

잡음 백색화와 Kalman 필터를 이용한 잡음제거

정상배, 한민수

한국정보통신대학원대학교 공학부

Noise reduction by whitening of colored noise and Kalman filter

Sang-Bae Jeong, Minsoo Hahn

School of Engineering, Information and Communications Univ.

sangbae@icu.ac.kr, mshahn@icu.ac.kr

요약

음성신호에 섞인 잡음을 처리하기 위해서 단일 마이크로폰을 이용한 방법이 많이 연구되고 있는데, 그 중에서 Kalman 필터를 이용한 방법은 먼저 음성신호의 모델을 검출하고 잡음이 섞인 신호에서 표준 Kalman 필터를 이용해서 음성신호 성분만을 검출하게 된다. 본 논문에서는 음성신호에 섞인 유색잡음을 백색화하는 방법을 적용하여 Kalman 필터의 잡음제거 성능을 향상시키는 방법을 제안하였다.

1. 서론

신호에 잡음이 섞인 상태에서 신호를 검출하는 방법은 2개의 마이크로폰을 이용하는 것이 일반적이다[1]. 구체적으로, 잡음원 근처에 또 다른 하나의 마이크로폰을 설치하여 실제 신호

에 섞인 잡음과 상관도가 큰 잡음만을 수신하여 적응필터를 동작시킴으로써 가능하다. 그러나 실제적인 상황에서, 음성신호에 잡음이 섞여 있을 경우, 잡음의 위치를 정확히 파악하여 마이크로폰을 설치한다는 것은 불가능한 경우가 많다. 따라서 약간의 성능 저하는 불가피하지만 보다 일반적으로 적용할 수 있는 단일 마이크로폰을 이용한 잡음처리 방법들이 연구되고 있다.

그러한 여러가지 방법들 중에 Kalman 필터를 이용한 방법은 먼저 음성신호에 대한 모델 파라미터들을 찾은 후에 표준 Kalman 필터를 이용해서 음성신호만을 검출하는 과정을 거치게 된다[2].

Kalman 필터는 상태변수와 모델 파라미터에 의해 원래의 신호가 발생되고, 수신된 신호는 원래의 신호와 백색 가산성 잡음이 섞인다고 가정할 때, 섞인 백색 잡음을 효과적으로 억제할 수 있는 필터이다. 일반적으로 음성신호의 경우

전극(all-pole)형태로 모델링이 가능하기 때문에 상태변수와 모델 파라미터에 의한 신호생성의 가정이 가능하다. 그리고 음성신호에 섞이게 되는 잡음은 가산성으로 생각할 수 있으므로 Kalman 필터를 이용해서 잡음을 제거할 수 있다.

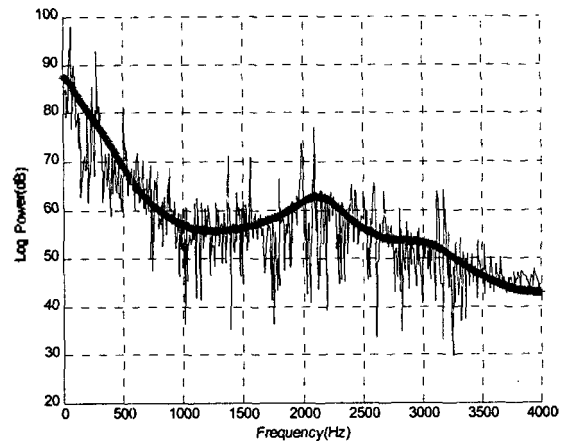
Kalman 필터의 성능을 향상시키기 위해 몇 가지 방법들이 제안되고 있으나, 실제 음성신호에 섞이는 유색잡음에 의한 필터의 성능저하에 대한 연구는 미흡하다고 하겠다[3]. 따라서 본 논문에서는 유색잡음의 백색화를 위해서 선형예측 후의 잔차신호가 백색잡음과 유사하다는 사실을 이용한다. 유색잡음에 대한 선형예측 파라미터는 수신되는 신호의 초기 구간 몇 프레임은 즉, 수십 msec 동안은 순수 잡음신호라고 가정 한 후에 추정하도록 한다. 일단 선형예측을 통해서 잡음의 백색화가 이루어지면, 표준 Kalman 필터를 동작시키게 된다.

2. 본론

2.1 사무실 환경에서의 잡음 특성

Kalman 필터를 사용하기 위해서는 수신된 신호에 섞이는 잡음을 백색으로 가정해야 하지만 대부분의 음성에 섞이는 가산성 잡음은 백색으로 가정하기 어렵기 때문에 필터의 성능을 최대한으로 발휘하기 어려운 단점이 있다. 예를 들어 사무실 환경의 잡음은 저주파 성분이 강조된 형태의 저대역 통과 필터의 주파수 응답과 유사하다. 그림 1에 실제 측정된 사무실 환경 잡음의 스펙트럼을 예로 보았다. 이 경우 저주파와 고주파 대역의 전력차가 30dB 이상임을 알 수 있다.

