

음성인식을 위한 복합형잡음제거필터와 최적특징추출에 관한 연구

A study on the Optimal Feature Extraction and
Complex Adaptive Filter for a speech recognition

차태호* · 장승관** · 최응세** · 최일홍*** · 김창석****

(T.-H. Cha · S.-K. Jang · U.-S. Choi · I.-H. Choi · C.-S. Kim)

ABSTRACT

In this paper, a novel method of noise reduction of speech based on a complex adaptive noise canceler and method of optimal feature extraction are proposed. This complex adaptive noise canceler needs simply the noise detection, and LMS algorithm used to calculate the adaptive filter coefficient. The method of optimal feature extraction requires the variance of noise.

The experimental results have shown that the proposed method effectively reduced noise in noisy speech. Optimal feature extraction has shown similar characteristics in noise-free speech.

Keywords : adaptive noise canceler, feature extraction, LMS algorithm

1. 서 론

최근 음성인식시스템의 실용화를 위해서 연구되고 있는 분야는 잡음환경에서의 음성인식이다[1]. 잡음이 없거나 비교적 조용한 환경에서 우수한 성능을 나타내는 음성인식시스템은 주위에 잡음이 존재하는 환경에서는 성능이 급격히 떨어져 사용할 수가 없다. 따라서 잡음환경에서 음성인식시스템의 성능을 향상시키는 것은 음성인식의 실용화를 위해 매우 중요한 과제이다. 음성인식시스템이 점차로 실용화된

* 한국생산기술연구원 산업기술교육센터 기획조정실

** 한국산업기술대학교 전기전자공학과

*** 진주전문대학 전자과

**** 명지대학교 전자정보통신공학부

에 따라 환경변화에 따른 음성인식시스템의 개발에 관한 연구가 많이 이루어지고 있다[2]. 음성인식시스템에 영향을 미치는 요소는 주변에 존재하는 잡음과 같은 부가적인 성분과 채널이나 마이크로폰을 통한 콘벌루션 특성을 갖는 성분으로 구분할 수 있다. 부가적인 잡음이라고 하여도 여러 가지 종류의 잡음에 강인한(robust) 음성인식시스템을 구성하기는 실제로 매우 어렵다.

잡음 환경에서의 음성인식을 위한 처리방식으로는 전처리과정을 통해 잡음을 제거하는 음질개선방식과 청각기관의 모델 등에 근거한 잡음에 강인한 거리 측정법, 이미 만들어진 모델을 토대로 환경에 따라 파라미터를 보상하는 모델적응화 기법 등이 있다[3,4]. 이들 방식 중에서 음질개선방식은 음성인식시스템의 전처리과정에서 잡음을 제거하는 것으로 기존의 인식시스템의 구조를 변화시키지 않고 처리가능하다는 장점이 있다. 음질개선방식의 구체적인 방법으로는 스펙트럼 차감법(spectral subtraction)[5], 적응필터(adaptive filter)[6,7] 등이 있다. 또한 잡음에 강인한 인식시스템 방식으로는 HMM을 이용하여 잡음이 없는 음성모델과 잡음으로부터 구한 모델을 결합하는 병렬모델결합(parallel model combination)[8] 방법이 있다.

본 논문에서는 잡음 환경에서 수집된 음성신호를 효과적으로 인식하기 위해서 잡음을 제거하는 복합형 잡음제거필터와 잡음이 섞인 음성신호에서 음성신호의 특징만을 최적으로 추출할 수 있는 최적특징을 추출하는 알고리즘을 제안하였다.

복합형 잡음제거 필터는 두 개의 입력과 하나의 출력으로 이루어진 잡음제거 필터로서 필터의 입력 중 하나는 잡음이 섞여있는 음성신호를 입력시키고, 다른 입력에는 잡음신호를 입력으로 하며, 출력은 잡음이 제거된 음성신호가 출력된다.

이 제안한 복합형 잡음제거필터는 두 개의 입력으로 들어오는 신호들의 특성 즉, 잡음이 섞여있는 음성신호와 잡음만의 신호들의 특성을 추정하여 이 추정된 값들을 이용한 적응필터의 필터계수를 추출함으로써 잡음을 제거하는 필터이다.

최적특징추출 알고리즘은 잡음 섞인 음성신호에서 음성신호만의 특징을 추출하는 알고리즘으로서, AR 모델에 의하여 음성신호의 특징을 추정할 때 잡음의 평균 에너지가 0이고, 음성신호와 상관관계가 없다는 통계학적인 특성을 이용하여 음성신호를 대표하는 최적의 LPC(linear prediction coefficient)를 추출할 수 있는 알고리즘이다.

2. 복합형 잡음제거필터

음성인식 시스템은 잡음이 있는 환경에서 사용하는 경우가 많으며, 이러한 경우 잡음을 제거하고 음성인식을 수행하여야 한다. 이렇게 잡음이 있는 환경에서 음성인식시스템을 사용하는 경우 음성을 발생하지 않은 시간에서 잡음을 검출할 수 있으며, 이렇게 검출된 잡음의 통계학적 특성은 음성신호에 포함되는 잡음특성과 동일한 특성을 갖는다. 따라서 이와 같은 특성을 이용하여 잡음이 섞인 음성신호로부

터 동일한 특징을 갖는 잡음을 제거하는 필터가 복합형 잡음제거필터이다.

