

ACHARF ANC를 채용한 화자인증시스템의 성능분석

이 현 승*, 최 홍 섭, 신 윤 기
대전대학교 전자공학과

Performance analysis of speaker verification system adopting the ACHARF ANC

Hyun Seung Lee*, Hong Sub Choi, Yoon Ki Shin
Dept. of Electronics, Daejin University
E-mail: hschoi@road.daejin.ac.kr

Abstract

The development of noise robust speech processing systems is becoming increasingly important as speech technology is currently widely applied in real world applications. Recently, to resolve such a noise problem, adaptive noise canceller(ANC) is frequently used, which is based upon adaptive filters. The adaptive recursive filters perform better than adaptive non-recursive filters due to the added poles, but the stability may be severely threatened. But these problems of adaptive recursive filters was solved by ACHARF algorithm. This paper presents a method which combines speaker verification system with ANC(Adaptive Noise Canceller) using the ACHARF algorithm. In the front-end stage, ANC is adopted to suppress the additive noise imposed on the speech signal. The results show that the performance of speaker verification system becomes better than before.

I. 서론

최근 들어 주변 환경과 잡음에 강인한 음성처리 알고리즘에 관한 연구가 활발히 진행되고 있으며 이는 제반 음성처리시스템을 상용화하는데 있어 매우 중요한 기반기술이 된다. 본 논문에서는 이러한 주변 잡음의 영향을 줄이기 위하여 화자인식시스템의 전처리과

정에 화자가 발생한 음성신호와 함께 입력으로 들어오는 잡음을 제거하기 위하여 적응신호처리 방법을 적용하였다.

일반적으로 적응신호처리 방법은 음성신호 처리기술이 적용된 시스템을 실용화하기 위해 잡음에 의해 손상된 음성에 대한 청취성을 높여주는 음질 개선 기술에 대한 해법으로 많은 연구 및 개발이 지속적으로 이루어져 왔다. 이러한 음질 개선에 있어서 배경 잡음, 또는 간섭 신호의 통계적 특성이 알려져 있지 않거나 신호와 동일한 대역을 점유할 경우는 기존의 고정 필터로는 신호와 잡음을 효과적으로 분리하기 어렵다. 이를 해결하기 위해 적응 필터를 이용한 적응 잡음 제거기(이하 ANC)가 매우 효과적으로 사용될 수 있으며 이에 대한 연구가 많이 진행되어 왔다[1,2]. 이러한 적응 잡음 제거기에는 주로 적응 비순환 필터가 사용되어 왔는데 이는 적응 비순환 필터로 구성된 ANC가 안정적이기는 하지만 시스템 모델링에 있어서 대략적인 접근 밖에 할 수 없을 뿐만 아니라 필터의 차수가 높아 계산량이 많고 수렴속도가 느려서 음성신호와 같이 빠르게 변화하는 신호에 적용하기에는 적절치 못한 단점이 있기 때문이다. 이에 비해 적응 순환 필터는 영점과 극점을 갖는 실제의 잡음에 대한 모델링을 정확히 할 수 있지만 안정성이 보장되지 않는다는 큰 단점이 있어 실제로 사용되지 않았으나[3], 근자에 적응 보상기를 채용한 초안정적 적응 순환 필터 알고리즘인 ACHARF(Adaptive Compensated Hyperstable Adaptive Recursive Filter)가 제시됨으로써 시스템에

실제 적용이 가능하게 되었다[4]. 따라서 본 논문에서는 ACHARF 알고리즘을 사용한 ANC를 화자인증 시스템의 전처리 과정에 사용하여 화자인증 시스템에 주는 성능향상에 대하여 고찰하여 보았다.

II. 적응순환필터를 이용한 ANC 시스템

그림 1의 (a)는 ANC 시스템의 개념도이며 잡음 $d(k)$ 와 $x(k)$ 는 동일한 잡음원 $n(k)$ 로부터 발생한다. 센서 1에는 원하는 신호 $s(k)$ 와 잡음 $d(k)$ 가 동시에 입력되지만 센서 2에는 잡음 $x(k)$ 만 입력되고 원하는 신호는 입력되지 않게 한다. 그림 (a)에서 경로 1과 경로 2의 전달 함수는 알 수 없으며 이들을 각각 $H_1(z)$, $H_2(z)$ 로 하면 $d(k)=H_1(q)n(k)$, $x(k)=H_2(q)n(k)$ 이므로 $d(k)=[H_1(q)/H_2(q)]x(k)$ 의 관계가 성립된다. 따라서 그림 (a)는 그림 (b)와 같은 구조로 모델링할 수 있다. 그림 (b)에서 적응 필터가 $B(z) \rightarrow H_1(q)$ 및 $1-A(z) \rightarrow H_2(q)$ 로 수렴하면 $y(k) \rightarrow d(k)$ 가 되어 $e(k) \rightarrow s(k)$ 로 수렴함으로써 잡음이 제거된 신호를 얻을 수 있는 것이다. 그러나 적응 순환 필터가 $e(k)$ 의 자승이 0으로 수렴하도록 적응되는 과정에서 극점인 $1-A(z)$ 의 근이 단위 원 밖으로 나갈 수 있으므로 이를 막기 위해 그림 (c)와 같이 적응 보상기를 채용함으로써 항상 초안정성을 만족하도록 한다[4]. 적응 순환 필터의 전달 함수는