2.2 표준 Kalman 필터 및 제안한 알고리즘 음성신호의 모델변수(AR 파라미터) 및 가



[그림 1: 사무실 환경 잡음]

산성 잡음에 대한 변수를 모두 알고 있다고 가정할 때, 표준 Kalman 필터의 알고리즘은 다음과 같다[4].

음성생성 모델(AR Model):

$$x(n) = \sum_{k=1}^p a(k)x(n-k) + w(n)$$

수신되는 신호($v(n)$): 백색 가산성 잡음):

$$y(n) = x(n) + v(n)$$

State Equation :

$$\mathbf{x}(n) = \mathbf{A}\mathbf{x}(n-1) + \mathbf{w}(n)$$

Observation Equation :

$$y(n) = \mathbf{C}^T \mathbf{x}(n) + v(n)$$

Initialization : $\bar{\mathbf{x}}(0|0) = E[\mathbf{x}(0)]$

$$\mathbf{P}(0|0) = E[\mathbf{x}(0)\mathbf{x}^H(0)]$$

Computation : For $n=1,2,3,\dots$

$$\bar{\mathbf{x}}(n|n-1) = \mathbf{A}\bar{\mathbf{x}}(n-1|n-1)$$

$$\mathbf{P}(n|n-1) = \mathbf{A}\mathbf{P}(n-1|n-1)\mathbf{A}^H + \mathbf{Q}_w$$

$$\mathbf{K}(n) = \mathbf{P}(n|n-1)\mathbf{C} / [\mathbf{C}^T \mathbf{P}(n|n-1)\mathbf{C} + \mathbf{Q}_v]$$

$$\bar{\mathbf{x}}(n|n) = \bar{\mathbf{x}}(n|n-1) + \mathbf{K}(n)[y(n) - \mathbf{C}^T \bar{\mathbf{x}}(n|n-1)]$$

$$\mathbf{P}(n|n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{C}^T] \mathbf{P}(n|n-1)$$

$$\mathbf{x}(n) = [x(n) \ x(n-1) \ \Lambda \ x(n-p+1)]^T$$

$$\mathbf{w} = [w(n) \ 0 \ 0 \ \Lambda \ 0]^T$$

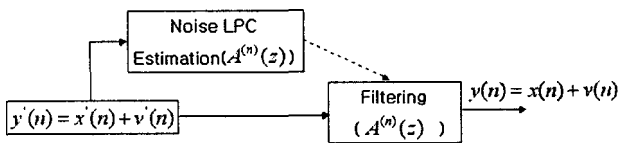
$$\mathbf{C}^T = [1 \ 0 \ 0 \ \Lambda \ 0]^T$$

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

$$Q_v = E[|v(n)|^2], \quad Q_w = E[w(n)w^T(n)]$$

\bar{x} : x 의 추정치

이때, 일반적인 상황에서는 수신된 신호에 백색이 아닌 유색잡음이 섞인다고 가정해야 하므로 수신된 신호의 초기 프레임들(순수 잡음구간으로 가정)에서 잡음의 LPC 계수를 구한 후 신호의 모든 구간을 필터링해서 유색잡음을 백색화시킨다. 그 과정을 그림 2에 보였다.

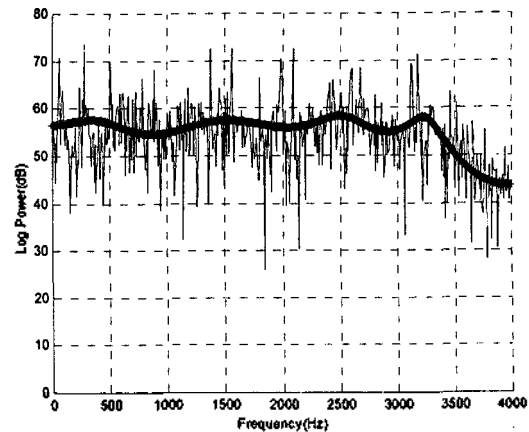


[그림 2: 유색잡음의 백색화]

$y'(n)$: 유색잡음($v(n)$)이 섞인 최초 수신신호
 $y(n)$: 유색잡음이 백색화($v(n)$)된 수신신호

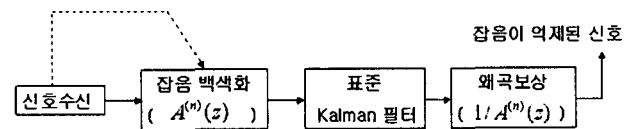
그림 3은 그림 1의 유색잡음을 5차의 AR 계수를 사용하여 모델링한 잔차신호에 대한 스펙트럼을 보인 것이다. 완벽하지는 못하더라도 백색화된 잡음에 가까운 스펙트럼을 가짐을 확인할 수 있다.