따라서 본 논문에서는 잡음의 특징을 추정하는 잡음특징 추정블럭, 잡음이 섞인 음성신호의 특징을 추정하는 잡음·음성신호 특징추정블럭과 복합필터 계수추정블럭으로 구성된 그림 2-1과 같은 복합형 잡음제거필터를 제안한다. 각 특징추정블럭에서는 입력되는 신호의 특징을 나타내는 $L+1$ 개의 적응필터계수와 1단계 지연요소로 구성되며, 각 블록별 계수추정알고리즘은 LMS(least mean square) 알고리즘을 사용한다. 잡음특징추정블럭은 입력되는 기준잡음을 처리하여 의하여 $L+1$ 개의 적응필터계수 $\bar{w}_0(k), \bar{w}_1(k), \dots, \bar{w}_L(k)$ 를 추정하고, 잡음·음성신호특징 추정블럭은 잡음이 섞인 음성신호 $x(k)$ 에 의하여 $L+1$ 개의 적응필터계수 $w_0(k), w_1(k), \dots, w_L(k)$ 를 추정한다. 이렇게 추정된 $2(L+1)$ 개의 적응필터계수를 이용하여 복합필터계수 추정블럭에서는 $L+1$ 개의 복합필터계수 $\tilde{w}_0(k), \tilde{w}_1(k), \dots, \tilde{w}_L(k)$ 를 추정한다. 이렇게 추정한 복합필터계수에 의하여 출력된 음성신호 $f(k)$ 와 잡음이 섞인 음성신호 $x(k)$ 의 차인 신호 $\hat{s}(k)$ 가 잡음이 제거된 음성신호이다.

본 논문에서 제안한 복합형 잡음제거필터의 출력과 각 블록별 필터계수추정을 위한 처리식은 다음과 같다.

그림 2-1에서 필터의 출력 $f(k)$, $\bar{f}(k)$ 는 각각 다음과 같다.

$$f(k) = \tilde{w}_0(k)x(k) + \tilde{w}_1(k)x(k-1) + \dots + \tilde{w}_L(k)x(k-L) \quad (1)$$

$$\bar{f}(k) = \bar{w}_0(k)\bar{n}(k) + \bar{w}_1(k)\bar{n}(k-1) + \dots + \bar{w}_L(k)\bar{n}(k-L) \quad (2)$$

또한 각각의 적응필터계수벡터 $\tilde{W}(k)$, $\bar{W}(k)$ 와 입력벡터 $X(k)$, $\bar{N}(k)$ 는

$$\tilde{W}(k) = [\tilde{w}_0(k) \ \tilde{w}_1(k) \ \dots \ \tilde{w}_L(k)]^T \quad (3)$$

$$\bar{W}(k) = [\bar{w}_0(k) \ \bar{w}_1(k) \ \dots \ \bar{w}_L(k)]^T \quad (4)$$

