칼만필터를 이용한 음성신호에 중첩된 유색잡음의 감쇠

An Application of the Kalman Filter for Attenuation of Colored Noise Superimposed on Speech Signal

구 본 웅*
(Bon Eung Gu*)

이 論文은 1992年度 敎育部 學術硏究助成費에 의하여 硏究되었음.

요 약

정체형 칼만필터와 간단한 음성·비음성 판별알고리즘을 사용하여 비정체형 유색잡음을 감쇠시키는 방법을 제안하였다. 종래의 잡음감쇠알고리즘들이 대부분 백색 또는 정체형 잡음을 다룬데 비하여 본 연구는 대부분의 실제 잡음환경, 즉, 비백색 비정체성 잡음을 다루었다는 점이 다르다. 잡음감쇠기로서는 AR모델에 의거한 벡터형 칼만필터를 사용하였고, 음성/비음성 판별에는 단구간에너지의 임계값논리를 사용하였다. 칼만필터에 필요한 잡음의 계수는 비음성구간에서 추산하였고, 음성의 계수는 EM반복법을 적용하여 추산하였다. 실험결과는 신호대 잡음비와 청취테스트로 제시하였다. 차량잡음을 사용한 실험결과, 비음성구간의 배경잡음은 거의 완전히 제거할 수 있었고, SNR이 0dB 내지 -5dB로 낮아짐에 따라 왜곡이 심화되는 경향을 보였으나, 음성의 명료도를 저하시키지는 않았다.

ABSTRACT

A speech enhancement algorithm which attenuates nonstationary colored noise is presented in this paper. The algorithm consists of a stationary Kalman filter and the simple speech/nonspeech detector. While the conventional enhancement systems are focused on a stationary and/or white background noise, this study is focused on the more realistic nonstationary and nonwhite noise. An AR model-based vector Kalman filter is used as a noise suppression system and a short-time energy threshold logic is used as a speech/nonspeech classifier. For Kalman filtering, noise coefficients are estimated in the nonspeech frame, and speech coefficients are estimated by applying the EM iteration algorithm. Simulation results using the car noise are presented based on the signal-to-noise ratio and informal listening tests. According to the experimental results, background noises in the nonspeech frames are eliminated almost completely, while some distortions are noticed in the speech frames. The distortion becomes severer as the SNR is reduced to 0dB and -5dB. Intelligibility, however, is not degraded significantly.

I. 서 론

음성통신이 배경잡음의 영향으로 인하여 그 성능

*경기대학교 전자공학과
접수일자 : 1993년 12월 21일

과 서비스에 지장을 받는 경우가 많이있다. 정도가 약한 사무실의 배경잡음에서부터 강한 엔진잡음을 포함하고 있는 자동차나 헬기내에서의 잡음에 이르기까지 잡음의 종류는 매우 다양하다. 일반적으로 부가잡음은 음성의 명료도를 저하시키고, 청각을 피로하게 한다. 부가잡음은 특히 고압축율의 음성코딩시스템과 인식시스템의 성능을 현저히 저하시키는 것으로 알려져 있다[1]-[3]. 이러한 잡음의 영향을 극소화시키고 음성품질을 개선하는 것은 많은 음성처리시스템의 실용화를 위해서는 필수적인 요소이다.

여러가지의 다양한 방법의 잡음감쇠시스템들이 발표된 바 있다[3]-[6]. 각 시스템의 구조와 성능은 용도, 대상 잡음의 종류, 잡음에 관하여 필요로 하는 정보, 또, 그것을 구하는 방법 등에 따라 달라진다. 특히, 사용되는 마이크의 수, 잡음의 백색여부는 적용방법선택의 중요한 기준이다. 종래에 많이 연구되어온 대표적인 잡음감쇠방법으로는 적응잡음제거[7]-[9], 캡스트럼차감방법[10], 스펙트럼차감방법[11], [12], 위너필터링방법[13], 칼만필터링방법[14]등이 있다. 그외에 다른 방법들도 있으나[15][16], 모든 각각의 방법은 나름대로 잡음감쇠효과가 충분치못하다거나 음질의 왜곡 등 단점을 갖고있다[6].

이러한 잡음감쇠시스템들은 모두 음성에 부가된 잡음이 WSS(Wide-Sense Stationary)이라는 가정하에서 개발되었다. 이 경우에는 처음의 일정시간동안은 잡음만 있다고 가정하고 필요한 통계적 계수들을 추정하여 사용할 수 있다. 그러나, 실제상황에서는 많은 경우에 잡음이 비정체형이므로, 잡음감쇠시스템을 실제상황에 응용하려면 잡음의 계수들을 수시로 다시 구해야 한다. 적응잡음제거를 제외한 다른 방법에 관한 대부분의 연구들은 잡음에 관한 정보를 미리 안다고 가정하는데, 잡음감쇠시스템을 실제상황에 적용하려면 잡음에 관한 정보를 음성이 없는 구간에서 추출하여 사용해야한다. 따라서, 부가잡음이 비정체성일 경우에, 음성과 잡음이 더해진 신호로부터 음성이 있는 구간과 잡음만 있는 구간을 판별하는 장치가 필요한데, 이 역할을 하는 것이 음성검출기이다.

잡음감쇠시스템의 성능은 음성검출기의 잡음에 대한 강인성과 정확도에 따라 좌우된다. 이전에 개발된 음성검출기들은 대부분 전화회선에서의 TASI시스템[17], [18] 및 음성인식에서의 끝점검출[19]-[21] 및 패턴인식기법을 이용한 유성-무성-침묵 구간판별[22]-[24] 또는 피치검출과 함께 LPC시스템의 일부로 사용하기위한 것들[2], [19]이므로 잡음에는 취약하다. 최근에와서 이동통신과 음성인식의 실용화에 따라 잡음에 강인한 음성검출기에 관한 연구도 발표되고 있으나[26]-[28], 비정체형잡음의 경우는 아직 미해결상태이다.