$$\frac{B(z)}{1-A(z)} = \frac{\sum_{j=0}^M b_j(k) z^{-j}}{1 - \sum_{i=1}^N a_i(k) z^{-i}} \quad (1)$$

이며 보상기는 다음과 같다.

$$C(z) = 1 - \sum_{l=1}^L c_l(k) z^{-l} \quad (2)$$

따라서 그림 1(c)의 추정 출력 $y(k)$ 는 다음과 같다.

$$\begin{aligned} y(k) &= B(q)x(k) + A(q)y(k) \\ &= \sum_{j=0}^M b_j(k)x(k-j) + \sum_{i=1}^N a_i(k)y(k-i) \end{aligned} \quad (3)$$

그리고 추정 오차 $e(k)$ 및 보상기 출력인 처리된 오차 $v(k)$ 는 각각 다음과 같다.

$$\begin{aligned} e(k) &= d_s(k) - y(k) \\ &= d_s(k) \\ &\quad - \left[\sum_{j=0}^M b_j(k)x(k-j) + \sum_{i=1}^N a_i(k)y(k-i) \right] \end{aligned} \quad (4)$$

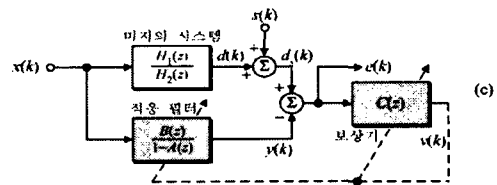
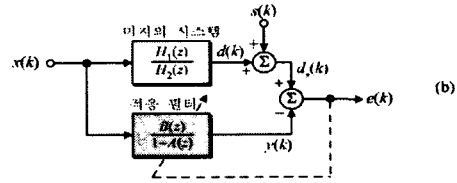
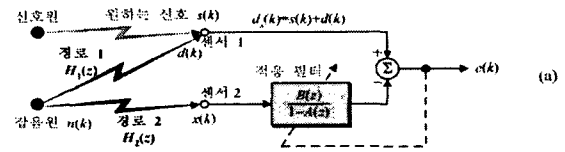


그림 1. 적응 순환 필터를 이용한 ANC 시스템의 (개념도 (b) 구조 (c) 보상기를 채용한 구조

$$v(k) = e(k) - \sum_{l=1}^L c_l(k) e(k-l) \quad (5)$$

이상으로 식 (1), (2)의 각 계수들에 대한 적응 알고리즘은 다음과 같다.

$$\begin{aligned} a_i(k) &= a_i(k-1) \\ &\quad + \frac{\mu_i}{1 + \hat{\Phi}_e^T(k) \Lambda_e \hat{\Phi}_e(k)} v^o(k) y(k-i) \\ b_j(k) &= b_j(k-1) \\ &\quad + \frac{\rho_j}{1 + \hat{\Phi}_e^T(k) \Lambda_e \hat{\Phi}_e(k)} v^o(k) x(k-j) \\ c_l(k) &= c_l(k-1) \\ &\quad + \frac{\gamma_l}{1 + \hat{\Phi}_e^T(k) \Lambda_e \hat{\Phi}_e(k)} v^o(k) e(k-l) \end{aligned} \quad (6)$$

단 식 (6)에서 사전 처리 오차 $v^o(k)$ 는

$$v^o(k) = e^o(k) \sum_{l=1}^L c_l(k-1) e(k-l) \quad (7)$$

로 정의되며 $e^o(k)$ 는 사전 오차로서 다음과 같다.

$$e^o(k) = d_s(k) - \left[\sum_{j=0}^M b_j(k-1)x(k-j) + \sum_{i=1}^N a_i(k-1)y(k-i) \right] \quad (8)$$

식(6)에서 데이터 벡터 $\hat{\Phi}_e(k)$ 는

$$\hat{\Phi}_e(k) = [x(k) \cdots x(k-M), y(k-1) \cdots y(k-N), e(k-1) \cdots e(k-L)]^T$$

(9)

로 정의된다. 그리고 이득 행렬 Λ_e 는

$$\Lambda_e = \text{diag}[\mu_1 \cdots \mu_N, \rho_0 \cdots \rho_M, \gamma_1 \cdots \gamma_L] \quad (10)$$