$$X(k) = [x(k) \ x(k-1) \ \dots \ x(k-L)]^T \quad (5)$$

$$\bar{N}(k) = [\bar{n}(k) \ \bar{n}(k-1) \ \dots \ \bar{n}(k-L)]^T \quad (6)$$

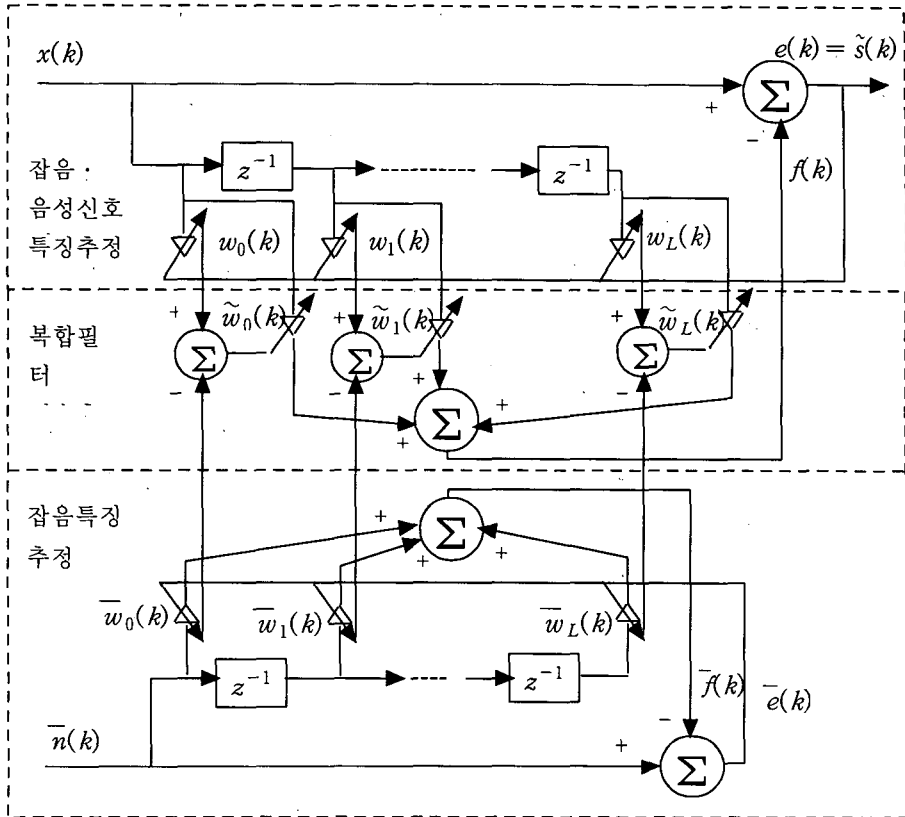


그림 2-1. 복합형 잡음제거필터

이므로 오차 $e(k)$, $\bar{e}(k)$ 는 다음과 같다.

$$\begin{aligned}
 e(k) &= x(k) - f(k) \\
 &= x(k) - \sum_{m=0}^L \tilde{w}_m(k) x(k-m) \\
 &= x(k) - X(k)^T \tilde{W}(k) = x(k) - \tilde{W}(k)^T X(k) \quad (7)
 \end{aligned}$$

$$\begin{aligned}
 \bar{e}(k) &= \bar{n}(k) - \bar{f}(k) \\
 &= \bar{n}(k) - \sum_{m=0}^L \bar{w}_m(k) \bar{n}(k-m) \\
 &= \bar{n}(k) - \bar{N}(k)^T \bar{W}(k) = \bar{n}(k) - \bar{W}(k)^T \bar{N}(k) \quad (8)
 \end{aligned}$$

또한 각각의 적응필터에 LMS 알고리즘을 사용한 적응필터계수는 다음과 같다.

$$w_m(k+1) = w_m(k) + \frac{(1-\beta)e(k)x(k)}{\sigma_x^2(k)} \quad (9)$$

$$\bar{w}_m(k+1) = \bar{w}_m(k) + \frac{(1-\beta)\bar{e}(k)\bar{n}(k)}{\bar{\sigma}_n^2(k)} \quad (10)$$

여기서 분산 $\sigma_x^2(k)$, $\sigma_n^2(k)$ 은 다음과 같고,

$$\sigma_x^2(k) = \beta\sigma_x^2(k-1) + (1-\beta)x^2(k), \quad 0 < \beta < 1 \quad (11)$$

$$\bar{\sigma}_n^2(k) = \beta\bar{\sigma}_n^2(k-1) + (1-\beta)\bar{n}^2(k), \quad 0 < \beta < 1 \quad (12)$$

여기서 β 는 다음 식에 의해서 주어진다[6].

$$\beta \approx 1 - \frac{1}{N} \quad (13)$$

여기서 N 는 데이터 수이다.

복합형잡음제거필터의 복합필터계수 $\tilde{w}_m(k)$ 는 다음 식에 의해서 추정하였다.

$$\tilde{w}_m(k) = w_m(k) - \bar{w}_m(k) \quad (14)$$

그러므로 잡음이 제거된 음성신호 $\hat{s}(k)$ 는 다음 식과 같다.

$$\hat{s}(k) = x(k) - \sum_{m=0}^L \tilde{w}_m(k)x(k-m) \quad (15)$$

식 (15)는 복합형 잡음제거필터로 입력되는 음성신호 $x(k)$ 와 복합필터 계수추정블럭의 출력신호와의 차로 잡음이 제거된 음성신호이다.

3. 최적특징추출알고리즘

음성인식 시스템에서 음성신호를 처리하기 위해서는 성도를 모델링하여 성도파라미터를 추정하는 방법으로서 선형예측방법을 많이 사용하고 있다. 음성신호에 잡음이 섞인 경우에는 잡음제거를 위한 적응필터를 이용하여 잡음을 제거한 후 음성특징추출을 하지만 적응필터를 이용하는 대부분의 경우 완벽하게 잡음을 제거할 수는 없고, 조금이라도 잡음이 있는 상태에서 LPC를 추정하게 되면 추정된 LPC는 잡음신호 때문에 왜곡된 성도의 파라미터가 추정되고 이 특징 파라미터를 이용하여 음성인식을 수행하면 인식률이 저하하게 된다.

따라서 본 논문에서는 효과적인 음성인식을 위하여 잡음이 섞인 음성신호에서 잡음신호에 의한 영향을 최소화 하여 최적의 성도 파라미터를 추정할 수 있는 최

적특징추출알고리즘을 제안한다.

잡음이 섞인 음성신호 $x(k)$ 는

$$x(k) = s(k) + n(k) \quad (16)$$

이고, $s(k)$ 는 깨끗한 음성신호, $n(k)$ 는 잡음이다. 잡음 $n(k)$ 는 분산이 σ^2 이고, 평균이 0인 백색잡음이며 $s(k)$ 와는 통계적으로 독립이다.