부가잡음은 그 종류가 매우 다양하여 모든 잡음을 일반적으로 다루리가 어려우므로, 본 연구에서는 차량잡음만을 다루었다. 차량잡음만을 이룬 연구논문도 있으나 그 내용은 위에서 기술한 방법의 범주내에 있다. 예를 들면, 적응잡음제거방법[29], 스펙트럼차감방법[30], 여러개의 부대역을 사용한 방법[31]등이 있다. 차량잡음은 엔진소음, 바람소리, 타이어의 마찰음, 통과차량의 소음 등이 복합된 것이므로 비백색이고, 또, 시간에따른 변화율이 일정치않은 비정체형잡음이므로 그 특성을 규정하기가 어렵다. 다만, 대부분의 에너지가 600Hz 이하의 저주파수 대역에 밀집되어있고, 또 음성신호에 비해서는 시간적인 변화속도가 느리다는 두가지 특성은 이전의 연구[29] 및 본 연구에서의 예비실험결과에 의해서도 분명하다고 할 수 있다.

본 연구에서는 칼만필터와 음성검출기를 사용하여 비정체형 유색잡음을 감쇠하는 방법을 제안하였다. 칼만필터는 유색잡음감쇠를 위한 백터형 칼만필터를 사용하였고, 단구간 에너지를 이용한 음성검출기를 고안하여 칼만필터의 잡음계수 추정에 사용하였다. 칼만필터에 요구되는 음성신호의 계수는 EM알고리즘을 적용하여 반복적 방법으로 구하는데, 음성신호와 잡음은 모두 짧은 시간동안은 정체형이라고 가정하여 프레임 단위로 처리하였다. 이러한 비정체성 잡음환경하에서의 잡음감쇠성능에 관한 연구는 종래의 연구들이 정체성잡음을 가정했다는 점에서 잡음감쇠시스템의 실용화에 더욱 근접한 시도라고 할 수 있다.

유색잡음용 칼만필터알고리즘은 II절에 제시하였고, III절에서는 에너지 임계값을 사용하는 음성검출알고리즘을 제시하였다. IV절에서는 실제음성과 차량잡음을 사용한 모의실험결과를 제시하였고, 제안된 시스템의 성능은 수동식 음성검출기를 사용했을 때의 성능과 비교하였다.

Ⅱ. 유색잡음 감쇄를 위한 칼만필터 알고리즘

순수한 음성신호 $x(i)$를 AR(p)모델로 표시하면,

$$x(i)=\sum_{j=1}^{p} a_j x(i-j)+w(i) \quad (1$$

이고, 여기서 $w(i)$는 평균이 0, 표준편차가 σ_w인 가우시안 i.i.d.라고 가정한다. 또, $x(i)$에 부가잡음 $v(i)$가 더해진 신호를 $s(i)$라 하고, $v(i)$는 $x(i)$와 무관하다고 가정한다. 잡음이 더해진 신호 $\{s(i)\}$와 잡음 $\{v(i)\}$의 자기상관계수가 주어졌을때 원신호 $\{x(i)\}$를 추정하는 것이 주어진 문제이다.

n번째 시간에서 N개의 연속적인 샘플로 이루어진 벡터를 각각 $\underline{x}(n)$, $\underline{s}(n)$, $\underline{v}(n)$이라하면, MMSE 추정벡터는

$$\hat{\underline{x}}(n)=E[\underline{x}(n)\mid \underline{s}(n), \underline{s}(n-1),\cdots] \quad (2)$$

이고, 이것은 벡타칼만필터로 계산할 수 있다.

신호모델을 벡타식으로 다시 쓰면,

$$\underline{x}(n)=\mathrm{A}_x \underline{x}(n-1)+\mathrm{B}_x \underline{w}(n) \quad (3)$$

$$\underline{s}(n)=\underline{x}(n)+\underline{v}(n) \quad (4)$$

이고, $\underline{w}(n)$은 평균이 0, covariance 행렬이 $Q_w=\sigma_w^2 I_N$인 백색 가우시안이다.

A_x, B_x는 $\{x(i)\}$의 선형예측계수로부터 계산된 $N\times N$ 행렬이다.

같은 방법으로 유색잡음의 AR모델을 벡타식으로 쓰면,

$$\underline{v}(n)=\mathrm{A}_v \underline{v}(n-1)+\mathrm{B}_v \underline{\nu}(n) \quad (5)$$

이고, $\underline{\nu}(n)$은 평균이 0, covariance 행렬이 $Q_\nu^2 I_N$인 백색 가우시안, A_v, B_v는 $\{v(i)\}$의 선형예측계수로부터 계산된 $N\times N$ 행렬이다.

식(3)과 식(5)의 상태벡터를 합성하여 $\bar{x}(n)=[\underline{x}^T(n)\ \ \underline{v}^T(n)]^T$, 구동잡음을 합성하여 $\bar{w}(n)=[\underline{w}(n)^T\ \underline{\nu}(n)^T]^T$라 하고 식(3)·(5)를 다시 쓰면,

$$\bar{x}(n)=\bar{A}\ \bar{x}(n-1)+\bar{B}\ \bar{w}(n) \quad (6)$$

$$\underline{s}(n)=\bar{C}\ \bar{x}(n) \quad (7)$$

이고, $\bar{A}=\begin{bmatrix} A_x & 0 \\ 0 & A_v \end{bmatrix}$, $\bar{B}=\begin{bmatrix} B_x & 0 \\ 0 & B_v \end{bmatrix}$ $\bar{C}=[I_N\ I_N]$ 인데, 이것은 측정이 완전한 경우에 백색 가우시안 벡타 $\bar{w}(n)$로 구동되는 선형시스템이다. $\bar{w}(n)$의 covariance행렬은

$$Q=E[\bar{w}(n)\cdot\bar{w}(n)^T]=\begin{bmatrix} Q_w & 0 \\ 0 & Q_v \end{bmatrix}=\begin{bmatrix} \sigma_w^2 I_N & 0 \\ 0 & \sigma_v^2 I_N \end{bmatrix} \quad (8)$$

이다.