로서 모두 상수이며 보편적으로 다음과 같이 정한다.

$$\begin{pmatrix} \mu_1 = \cdots = \mu_N = \mu \\ \rho_0 = \cdots = \rho_M = \rho \\ \gamma_1 = \cdots = \gamma_L = \gamma \end{pmatrix} \quad (11)$$

III. GMM을 이용한 화자인식 시스템

화자인증은 발생된 음성이 원하는 화자, 즉 의뢰인지 또는 사칭자인지를 구분해 내는 것으로 의뢰인에 대한 초기등록이 요구된다. 초기등록은 화자의 음성을 모델링하는 것으로 논문에서는 문장독립형 화자 인식기에서 가장 많이 사용하는 방법인 GMM을 사용하였다[5]. 인증 시스템은 그림 1과 같이 입력 음성신호와 그 음성을 발생한 화자에 대한 인증번호가 같이 주어지며, 시스템은 그 인증번호에 해당되는 화자에 대한 참조 패턴과 입력된 음성신호가 일치하는지를 검사하여 일치 또는 불일치라는 판정을 내리게 된다. 즉 입력음성에 대한 화자인증은 다음과 같은 기준에 의해 판정을 내리게 된다.

화자 인증 : if $d(T(x), P(x)) \leq$ 임계치

화자 거부 : if $d(T(x), P(x)) >$ 임계치

$T(x)$: 검사할 입력패턴

$P(x)$: 기준화자에 대한 참조패턴

$d(T(x), P(x))$: 입력 패턴과 참조패턴과의 거리

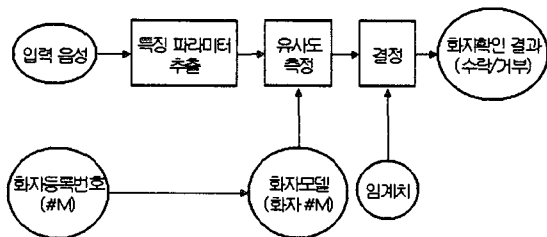


그림 2. 화자인증시스템의 구성

IV. 실험 및 고찰

4.1. 음성 DB

음성 DB로는 YOHO 표준 DB를 사용하였다. YOHO DB는 문장중속 화자인식 실험을 위해 만들어졌으며, 총 138명의 화자 데이터로 구성되었다. 각각의 화자마다 4번의 등록세션과 10번의 테스트세션이 있

며, (35-72-41)과 같이 숫자를 발음한 파일이 등록세션에는 24개씩, 그리고 화자인증실험을 위한 테스트세션에는 4개씩의 음성파일이 있다. 데이터들은 21-97까지의 56개의 숫자들의 조합으로 구성되었으며, 고급 전화 마이크를 사용하여 조용한 환경의 사무실에서 녹음하였다.

본 실험에서는 반복적인 실험을 위하여 계산량과 데이터량을 고려하여 대상 화자의 수를 20명으로 제한하였다. 기준화자의 등록을 위하여 등록세션에 있는 24개 파일을 모두 합친 약 1분 40초 길이의 데이터를 사용하였고, 테스트에는 YOHO DB의 각 테스트세션의 4개의 파일을 합친 20초 데이터를 만들어 세션별로 10회 실험하였다.

4.2. 시뮬레이션 및 실험방법

미지의 잡음에 대한 경로설정은 표 1의 실험조건에 전달함수로 가정하였으며, 화자인증시스템에 들어오는 잡음을 시뮬레이션을 통하여 구하고 이를 음성신호에 부가하여 음성의 SNR이 0dB가 되도록 만들었다. 먼저 잡음이 부가되지 않은 원음에 대하여, 다음으로 ACHARF ANC를 통과한 음성신호를 실험하였다. 그림 3은 원래의 음성신호의 파형을 나타내고 있으며, 합성된 잡음을 부가하여 SNR이 0dB인 음성신호는 그림 4에 있다. 그림에서 보아도 원래의 음성신호의 형태가 잡음에 의해 완전히 가려진 것을 볼 수 있으며 청취 테스트를 통하여도 원래의 음성을 들을 수가 없을 정도로 음질이 좋지 않았음을 알 수 있다. 이렇게 심하게 왜곡된 음성신호를 ACHARF ANC를 통과하여 처리한 신호의 파형이 그림 5에 표시하였는데, 초반부의 잡음성분은 순환적응필터가 수렴하는 동안에 나오는 오차신호에 의한 것임을 알 수 있으며, 대략 2000-2500샘플 정도의 시간이 소요되어 NLMS계열의 비순환적응필터에 비해 탁월한 수렴속도를 보여줄 수 있다.