성도를 AR모델로 설정하여 성도의 특징추출은 식 (17)의 AR모델의 평가 식을 사용하여 식 (17)에서 AR모델의 p 개 계수 a_i 를 성도파라미터인 LPC라 한다. 임의의 시점 k 에서 잡음 섞인 음성신호 $x(k)$ 로부터 AR모델에 의하여 평가한 $\hat{x}(k)$ 는 그 전 출력들의 선형조합으로 추정할 수 있으며

$$\hat{x}(k) = \sum_{i=1}^p a_i x(k-i) \quad (17)$$

와 같이 표현할 수 있으며 평균자승오차 $e_x^2(k)$ 는 다음 식과 같다.

$$\begin{aligned} e_x^2(k) &= E[\{x(k) - \hat{x}(k)\}^2] \\ &= E[\{s(k) + n(k) - (a_1 s(k-1) + a_1 n(k-1) + a_2 s(k-2) \\ &\quad + a_2 n(k-2) + \dots + a_p s(k-p) + a_p n(k-p))\}^2] \quad (18) \end{aligned}$$

식 (18)에서

$$\hat{s}(k) = a_1 s(k-1) + a_2 s(k-2) + \dots + a_p s(k-p) = \sum_{i=1}^p a_i s(k-i) \quad (19)$$

$$\hat{n}(k) = a_1 n(k-1) + a_2 n(k-2) + \dots + a_p n(k-p) = \sum_{i=1}^p a_i n(k-i) \quad (20)$$

이라 놓으면 식 (19)와 식 (20)은 음성신호와 잡음의 AR모델로서 식 (18)은 다음과 같이 표현할 수 있다.

$$\begin{aligned} e_x^2(k) &= E[\{s(k) - \hat{s}(k)\}^2] + 2E[\{s(k) - \hat{s}(k)\}\{n(k) - \hat{n}(k)\}] \\ &\quad + E[\{n(k) - \hat{n}(k)\}^2] \quad (21) \end{aligned}$$

여기서 $s(k)$ 와 $n(k)$ 가 독립이므로 식 (21)은

$$e_x^2(k) = E[\{s(k) - \hat{s}(k)\}^2] + 2E[s(k) - \hat{s}(k)]E[n(k) - \hat{n}(k)] + E[\{n(k) - \hat{n}(k)\}^2] \quad (22)$$

와 같이 표현할 수 있으며 두 번째 항에서

$$E[n(k) - \hat{n}(k)] = E[n(k)] - E[\hat{n}(k)] = 0 \quad (23)$$

이고, 세 번째 항에서

$$\begin{aligned} & E[\{n(k) - \hat{n}(k)\}^2] \\ &= E\left[\{n(k)\}^2 - 2n(k) \sum_{i=1}^p a_i n(k-i) + \sum_{i=1}^p \sum_{j=1}^p a_i a_j n(k-i)n(k-j)\right] \\ &= \sigma^2(1 + a_1^2 + a_2^2 + \dots + a_p^2) \end{aligned} \quad (24)$$

이므로 식 (22)는 식 (21)과 식 (22)에 의해서 다음 식과 같이 표현된다.

$$\begin{aligned} e_x^2(k) &= E[\{x(k) - \hat{x}(k)\}^2] \\ &= E[\{s(k) - \hat{s}(k)\}^2] + \sigma^2(1 + a_1^2 + a_2^2 + \dots + a_p^2) \end{aligned} \quad (25)$$

또한 식 (25)에서

$$\begin{aligned} E[\{s(k) - \hat{s}(k)\}^2] &= E[\{x(k) - \hat{x}(k)\}^2] - \sigma^2(1 + a_1^2 + a_2^2 + \dots + a_p^2) \\ & \quad (26) \end{aligned}$$

이 되고, 식 (26)의 첫 번째 항은 다음 식과 같다.

$$\begin{aligned} & E[\{x(k) - \hat{x}(k)\}^2] \\ &= E\left[\{x(k)\}^2 - 2 \sum_{i=1}^p a_i x(k-i)x(k) + \sum_{i=1}^p \sum_{j=1}^p a_i a_j x(k-i)x(k-j)\right] \end{aligned} \quad (27)$$

그리고 $x(k)$ 의 공분산(covariance) $\Phi_{xx}(i, j)$ 는 다음 식과 같이 표현할 수 있다.

$$\Phi_{xx}(i, j) = E[x(k-i)x(k-j)] \quad (28)$$

식 (26)에서 음성신호의 평균자승오차 $e_s^2(k)$ 는 식 (28)의 공분산에 의해서 다음과 같이 된다.

$$\begin{aligned} e_s^2(k) &= E[\{s(k) - \hat{s}(k)\}^2] \\ &= \Phi_{xx}(0, 0) - 2 \sum_{i=1}^p a_i \Phi_{xx}(i, 0) + \sum_{i=1}^p \sum_{j=1}^p a_i a_j \Phi_{xx}(i, j) \\ & \quad - \sigma^2(1 + a_1^2 + a_2^2 + \dots + a_p^2) \end{aligned} \quad (29)$$