칼만필터의 error covariance행렬의 singularity를 제거하기 위하여 축변환을 해야한다. 변환행렬을 $T=\begin{bmatrix} I_N & I_N \\ I_N & 0_N \end{bmatrix}$로 하면, $\tilde{x}(n)=T\bar{x}(n)=\begin{bmatrix} \underline{s}(n) \\ \underline{x}(n) \end{bmatrix}$ 이 되고, 식(7)에 T를 곱하면

$$\tilde{x}(n)=\tilde{A}\ \tilde{x}(n-1)+\tilde{B}\ \bar{w}(n) \quad (9)$$

$$\underline{s}(n)=\tilde{C}\ \tilde{x}(n) \quad (10)$$

이고, $\tilde{A}=T\bar{A}T^{-1}$, $\tilde{B}=T\bar{B}$, $\tilde{C}=\bar{C}T^{-1}=[I_N\ 0_N]$ 이다.

확장상태벡타 $\tilde{x}(n)$의 추정벡타 $\check{x}(n)$은

$$\check{x}(n)=\check{x}(n|n-1)+K(n)\{\underline{s}(n)-\tilde{C}\check{x}(n|n-1)\} \quad (11)$$

로 계산하고, 여기서 $\check{x}(n|n-1)=\tilde{A}\check{x}(n-1)$는 확장예측벡타이다. 이득벡타 및 error covariance행렬을 구하는 식은

$$K(n)=P(n|n-1)\ \tilde{C}^T[\tilde{C}P(n|n-1)\ \tilde{C}^T]^{-1} \quad (12)$$

$$P(n|n-1)=\tilde{A}P(n-1)\ \tilde{A}^T+\tilde{B}Q\tilde{B}^T \quad (13)$$

$$P(n)=[I-K(n)\tilde{C}]P[n|n-1) \quad (14)$$

이다. 추정벡타와 이득벡타의 크기는 각각 $2N$, $2N\times N$인데, 이것은 다음과 같이 각각 N, $N\times N$으로 축소할 수 있다.

$2N\times 2N$ 행렬 $P(n|n-1)$을 다음과 같이 4개의 $N\times N$ 부행렬로 표시하면

$$P(n|n-1)=\begin{bmatrix} P_{11}(n|n-1) & P_{12}(n|n-1) \\ P_{21}(n|n-1) & P_{22}(n|n-1) \end{bmatrix}$$

이므로, 식(12)는 다음과 같이 된다.

$$K(n)=\begin{bmatrix} P_{11}(n|n-1) \\ P_{21}(n|n-1) \end{bmatrix} P^{-1}{}_{11}(n|n-1) \qquad (15)$$

$$=\begin{bmatrix} I_N \\ P_{21}(n|n-1)\ P^{-1}{}_{11}(n|n-1) \end{bmatrix}=\begin{bmatrix} I_N \\ K_2(n) \end{bmatrix}$$

또, 확장예측벡타를 풀어서 쓰면

$$\check{x}(n|n-1)=\hat{A}\ \check{x}(n-1)$$

$$=\begin{bmatrix} A_s \underline{\hat{s}}(n-1)+A_v\{\underline{\hat{s}}(n-1)-\underline{\hat{x}}(n-1)\} \\ A_v \underline{\hat{x}}(n-1) \end{bmatrix}$$

$$\equiv\begin{bmatrix} \underline{\hat{s}}(n|n-1) \\ \underline{\hat{x}}(n|n-1) \end{bmatrix} \qquad (16)$$

이므로, 확장상태벡타의 추정벡타는

$$\check{x}(n)\equiv\begin{bmatrix} \underline{\hat{s}}(n) \\ \underline{\hat{x}}(n) \end{bmatrix}$$

$$=\begin{bmatrix} \underline{s}(n) \\ \underline{\hat{x}}(n|n-1)+K_2(n)\{\underline{s}(n)-\underline{s}(n|n-1)\} \end{bmatrix}$$

가 되고, 따라서 $\underline{\hat{s}}(n)=s(n)$이고, 추정벡타는

$$\underline{\hat{x}}(n)=\underline{\hat{x}}(n|n-1)+K_2(n)\{\underline{s}(n)-\underline{\hat{s}}(n|n-1)\} \qquad (17)$$

이다. 여기서, $\underline{\hat{x}}(n|n-1)$, $\underline{\hat{s}}(n|n-1)$는 식(16)에 있고, 식(15)로부터

$$K_2(n)=P_{21}(n|n-1)P^{-1}{}_{11}(n|n-1) \qquad (18)$$

이다. 식(13), 식(14)에서 $P(n|n-1)$의 모든 부행렬이 $P(n)$의 계산에 사용되므로 이들 행렬의 찻수는 축소할 수 없다.

요약하면, 유색잡음감쇠를 위한 축소된 찻수의 칼만필터알고리즘은 초기조건을 $\underline{\hat{x}}(0)=0$, $P(0)=0$으로 하여 식(13), 식(18), 식(17), 식(14)의 순서로 반복한다.