유색잡음의 백색화가 끝나면, 앞서 언급한 표준 Kalman 필터를 통과시켜서 잡음이 섞이지 않은 신호($x(n)$)의 추정치인 $\bar{x}(n)$ 을 구하도록 한다. 여기서, 추정된 신호 $\bar{x}(n)$ 는 초기의 잡음 백색화 과정에 의해서, 잡음은 억제되었을지라도, 왜곡이 존재하는 신호이다. 그렇지만 신호가 $A^{(n)}(z)$ 로 왜곡되었음을 알고 있으므로



[그림 3: 백색화된 유색잡음]

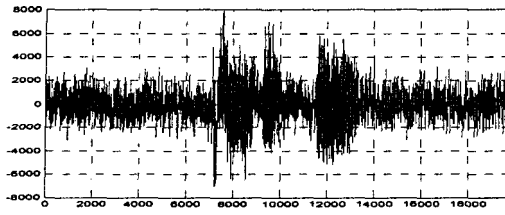
$1/A^{(n)}(z)$ 를 사용하여 필터링해 줌으로써 보상해 주도록 한다. 그림 4는 전체적인 알고리즘의 구성도이다.



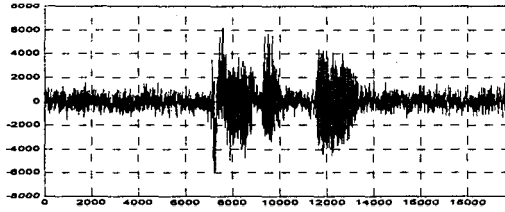
[그림 4: 전체적인 알고리즘 구성도]

3. 실험 및 결과

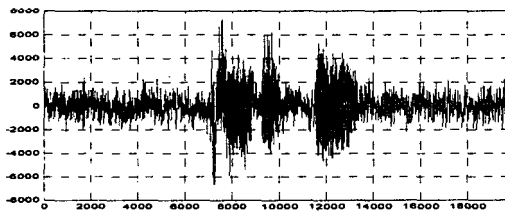
그림 1의 사무실환경 잡음과 유사한 스펙트럼을 갖는 가산성 인공잡음을 생성하고, 조용한 곳에서 녹음한 음성신호에 대해서 SNR이 5dB가 되도록 잡음을 섞은 후에 제시한 방법 및 잡음백색화를 하지 않았을 경우에 대해서 실험을 행하였다. 사용한 신호는 모두 8kHz, 16bit로 표본화하였다. Kalman 필터의 수행에 필요한 음성신호에 대한 AR 파라미터는 미리 알고 있다고 가정할 수 없으므로 잡음백색화가 끝난 후에 나온 신호를 사용하여 추정하였다. 사용한 차수는 10차이다. 잡음백색화는 음성신호의 초기 10프레임에 대해 6차의 모델링을 통해서 필요한 파라미터를 추정하였으며 프레임의 크기는 16msec이다. 실험결과는 그림 5, 6, 7과 같다.



[그림 5: 유색잡음이 섞인 신호 /불국사/]



[그림 6: 제안된 방법에 의한 잡음제거]



[그림 7:잡음백색화를 하지 않았을 때]

제안된 방법에 의한 SNR 은 10.8dB 였으며, 잡음 백색화를 하지 않았을 때 보다 약 2dB 의 SNR 향상이 관측되었다.

잡음제거에 의한 SNR 의 향상은 기본적으로 음성인식기에 사용되는 음성검출기(speech detector)의 성능을 높일 것으로 생각할 수 있는데, 실제 인식기에서 인식 파라미터로 사용되는 것은 스펙트럼 정보이다. 따라서 본 논문에서는 단위 프레임당 캡스트럼(Cepstrum) 거리를 구하여 스펙트럼 왜곡의 향상 정도를 제시한다. 단위 프레임당 캡스트럼 거리는 아래의 식과 같이 정의된다.

$$CD(i) = \sum_{\tau=1}^M (c_{\tau}^{ref} - c_{\tau}^{ts'})^2$$

$$CD^{mean} = \frac{1}{L} \sum_{i=0}^{L-1} CD(i)$$

$CD(i)$: 프레임 i 에서 캡스트럼 거리

10 차의 캡스트럼을 각각 구하여 실험한 결과 제안된 방법을 사용한 것은 잡음백색화를 사용하지 않은 것 보다 0.0011 정도의 거리 향상이 측정되었다.

4. 결론

잡음백색화 과정을 사용하였을 때, Kalman 필터를 이용한 잡음제거시 SNR 의 향상 및 스펙트럼 왜곡의 감소가 관측되었다. 이는 잡음이 섞인 음성의 음질 개선 뿐만 아니라 잡음환경에서의 음성인식기의 인식률을 개선하는데 기여할 것으로 예상된다.

본연구는 ㈜한국 엑시스에서 지원한 연구비로 수행되었습니다.

5. 참고문헌

- [1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, 1985.
- [2] K. K. Pailwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. ICASSP'87*, pp.177-180
- [3] M. Gabrea, E. Grivel and M. Najim, "A Single Microphone Kalman Filter-Based Noise Canceller," *IEEE Signal Proc. Lett.*, Vol. 6, No. 3, March 1999.
- [4] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, 1996.