식 (29)로부터 최적특징 LPC는 평균자승오차 $e_s^2(k)$ 를 a_i 에 대하여 각각 편미분하고, 0으로 놓으면 추출할 수 있다.

$$\begin{aligned} \frac{\partial e_s^2(k)}{\partial a_i} &= -2\Phi_{xx}(i, 0) + \sum_{l=1}^p a_l \Phi_{xx}(l, i) + \sum_{j=1}^p a_j \Phi(i, j) - 2\sigma^2 a_i \\ &= 0 \quad (i=1, 2, \dots, p) \end{aligned} \quad (30)$$

또한 식 (30)에서

$$\sum_{l=1}^p a_l \Phi_{xx}(l, i) + \sum_{j=1}^p a_j \Phi(i, j) - 2\sigma^2 a_i = 2\Phi_{xx}(i, 0) \quad (31)$$

이 되고, 식 (31)을 행렬 형태로 바꾸면 다음과 같다.

$$\begin{aligned} &\begin{bmatrix} 2\Phi_{xx}(1, 1) - 2\sigma^2 & \Phi_{xx}(1, 2) + \Phi_{xx}(2, 1) & \cdots & \Phi_{xx}(1, p) + \Phi_{xx}(p, 1) \\ \Phi_{xx}(2, 1) + \Phi_{xx}(1, 2) & 2\Phi_{xx}(2, 2) - 2\sigma^2 & \cdots & \Phi_{xx}(2, p) + \Phi_{xx}(p, 2) \\ \Phi_{xx}(3, 1) + \Phi_{xx}(1, 3) & \Phi_{xx}(3, 2) + \Phi_{xx}(2, 3) & \cdots & \Phi_{xx}(3, p) + \Phi_{xx}(p, 3) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{xx}(p, 1) + \Phi_{xx}(1, p) & \Phi_{xx}(p, 2) + \Phi_{xx}(2, p) & \cdots & 2\Phi_{xx}(p, p) - 2\sigma^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} \\ &= 2 \begin{bmatrix} \Phi_{xx}(1, 0) \\ \Phi_{xx}(2, 0) \\ \Phi_{xx}(3, 0) \\ \vdots \\ \Phi_{xx}(p, 0) \end{bmatrix} \end{aligned} \quad (32)$$

식 (28)에 의하여 $\Phi_{xx}(i, i+1) = \Phi_{xx}(i+1, i)$ 이므로 식 (32)는 다음과 같이 쓸 수 있다.

$$\begin{aligned} &\begin{bmatrix} \Phi_{xx}(1, 1) - \sigma^2 & \Phi_{xx}(1, 2) & \Phi_{xx}(1, 3) & \cdots & \Phi_{xx}(1, p) \\ \Phi_{xx}(2, 1) & \Phi_{xx}(2, 2) - \sigma^2 & \Phi_{xx}(2, 3) & \cdots & \Phi_{xx}(2, p) \\ \Phi_{xx}(3, 1) & \Phi_{xx}(3, 2) & \Phi_{xx}(3, 3) - \sigma^2 & \cdots & \Phi_{xx}(3, p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi_{xx}(p, 1) & \Phi_{xx}(p, 2) & \Phi_{xx}(p, 3) & \cdots & \Phi_{xx}(p, p) - \sigma^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} \\ &= \begin{bmatrix} \Phi_{xx}(1, 0) \\ \Phi_{xx}(2, 0) \\ \Phi_{xx}(3, 0) \\ \vdots \\ \Phi_{xx}(p, 0) \end{bmatrix} \end{aligned} \quad (33)$$

공분산 함수 $\Phi_{xx}(i, j)$ 와 자기상관함수 $R_{xx}(i-j)$ 에는 다음과 같은 관계식이 성립된다.

$$\Phi_{xx}(i, j) = R_{xx}(i - j) \quad (34)$$

또한 자기상관함수는 $R_{xx}(-k) = R_{xx}(k)$ 이므로 식 (32)는 다음과 같이 다시 표현할 수 있다.

$$\begin{bmatrix} R_{xx}(0) - \sigma^2 & R_{xx}(1) & R_{xx}(2) & \cdots & R_{xx}(p-1) \\ R_{xx}(1) & R_{xx}(0) - \sigma^2 & R_{xx}(1) & \cdots & R_{xx}(p-2) \\ R_{xx}(2) & R_{xx}(1) & R_{xx}(0) - \sigma^2 & \cdots & R_{xx}(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & R_{xx}(p-3) & \cdots & R_{xx}(0) - \sigma^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ R_{xx}(3) \\ \vdots \\ R_{xx}(p) \end{bmatrix} \quad (35)$$

따라서 식 (35)으로부터 최적특징파라미터 a_i 를 추정할 수가 있고, a_i 는 음성인식을 위한 특징벡터이다.