Ⅲ. 음성-비음성 판별

주어진 신호 $\{s(i)\}$로부터 음성신호가 있는 구간과 없는 구간을 구분하는 것이 음성검출알고리즘인데, 이것은 임계값을 사용하는 방법과 패턴인식적인 기법을 사용하는 방법으로 크게 분류할 수 있다. 본 연구에서 다루는 음성검출의 목적은 잡음감쇠에 필요한 잡음의 계수를 얻는데 있고 계산량 및 지연시간이 짧을 것이 요구되므로 본 연구에서는 비교적 간단한 임계값 방법을 사용하였다.

이 경우에 사용되는 특징파라메타로에는 여러가지가 있으나 가장 흔히 사용되는 것은 단구간에너지와 영교차율이다. 이것들은 배경잡음의 에너지가 음성의 것보다 상대적으로 적은 경우에 효과적인데, 단구간에너지는 에너지가 큰 유성구간 검출에, 영교차율은 에너지가 작은 무성구간 검출에 각각 사용되어왔다. 그러나, 배경잡음의 에너지가 크고(5dB 이하) 자동차잡음과 같이 에너지가 저주파대역에 집중되어 있는 경우에는 무성음구간과 같은 작은 에너지의 신호는 배경잡음에 묻혀버리고, 또, 비정체형 유색잡음의 경우에는 영교차율의 변화패턴이 음성구간의 것과 구분하기가 어려워서, 영교차율은 판별기능을 상실하게 된다. 예비실험결과에서도 영교차율은 음성-비음성 구간판별에 효과가 별로 없는 것으로 판명되었고, 이점은 다른 논문에서도 지적된 바 있다[20], [23]. 단구간에너지를 사용할 경우의 단점은 배경잡음의 에너지가 어느 정도(0dB)이상 커지면 일부 유성구간마저 잡음구간과 구분하기가 어려워진다는 것이다. 그러나, 음성구간의 에너지가 배경잡음의 것보다 상대적으로 클 경우에 단구간에너지는 유력한 특징파라메타이다.

프레임의 길이를 L이라 하면, n번째 프레임의 단구간에너지 $STE(n)$은

$$STE(n)=\frac{1}{L}\sum_{i=1}^{L} s^2(i)$$

인데, 계산상의 편의를 위하여 절대값에너지, 즉,

$$STE(n)=\frac{1}{L}\sum_{i=1}^{L} |s(i)|$$

를 사용할 수도 있다.

이외에도 음성인식을 위한 끝점검출이나 패턴인식 기법에서 사용되는 판별파라메타에는 여러가지가 있는데, 많이 사용되는 것으로는 자기상관계수, 선형예측계수 등이 있다. 이것들은 모두 스펙트럼 추산을 위한 모델계수들로서 음성신호의 스펙트럼과 그 형태가 판이한 백색잡음에는 효과적일 수 있으나, 자동차잡음과 같이 음성신호와 유사한 대역에 많은 에너지를 갖고 있고, 또, 그 형태가 시간에 따라 변하는 비정체성 유색잡음에는 적합치 못한 것으로 알려져 있다[25], [27].

본 연구에서는 단구간에너지와 임계값논리를 사용하였는데, 이것은 알고리즘이 시작된 후 처음의 몇 프레임은 잡음구간이라고 가정하고 에너지임계값을 구한 다음, 이후의 단구간에너지가 이것보다 크면 음성, 작으면 비음성구간으로 판정하는 방법이다. 이 방법은 잡음의 비정체성에 대한 고려가 전혀 되어있지 않으므로 비정체성 잡음에 대한 판별능력은 미약하다. 잡음의 비정체성에 대처하려면 임계값을 적응적으로 재추산해야 하는데, 잡음에너지가 감소할 경우에는 재추산이 용이하지만, 증가할 경우에 재추산하는 방법은 문헌조사결과 아직 없는 것으로 조사되었다. 다음 절에서는 이 방법에 대한 실험결과를 제시하고 실험자가 수동식으로 판별한 결과와 비교하였다.

Ⅳ. 실험결과

실험에는 2초동안 발성된 문장과 자동차 운전석에서 녹음된 잡음을 사용하였다. SNR을 조정하기 위하여 각각 따로 녹음하여 컴퓨터로 합성하였다. 각 신호는 4kHz의 차단 주파수를 갖는 저대역필터를 거친 다음 8kHz로 샘플링하여 14bit로 A/D변환하였다. 프레임의 길이는 30ms(=240개의 샘플)이고, 해닝윈도우를 사용하였고, 15ms씩 중첩하였다.

실험에 사용된 문장은 다음의 두가지이고, 각각 2초동안 발성되었다.

문장1 ; "The pipe began to rust while new."(여성)
문장2 ; "하나, 둘, 셋"(남성)

그림1의 (a)와 (b)는 각각 실험에 사용된 문장1과 문장2의 파형이다. 문장1은 음성구간과 비음성구간이 비교적 자주 바뀌고 비음성구간이 연속되는 길이가 짧게 여러 곳에 분산 되어 있어서 정교한 판별을 요하는 신호라고 할 수 있다. 이와는 대조적으로 문장2는 단순한 모양을 갖고 있고 비음성구간의 길이도 비교적 같다. 그림2의 (a)와 (b)는 각각 문장1과 문장2의 단구간에너지가 변화하는 모습이다. 에너지의 변화폭이 커서 무성음구간과 같은 에너지가 적은 부분은 그림에 나타나지 않았다. 따라서, 이러한 구간은 잡음구간과의 구분이 어려울 것임을 예상할 수 있다. 음성구간은 최소 다섯 프레임, 즉, 75ms 이상 지속됨을 알 수 있다.

