k)=L\hat{Y}(k)\qquad L\in\mathbb{R}^{1\times n} \tag{33}$$

4. 실험 및 결과

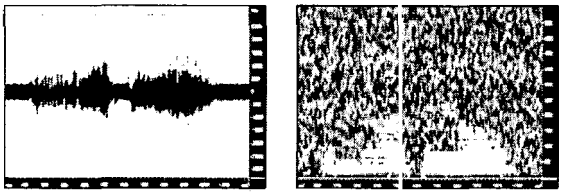
제안된 방법은 SNR이 각각 0dB, 5dB, 10dB에서 백색 잡음이 부가되었을 때 음성향상 결과를 보여주고 있다.

학습은 "안녕하세요"로 발성된 6개의 문장을 사용하였고, 테스트는 동일 문장으로 학습에 사용하지 않은 문장을 사용하였다. 이 실험에서 샘플링 주파수는 11,025Hz이고, 깨끗한 음성의 AR 모델은 15차이다. HMM에서 7상태, 2혼합을 사용하였고, Pentium IV 2.4GHz PC에서 실험하였다.

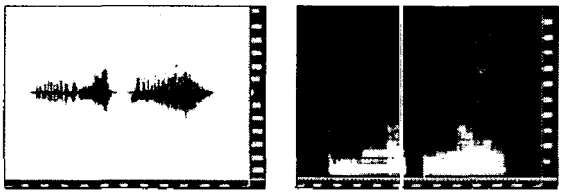
그림1은 SNR 5dB 환경에서 음질향상 결과를 보여주고 있다. (a), (b)는 각각 깨끗한 음성과 잡음이 부가된 음성이고, (c), (d)는 기존의 방법과 제안된 방법으로 음질을 향상 시킨 결과이다. 표1은 각 dB별로 음질이 향상된 결과를 비교하고 있다. 전체적으로 기존의 방법에 비해 계산량은 약간 증가하나, SNR은 0.5 ~ 2.5dB 향상되었다. 기존의 방법은 $\hat{Y}_{h_t|s_t}(k)$을 추정하는데 칼만 필터를 사용하였다.



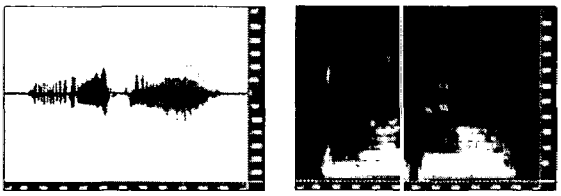
(a) Clean Speech



(b) 5dB Noisy Speech



(c) Enhanced Speech(conventional method)



(d) Enhanced Speech(proposed method)

그림1. 음질향상 결과 파형

	conventional method		proposed method	
Input SNR[dB]	Output SNR[dB]	시간 (초)	Output SNR[dB]	시간 (초)
0	7.05	6.0	7.52	8.8
5	8.84		10.44	
10	10.80		13.12	

표1. 기존 방법과 제안된 방법의 음질향상 결과 비교

5. 결론

본 논문에서는 음성의 통계적 특성을 이용하여 모델 파라미터를 추정하는 은닉 마코프 모델과 잡음의 변화에 강인한 H_∞ 알고리즘을 사용해서, 다중 H_∞ 필터에 의한 강인한 음성향상 방법을 제안하였다. 기존 방법보다 계산량은 약간 증가하나 SNR성능이 좋아짐을 알 수 있다.

참고문헌

1. J.S.Lim and A.V.Oppenheim, All-pole modeling of degraded speech, IEEE Trans. Acoust. Speech Signal Processing, vol.ASSP-26, pp.197-210, 1978

2. Y.Ehpraim, A Bayesain approach for speech enhancement using hidden Markov models, IEEE Trans. Singnal Processing, vol.41, pp.725-735, 1992

3. U.Shaked and Y.Theodor, H_∞-optimal estimation: A tutorial, in Proc. 31st IEEE CDC, pp.2278-2286, 1992

4. X.Shen and L.Deng, A Dynamic System Approach to Speech Enhancement Using the H_∞ Filtering Algorithm, IEEE Trans. Speech and Audio Processing, vol.7, no.4, pp.391-399, 1999

5. K.Y.Lee and J.Y.Rheem, Smoothing approach using forward - backward Kalman filter with Markov Switching Parameters for Speech Enhancement, IEEE Trans. Signal Processing, vol.80, pp.2549-2588, 2000

6. K.Y.Lee and Katsuhiko Shirai, Efficient Recursive Estimation for Speech Enhancement in Colored Noise, IEEE Signal Processing Letters, vol.3, no.7, pp.196-199, 1996