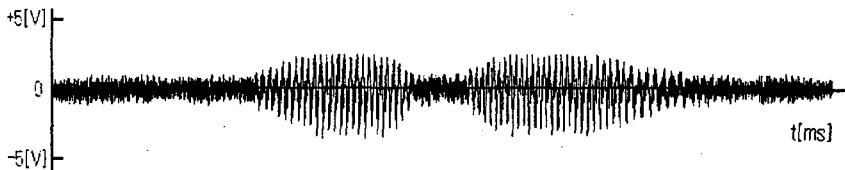
4. 실험 및 고찰

본 실험을 위해 사용된 음성데이터는 50개의 단어이다. 녹음은 조용한 사무실 환경에서 이루어졌고, 발음한 음성신호는 16 bit, 11.025 kHz로 A/D변환하였다. 잡음은 백색잡음을 사용했다.

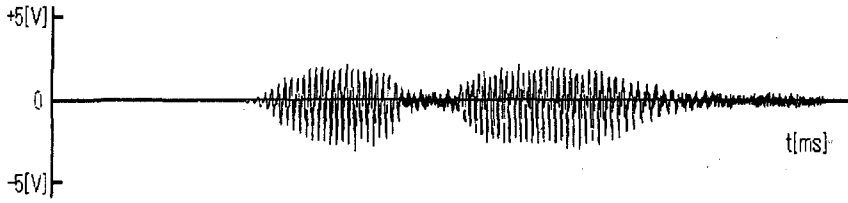
4.1 복합형 잡음제거필터에 의한 잡음처리

식 (1)과 식 (2)에서 잡음제거를 위한 복합형잡음제거필터의 차수 L 은 10차로 설정하였고, 백색잡음이 각각 0, 5, 10, 20, 30 dB로 섞인 음성신호에 복합형 잡음제거필터를 적용하여 잡음을 제거한 후 주파수 평면에서 스펙트럼 분석과 시간평면에서 청취 실험을 수행하였다.

그림 4-1의 (a)는 백색잡음이 5 dB 섞인 음성신호 “아기”이고, (b)는 복합형 잡음제거필터에 의해서 잡음이 제거된 음성신호이다.