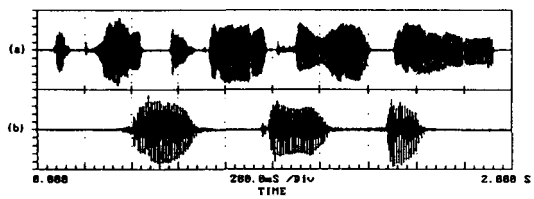

그림. 1. 실험에 사용된 음성신호의 파형
(a) 문장1("The pipe began to rust while new." : 여성)
(b) 문장2("하나, 둘, 셋" : 남성)

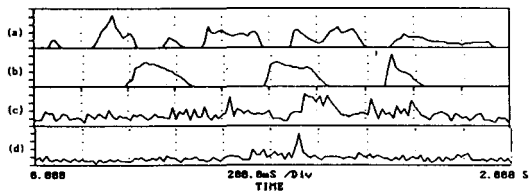

그림 2. 실험에 사용된 신호들의 단구간에너지
(a) 문장1 (b) 문장 2 (c) 잡음1 (d) 잡음2

자동차잡음도 두가지를 사용하였는데, 잡음1은 저속 내지 중속에서 주행중일 때(대략 시속 50Km 내외)의 잡음이고, 잡음2는 고속행중일 때(대략 시속 10Km)의 잡음이다. 그림 2의 (c)와 (d)는 각각 잡음1과 잡음2의 단구간에너지가 변화하는 모습으로서, 상당히 비정체적임을 알 수 있다.

1. 음성-비음성 판별실험

음성검출에는 단구간에너지를 사용하였다. 단구간에너지의 임계값은 실제상황에서는 처음의 일정시간동안의 잡음구간에서 계산하면 되는데, 본 연구에서는 문장1의 경우는 처음 5개의 구간에서, 문장2의 경

우는 처음 20개의 구간에서 계산된 값의 최대값을 사용하였다. 다음 프레임의 에너지가 임계값보다 크면 음성, 작으면 비음성으로 판별하였다. 또, 임계값 논리를 사용하면 고립판별오류가 생길 수 있는데, 이를 제거하기 위하여 3점 스무더(3-Point Smoother)를 사용하였다. 따라서, 한 프레임만큼의 시간지연이 있게 되는데, 이는 비슷한 목적으로 사용되는 이중 임계값[20]보다는 훨씬 시간지연이 적은 것이다. 음성검출알고리즘은 실험에 사용된 문장과 잡음에 대한 어떤 종류의 사전정보도 사용하지 않았다.

그림3은 문장1에 대한 음성-비음성 판별실험결과이다. (a)는 잡음이 섞이지 않은 순수한 음성신호로부터 실험자의 관찰로 판별한 수동식 결과이고, (b)부터 (h)까지는 앞에서 말한 방법1, 즉, 단구간에너지를 사용했을 때의 자동판별결과이다. (b)는 그림1(a)에 있는 잡음이 섞이지 않은 원래신호에 적용했을 때, 즉, SNR이 무한대일 때의 판별결과로써 (a)의 수동식 판별결과와 거의 비슷하다. (c), (d), (e)는 문장1에 잡음1이 더해진 경우로써, SNR이 각각 5dB, 0dB, -5dB인 경우의 판별결과이다. SNR이 적어질수록 오류가 증가함을 알 수 있는데, 5dB, 0dB까지는 약간의 오류는 있으나 전반적으로 수동식과 거의 유사한 결과를 얻었으나 -5dB에서는 오류가 현저히 증가하였는데, 특히, 시작 부분의 유성음구간("The")과 끝부분("new")에서 판별오류가 발생하였다. 오류의 정도는 잡음에너지가 증가할수록 에너지가 적은 무성음구간의 상실로부터 일부 유성음구간의 상실로 악화됨을 알 수 있다. (f), (g), (h)는 문장1에 잡음2(시속 100Km)가 더해진 경우로써, SNR이 각각 5dB, 0dB, -5dB일때의 판별결과이다. 잡음1의 경우보다 전반적으로 판별결과가 부정확함을 알 수 있다. 그 이유는, 입력신호의 처음 다섯개 구간에서 계산된 에너지 임계값을 전구간에 적용했기 때문에 잡음에너지의 비정체성에 대처하지 못한데 있다.

그림4는 문장2에 대한 실험결과이다. (a)는 수동식 판별결과이고, (b)부터 (h)까지는 방법1, 즉, 단구간에너지를 사용했을 때의 자동판별결과이다. (b)는 그림1(b)에 있는 원래신호에 적용했을 때의 판별결과로써 (a)의 수동식 판별결과와는 달리 "둘"과 "셋" 사이의 비음성구간이 음성구간으로 오판되어있다. 그 이유는 그림1(b)에서 보듯이 이 구간의 에너지가 문장 시작부분의 에너지보다 약간 커서 임계값으로

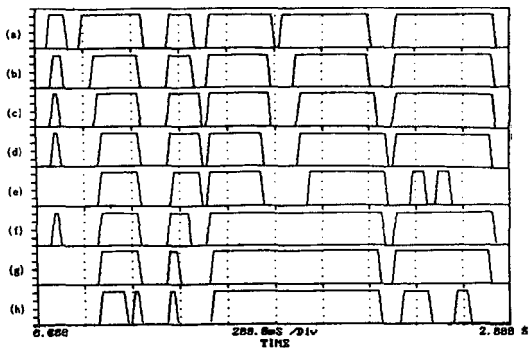

그림 3. 문장1에 대한 음성검출 실험결과
(a) 수동 (b)-(h) 자동

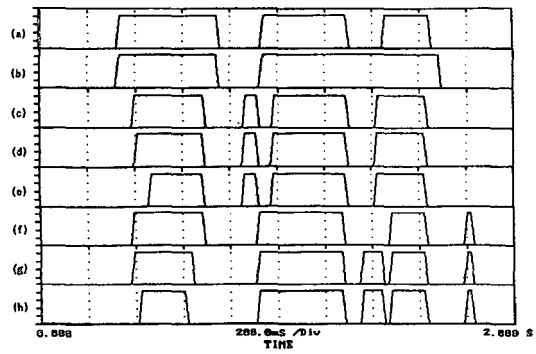

그림 4. 문장2에 대한 음성검출 실험결과
(a) 수동 (b)-(h) 자동

잡히지 않았기 때문인데, 이것은 발성시의 호흡때문인 것으로 추측된다. (c), (d), (e)는 문장2에 잡음1(시속 50Km 내외)이 더해진 경우로써, SNR이 각각 5dB, 0dB, -5dB인 경우의 판별결과이다. 가운데 부분에 비음성을 음성으로 오판한 부분 이외에는 5dB, 0dB까지는 양호한 결과를 얻었으나 -5dB에서는 앞부분의 일부 유성음구간의 축소되었다. 오류의 정도는 잡음에너지의 증가에 상당히 강인함 알 수 있다. (f), (g), (h)는 문장2에 잡음2(시속 100Km)가 더해질 경우로써, SNR이 각각 5dB, 0dB, -5dB일때의 판별결과이다. 잡음1의 경우와 비슷하거나 약간 못한 판별결과라고 볼 수 있는데, 이는 문장1의 경우와 같은 양상이라 할 수 있다.

이상의 실험결과로부터, SNR이 5dB일때는 단구간에너지만으로 상당히 정확한 판별이 가능하나, 0dB 이하에서는 정확도가 저하됨을 알 수 있다.

2. 잡음감쇄 실험결과

실제 상황에서는 신호의 길이가 일정치 않고, 음성은 물론 잡음도 비정체성이므로 적응 필터를 사용해야 하는데, 짧은 시간동안은 정체적이라고 간주할 수 있으므로 프레임단위로 처리하였다. 이와같이 프레임의 길이가 짧을때의 최적의 MMSE 추산기는 칼만필터이다. AR모델에 의거하여 설계된 칼만필터를 사용하여 정체적인 유색잡음을 감쇠하는 방법은 이미 발표된 바 있다[4]. 이상적인 칼만필터의 계수는 순수한 음성신호와 잡음의 AR계수로 구성되는데, [14]에서는 잡음이 정체적인 경우로써 잡음의 AR계수를 전체 잡음시퀀스로부터 직접 계산하여 사용하였다. 물론 이것은 실제 상황에서는 구할 수 없다. 본 연구에서는 잡음의 AR계수를 매 프레임마다 계산하는데 [14]와는 달리 비정체성 잡음을 사용하여 프레임단위로 처리하였다. 단, 주어진 프레임이 음성구간이라고 판단되면 이전의 마지막 잡음 프레임에서 계산된 AR계수를 그대로 사용하고, 잡음 프레임이면 잡음의 AR계수를 새로 계산한다. 음성-비음성 판별에는 IV.1절의 방법을 사용하였다.

음성의 AR계수와 필터출력은 ML(Maximum Likelihood) 추정을 위한 EM기법[32]을 적용하여 반복적으로 계산하는데, 이러한 방법은 두개의 마이크를 사용하는 적응잡음제거에도 이용된 바 있고[33], 2회내지 3회의 반복만으로 최적해에 근접하는 성능을 보인 바 있다. 사용된 EM알고리즘은 다음과 같다.

(1) 입력프레임으로부터 음성의 AR계수 초기값 추산
(2) 칼만필터의 출력 계산
(3) 칼만필터의 출력으로부터 다시 AR계수 추산
(4) (2)-(3) 반복

본 실험에서 음성과 잡음은 각각 AR(10)으로 모델링하였고, 2회 반복하였다. 3회이상 반복하면 SNR의 개선은 매우 적은 반면에 왜곡이 심화되었다.

입력신호의 SNR은 5dB, 0dB, -5dB로 조절하였다. 마이크를 운전석 앞의 계기판에 설치했을 때의 SNR은 -5dB까지도 내려간다는 보고도 있으나[35], 예비실험결과에 의하면 보통 전화하듯이 입 가까이 대고 발성하는 경우에는 5dB이상 될 것으로 추정된다.

잡음감쇄 실험결과 SNR을 표1에 보였다. SNR은

표 1. Kalman Filter 출력의 SNR
(괄호안은 Segmental SNR의 평균값)

문장 번호	잡음 번호	KF Input	KF Outout			
			Case-1	Case-2	Case-3	Case-4
1	1	5(-4.70)	8.60(5.65)	8.51(5.46)	7.97(3.54)	7.86(3.17)
		0(-9.69)	5.85(3.80)	5.81(3.72)	4.72(0.65)	4.57(-0.10)
		-5(-14.71)	3.90(2.54)	3.63(2.22)	1.78(-2.15)	0.47(-3.81)
	2	5(-4.78)	8.47(5.37)	8.48(5.37)	7.56(3.10)	7.72(2.44)
		0(-9.76)	5.61(3.50)	5.61(3.51)	4.21(-0.22)	4.14(-0.61)
		-5(-14.80)	3.59(2.18)	3.55(2.18)	0.93(-3.53)	1.16(-3.62)
2	1	5(-4.70)	9.67(3.66)	9.61(3.58)	8.52(2.58)	8.57(1.84)
		0(-9.69)	6.28(2.31)	6.24(2.26)	4.67(0.32)	4.60(-0.55)
		-5(-14.71)	3.72(1.36)	3.53(1.27)	1.96(-2.12)	1.05(-3.42)
	2	5(-4.78)	9.87(3.65)	9.59(3.51)	9.17(2.65)	8.78(2.15)
		0(-9.76)	6.35(2.25)	5.87(2.14)	5.20(0.20)	4.49(-0.76)
		-5(-14.80)	3.71(1.26)	2.92(1.08)	1.74(-2.65)	0.32(-3.81)

문장의 전 구간에서 계산된 값이고, 괄호안의 숫자는 segmental SNR(각 프레임의 SNR의 평균값)이다. 표에서 다룬 네가지 경우는 다음과 같다.