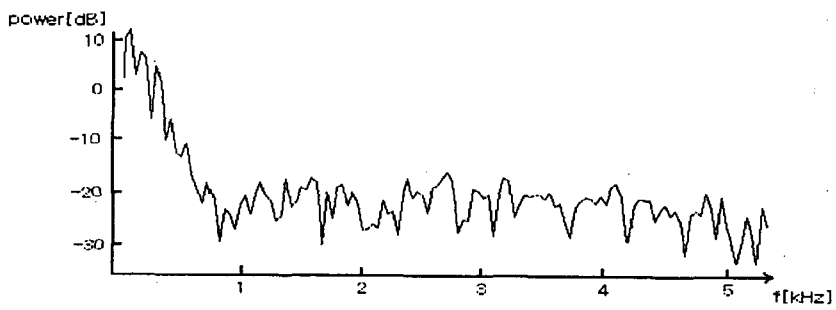


(a) 잡음 섞인 음성신호

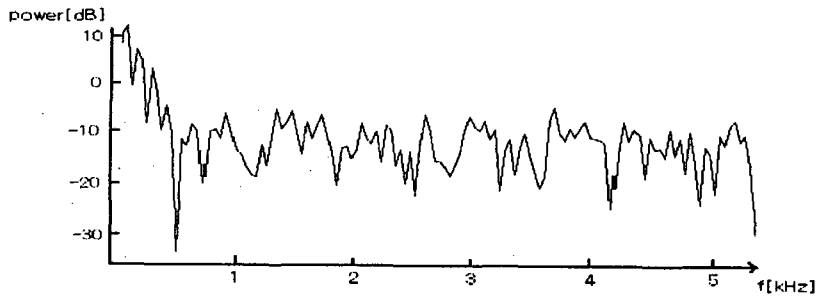


(b) 잡음이 제거된 음성신호

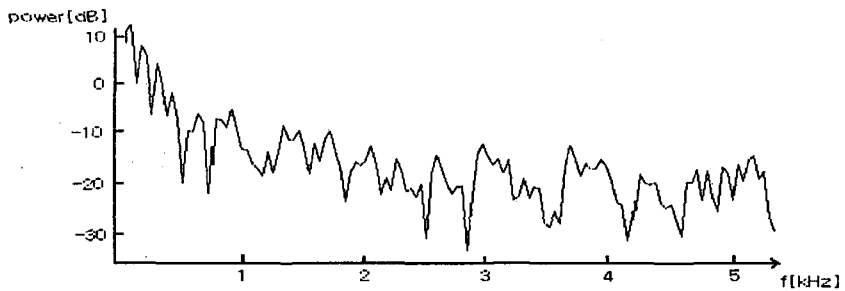
그림 4-1. 복합형잡음제거필터에 의한 잡음 제거된 음성신호



(a) 잡음이 없는 음성신호의 스펙트럼



(b) 잡음 섞인 음성신호의 스펙트럼



(c) 잡음이 제거된 음성신호의 스펙트럼

그림 4-2. 복합형잡음제거필터에 의한 음성신호의 스펙트럼

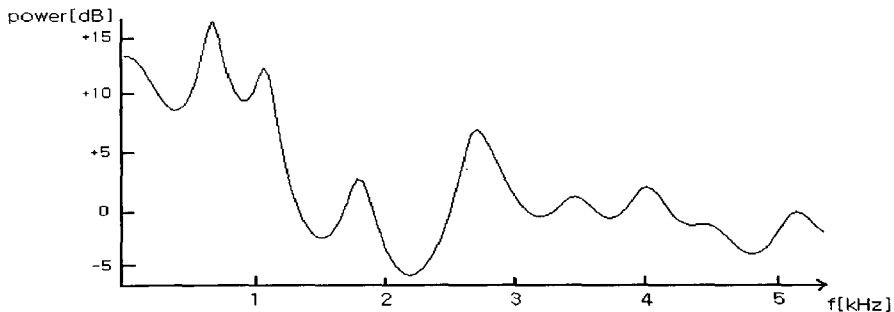
복합형 잡음제거필터에 의한 특성은 단어 음성신호를 256개의 프레임단위로 분할한 후 프레임별로 복합형 잡음제거필터에 의한 스펙트럼 분석을 하였으며, 그림 4-2는 음성 “아기”의 30번째 프레임 데이터에 복합형 잡음제거필터를 적용하여 얻어진 스펙트럼 결과이다. 그림 4-2에서 (a)는 잡음이 없는 음성신호의 스펙트럼이고, (b)는 잡음 섞인 음성신호의 스펙트럼, (c)는 잡음이 제거된 음성신호의 스펙트럼이다.

그림 4-2의 결과로부터 그림 (b)의 잡음 섞인 음성신호의 스펙트럼 평가는 백색잡음 성분에 의하여 주파수 전 대역에 걸쳐 스펙트럼 성분이 -5 dB ~ 10 dB에 존재하고, 그림 (c)는 본 논문에서 제안한 복합형 잡음제거필터에 의하여 백색잡음의 스펙트럼 성분이 많이 제거되어 -15 dB ~ -25 dB로 스펙트럼이 (b)보다 낮아졌음을 알 수 있다.

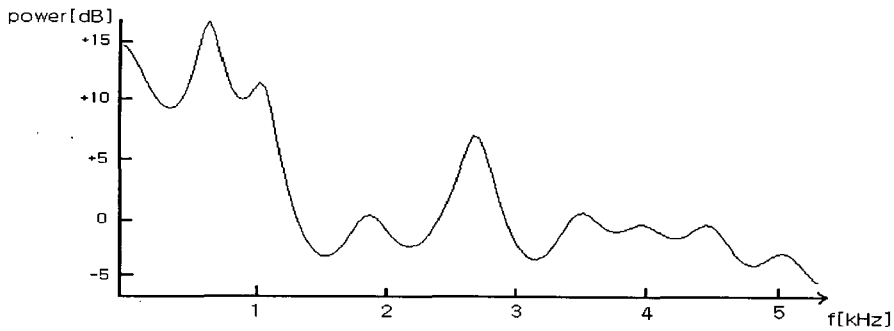
4.2 최적특징추출알고리즘에 의한 특징추출

최적특징추출 알고리즘에 의한 음성신호의 특징추출은 잡음이 없는 음성신호, 백색잡음이 5 dB 섞인 음성신호, 복합형 잡음제거필터에 의하여 잡음이 제거된 음성신호에 적용하여 특징을 추출한 후 각각의 특징을 비교·분석하였다.

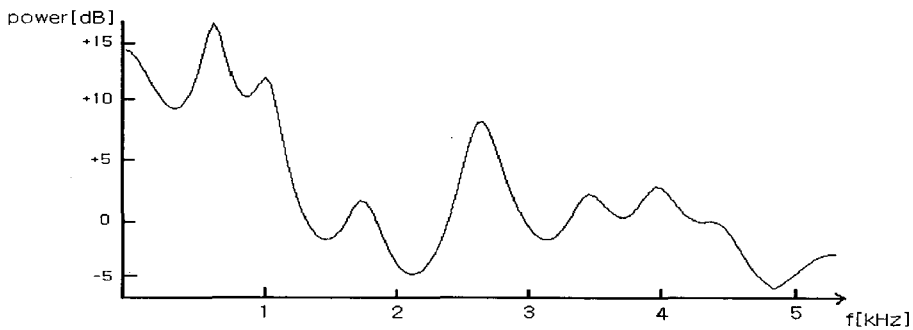
그림 4-3 (a)는 잡음이 없는 음성 “아기”에서의 LPC 스펙트럼이고, 그림 (b)는 잡음 섞인 음성신호에 제안한 최적특징추출 알고리즘을 적용하여 추출한 LPC 스펙트럼이고, 그림 (c)는 음성 “아기”에 5 dB의 잡음 섞인 음성신호를 복합형 잡음제거필터에 의하여 잡음을 제거한 후 제안한 최적특징추출 알고리즘에 의하여 추출한 LPC 스펙트럼이다. 그림 4-2로부터 잡음 섞인 음성신호에서 추출한 그림 (b)의 LPC 스펙트럼과 잡음이 없는 음성신호에서 추출한 그림 (a)의 LPC 스펙트럼과 비교하면 약 4 kHz 부근에서 스펙트럼이 왜곡되어 있음을 알 수 있다. 그림 (c)는 잡음 섞인 음성신호를 복합잡음제거필터를 통과시켜서 잡음을 제거한 후 추출한 LPC 스펙트럼으로 그림 (a)를 비교하면 커다란 차이가 없음을 알 수 있다. 따라서 본 논문에서 제안한 최적특징추출 알고리즘은 잡음 섞인 음성신호에서 잡음의 영향을 줄여서 음성신호만의 특징을 추출하는 알고리즘임을 알 수 있다.