Case-1 : Ideal Speech Parameter, Ideal Speech Detection

Case-2 : Ideal Speech Paramenter, Practical Speech Detection

Case-3 : Practical Speech Parameter, Ideal Speech Detection

Case-4 : Practical Speech Parameter, Practical Speech Detection

여기서, 'Ideal Speech Parameter'는 음성의 AR계수를 원래의 음성신호에서 계산한 것이고, 'Practical Speech Parameter'는 EM반복법을 사용한 것이고, 'Practical Speech Detection'은 IV.1절에서 제안한 방법을 사용한 것을 말한다. 따라서, Case-1은 이상적인 음성 및 잡음의 계수를 사용한 경우로서 실제로는 사용할 수 없으나, 본 연구에서 제안된 칼만필터를 사용했을 때의 성능의 한계점으로서 그 의의가 있다. Case-3은 음성-비음성 판별이 이상적인 경우이고, Case-4는 음성과 잡음의 계수를 모두 모른다고 가정했으므로 가장 현실적인 경우이다.

표1에서, SNR이 5dB, 0dB, -5dB일때 Case-1, -2의 경우, 대략 3 내지 5dB, 5 내지 6dB, 8 내지 9dB의 개선이 있고, Case-3, -4의 경우는 앞의 경우들보다 대략 1 내지 2dB 정도 개선이 덜 됨을 알 수 있다. SNR이 5dB 및 0dB일때, Case-1과 2를 비교해보면, 차이가 대체로 0.2dB 미만으로 매우 적고, Case-3과 4를 비교해봐도 차이가 적은데, SNR이 -5dB일때는 차이가 심한 경우 1dB이상 벌어짐을 알 수 있다. 이는 SNR이 -5dB일때의 음성검출결과가 부정확한데 기인하는 것이다.

청취실험결과 배경잡음은 거의 완전히 제거되었고, Case-4의 경우에는 어느 정도의 왜곡을 감지할 수 있었다. Case-1의 경우에도 약간의 왜곡은 있었으나 그 정도는 미미하였다. 스펙트럼차감법을 사용할 때의 문제점인 'tonal' 잡음은 전혀 없었다.

그림5의 (a)는 SNR이 0dB일때 문장1에 잡음1을 더했을 때의 파형이고, (b)는 Case-4의 출력 파형으로서 배경잡음이 현저히 줄었으나, 그림1(a)의 원신호와 비교해보면 왜곡이 있음을 알 수 있다. 그 이유로서는 잡음의 계수추산값의 바이어스와 음성-비음성 판별오류에 기인하는 것으로 생각된다.

그림6은 그림5의 (a), (b)에 있는 파형들의 SNRSEG이 시간에 따라 변화하는 모습이다. 모든 구간에서 SNR의 개선정도를 알 수 있는데, 특히 비음성구간에서의 개선이 현저함을 알 수 있다.

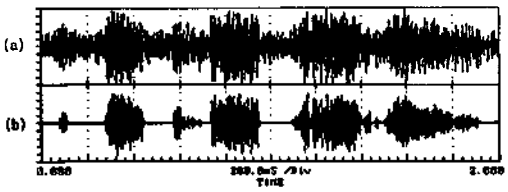

그림 5. 문장1 + 잡음1 파형
(a) 입력(SNR = 0dB) (b) 출력((SNR = 4.57dB)

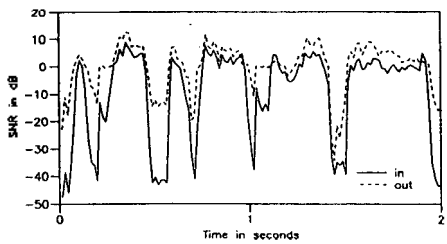

그림 6. SNRSEG vs. Time
(in : SNR = 0dB, out : SNR = 4.57dB)

V. 결 론

본 연구에서는 정체형 칼만필터와 간단한 음성-비음성 판별알고리즘을 사용하여 비정체형 유색잡음을 감쇠시키는 방법을 제안하였다. 기존의 잡음감쇠알고리즘들이 대부분 백색 또는 정체형 잡음을 다룬데 비하여 본 연구는 대부분의 실제 잡음환경, 즉 비백색 비정체성 잡음을 다루었다는 점이 다르다.

칼만필터에 필요한 잡음의 계수는 비음성구간에서 추산하였고, 음성의 계수는 EM반복법을 사용하여 추산하였다. 실험결과, SNR이 -5dB일때에도 5dB 내지 6dB의 SNR 개선이 있었다. 청취실험결과에 의하면, 비음성구간의 배경잡음은 거의 완전히 제거할 수 있었으나, SNR이 0dB 내지 -5dB로 내려갈수록 왜곡

이 심화되었다. 이 왜곡은 음성구간 판별의 오류 및 잡음계수 추산값의 바이어스에 의한 것으로 추정되는데, 청각적으로 불편할 정도이기도 하지만 음성의 명료도를 저하시키지는 않았다. 또, 스펙트럼차감법에서와 같은 'toanl'잡음은 전혀 감지되지 않았다. 잡음계수를 여러개의 구간에서 추산하고, 정확한 음성-비음성 판별알고리즘을 사용하면 왜곡은 감소시킬 수 있을 것이다.

본 연구의 실험에서는 차량잡음만을 사용했으나, 본 연구에서 제안한 방법은 차량잡음의 특성을 전혀 이용하지 않았으므로 일반적인 모든 비백색 비정체성 잡음에 적용할 수 있다.

참 고 문 헌

1. R. L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IEEE Trans. Inform. Th.*, vol. IT-8, pp. 293-304, 1962.
2. C. F. Teacher and D. Coulter, "Performance of LPC vocoders in a noisy environment," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, pp.216-219, April 1979.
3. *Speech Enhancement*, J. S. Lim, Ed., Englewood Cliffs, NJ : Prentice-Hall, 1983.
4. J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, pp. 1586-1604, Dec. 1979.
5. Y. Ephraim, "Statistical-model-based speech enhancemen systems," *Proc. IEEE*, vol. 80, pp. 1526-1555, Oct. 1992.
6. J. R. Deller, Jr., J. G. Proakis and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, New York : Macmillan, 1993.
7. B. Widrow et al., "Adaptive noise cancelling : Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.
8. S. F. Boll and D. C. Pulsipher, "Suppression of acoustic noise in speech using two microphone adaptive noise cancellation," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. ASSP-28, pp. 752-753, Dec. 1980.
9. W. A. Harrison, J. S. Lim and E. Singer, "A new application of adaptive noise cancellation," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. ASSP-34, No.1, pp.21-27, Feb. 1986.
10. M. R. Weiss, E. Aschkenasy, and T. W. Parsons, "Study and development of INTEL technique for improving speech intelligibility," *Nicolet Scientific Corp., Final Tech Rep.*, RADCTR-75-155, June 1975.
11. J. S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-26, No.5, pp.471-472, Oct. 1978.
12. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. ASSP-27, No.2, pp. 113-120, April 1979.
13. J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. ASSP-26, pp. 197-210, June 1978.
14. J. D. Gibson, B. Koo and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Proc.*, vol 39, pp. 732-1742, Aug. 1991.
15. R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. ASSP-28, no. 2, pp. 137-145, April 1980.
16. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square erorr short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.
17. P. G. Drago, A. M. Molinari, and F. C. Vagliani, "Digital dynamic speech detectors," *IEEE Trans. Commu.*, vol. COM-26, pp. 140-145, Jan. 1978.
18. Y. Yatsuzuka, "Highly sensitive speech detector based on sign bit sequence manipulations," *Trans. IECE Japan*, vol. 63-A, pp. 413-419, July 1980.
19. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, N. J. : Prentice-Hall, 1978.
20. L. F. Lamel, L. R. Rabiner, A. E. Rosenberg, and J. G. Wilpon, "An improved endpoint detector for isolated word recognition," *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol. ASSP-29, no. 4, pp. 777-785, Aug. 1981.
21. M. Savoji, "A robust algorithm for accurate

endpointing of speech," *Speech Communications*, vol. 8, pp. 45-60, 1989.

22. B. S. Atal and L. R. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-24, no. 3, pp. 201-212, June 1976.
23. L. R. Rabiner, C. E. Schmidt, and B. S. Atal, "Evaluation of a statistical approach to voiced-unvoiced-silence analysis for telephone-quality speech," *Bell System Tech. J.*, vol. 56, no. 3, pp. 455-482, March 1977.
24. L. J. Siegel, "A procedure for using pattern classification techniques to obtain a voiced/unvoiced classifier," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-27, no. 1, pp. 83-89, Feb. 1979.
25. H. Kobatake, "Optimization of voiced/unvoiced decisions in nonstationary noise environment," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. ASSP-35, no. 1, pp. 9-18, Jan. 1987.
26. Y. Ephraim, J. G. Wilson, and L. R. Rabiner, "A linear predictive front-end processor for speech recognition in noisy environment," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, pp. 31.1.1-31. 1.4, 1987.
27. D. A. Krubsack and R. J. Niederjohn, "An autocorrelation pitch detector and voicing decision with confidence measures developed for noise-corrupted speech," *IEEE Trans. Sig. Proc.*, vol. 39, No. 2, pp. 319-329., Feb. 1991.
28. B. Mak, J. Junqua, and B. Reaves, "A robust speech/nonspeech detection algorithm using time and frequency-based features," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, vol. I, pp. 269-272, 1992.
29. M. M. Goulding and J. S. Bird, "Speech enhancement for mobile teleophony," *IEEE Trans. Vehi. Tech.*, vol. 39, no. 4, Nov. 1990.
30. D. Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," *Signal Processing*, vol. 15, pp. 43-56, 1988.
31. J. Yang, "Frequency domain noise suppression approaches in mobile telephone systems," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, vol. II, pp. 363-366, 1993.
32. A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Annals of Royal Stat. Soc.*, pp. 1-38, Dec.1977.
33. M. Feber, *et. al.*, "Maximum likelihood noise cancellation using the EM algorithm," *IEEE Trans. Acoustics, Speech and Signal Proc.*, vol. 37, No. 2, pp. 204-216, Feb. 1989.

▲具 本 勝(정회원) 1953년 8월 2일생

1975년 2월 : 서울대학교 공과대학 공업교육학과 전자전공(학사)
1984년 12월 : Texas A&M Univ. 전기공학과(석사)
1988년 12월 : Texas A&M Univ. 전기공학과(박사)
1977년 1월~1982년 7월 : 한국원자력연구소(연구원)
1989년 3월~현재 : 경기대학교 전자공학과(부교수)
※주관심분야 : Speech Coding, Speech Enhancement