(a) 잡음이 없는 음성신호의 LPC 스펙트럼



(b) 잡음 섞인 음성신호의 LPC스펙트럼



(c) 잡음을 제거한 음성신호의 LPC 스펙트럼

그림 4-3 최적특징알고리즘에 의한 LPC 스펙트럼

5. 결론

본 논문에서는 잡음 환경에서 수집된 음성신호를 효과적으로 잡음을 제거하기 위해서 잡음을 제거하는 복합형 잡음제거필터와 잡음 섞인 음성신호에서 음성신호의 특징만을 최적으로 추출할 수 있는 최적특징추출 알고리즘을 제안하였으며, 제안한 방법의 타당성을 검토하기 위하여 백색잡음이 섞인 음성신호로부터 잡음제거와 최적특징 추출실험을 수행하였다.

복합형 잡음제거필터를 적용한 실험에서는 백색잡음이 섞인 음성신호를 효과적으로 제거하였고, 5 dB의 백색잡음이 섞인 음성신호에서 약 7 dB의 잡음제거가 이루어졌다.

최적특징추출 알고리즘의 실험은 잡음 섞인 음성신호와 복합형 잡음제거필터에 의하여 잡음이 제거된 음성신호에 제안한 알고리즘을 적용하여 최적특징 파라미터 LPC를 추출하여 주파수 평면에서 LPC 스펙트럼을 비교·분석하였고, 이 결과 제안한 알고리즘에 의하여 추출한 특징이 잡음이 없는 음성신호의 스펙트럼과 매우 유사함을 확인하였다.

참 고 문 헌

- [1] A. Acero and R. M. Stern. 1990. "Environmental Robustness in Automatic Speech Recognition." in *Proc. ICASSP*, 849-852.
- [2] Y. Gong. 1995 "Speech Recognition in Noisy Environments : A Survey." *Speech Communication*, 16, 261-291.
- [3] B. H. Juang. 1991 "Speech recognition in adverse environments." *Computer Speech and Language*, 5, 275-294.
- [4] J. C. Junque and J. P. Hato. 1996. *Robustness in Automatic Speech Recognition*. Kluwer Academic Publishers.
- [5] S. F. Boll. 1979 "Suppression of Acoustic Noise in Speech Using Spectral Subtraction." *IEEE Trans. Acoust., Speech Signal Processing*, ASSP-27(2), 113-120.
- [6] B. Widrow and S. D. Stearns. 1985. *Adaptive Signal Processing*. Prentice-Hall, Inc. Englewood Cliffs.
- [7] J. S. Lim. 1982. *Speech Enhancement*. Prentice Hall Inc.
- [8] M. J. Gales and S. J. Young. 1993. "Cepstral Parameter Compensation for HMM Recognition in Noise." *Speech Communication*, 12(3), 231-239.
- [9] 김창석 외. 1992. "음성신호의 최적특징을 적응적으로 추출하는 방법에 관한 연구." *한국통신학회논문지*, 19(2), 373-380

접수일자 : '98. 9. 12.

게재결정 : '98. 11. 1.

▲ 차태호

서울 금천구 가산동 219-5

한국생산기술연구원 산업기술교육센터 기획조정실(우 : 153-023)

Tel: (02) 8603-477 (O), Fax: (02) 8603-582

e-mail : lmscth@ined.ittc.re.kr

▲ 장승관

경기도 시흥시 정왕동 3가 101

한국산업기술대학교 전기전자공학과(우 : 429-450)

Tel: (0345) 4968-262 (O), Fax: (0345) 4968-279

e-mail : skjang@kup.ac.kr

▲ 최응세

경기도 시흥시 정왕동 3가 101

한국산업기술대학교 전기전자공학과(우 : 429-450)

Tel: (0345) 4968-261 (O), Fax: (0345) 4968-279

e-mail : uschoi@kup.ac.kr

▲ 최일홍

경남 진주시 문산읍 상문리 산270
진주전문대학 전자과(우 : 660-759)
Tel: (0591) 751-8193 (O), Fax: (0591) 761-7407
e-mail : ilchoi@namgang.chinju-u.ac.kr

▲ 김창석

경기도 용인시 남동 산38-2
명지대학교 전자정보통신공학부(우 : 449-728)
Tel: (0335) 37-2300 (O), Fax: (0335) 30-6374
e-mail : kimcs@wh.myongji.ac.kr