표 1. 실험조건

기준시스템		$H(z) = \frac{2 - Z^{-1}}{1 - Z^{-1} + 0.75Z^{-2}}$
입력신호	s(k)	음성신호
	x(k)	백색잡음
	SNR	0dB
ACHARF 적응순환필터		영점 차수 = 1차 극점 차수 = 2차 보상기 차수 = 15
		적용 이득

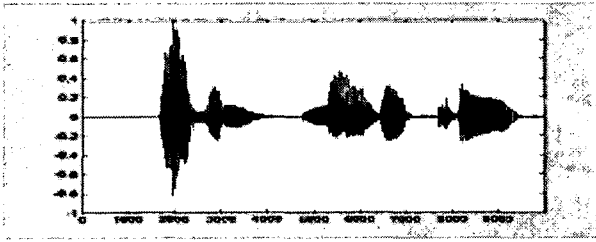


그림 3. 원래의 음성신호

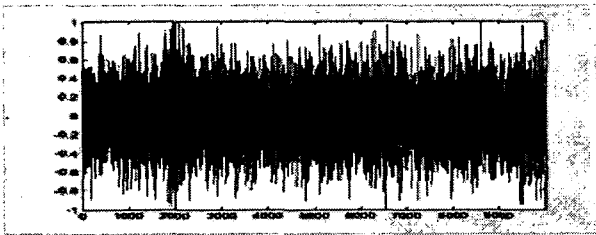


그림 4. 잡음이 부가된 음성신호의 파형 (SNR = 0dB)

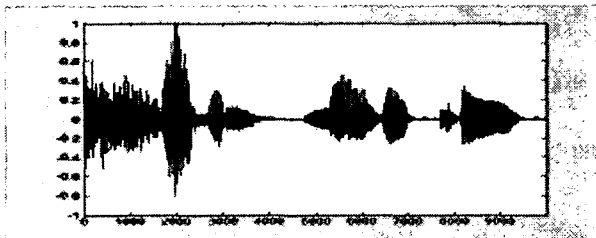


그림 5. 적응시스템에 의해 처리된 음성신호

아래의 표2는 원래의 음성신호와 적응시스템에 의해 복원된 음성신호를 화자인증 시스템을 통해서 성능평가를 한 결과를 보여준다. 화자인증에서 사용한 유사도는 2가지로 normalized log-likelihood 와 maximum log-likelihood를 사용하여 비교하여 보았다. 결과에는 보이지 않았지만 ANC를 이용하여 전처리를 하지 않은 0dB의 음성신호에 대한 인증실험의 결과는 거의 인식이 되지 않을 정도로 시스템의 성능이 떨어지는 것을 알 수 있었는데, ANC를 사용하여 전처리과정을 거친 음성신호는 EER이 normalized log-likelihood 유사도를 사용한 경우에는 2.85%로 원음의 결과인 1.17%에 근접하고 있음을 볼 수가 있다. maximum log-likelihood 유사도인 경우에도 복원된 음성에서는 3.75%와 원음에서는 1.57%를 나타내서 잡음처리의 효과를 확인할 수가 있었다.

표 2. 화자인증 시스템의 결과 EER

	normalized log-likelihood	maximum log-likelihood
원음	1.17	1.57
적용처리된 음성	2.85	3.75

V. 결론

본 논문에서는 화자인증시스템의 전처리 과정에서 음성과 함께 입력으로 들어오는 미지의 잡음을 제거하는 방법으로 적응순환필터를 이용한 ACHARF ANC를 적용하였다. 영점과 극점이 모두 있는 ACHARF ANC는 기존의 NLMS ANC에 비해 충분히 적은 차수로도 보다 나은 성능을 얻을 수 있을 뿐만 아니라 보상기의 존재로 적응순환필터의 초안정성을 보장해 주고 더욱 정확한 시스템 식별이 가능한 것이 특징이다. 특히 ACHARF ANC는 수렴속도가 매우 빨라, 음성관련 시스템에 적용할 시 NLMS ANC에 비해 시스템의 인식 성능을 향상시킴을 확인할 수 있었다. 화자인증시스템은 GMM방법으로 구성하였으며, 사용한 음성DB는 YOHO 데이터로 먼저 깨끗한 음성을 대상으로 훈련과 인식실험을 했을 경우의 결과는 EER이 1.17 %로 나왔다. 반면에 잡음을 섞어서 음성의 SNR을 0dB로 한 경우에는 전처리를 하지 않을 경우에는 음성이 잡음에 가려 청취가 힘들 정도로 인식이 되지 않았지만, 이를 ACHARF ANC를 통과시킨 이후의 음성신호에 대한 화자인식은 원음에 가까운 성능으로 EER이 2.85 % 정도로 개선되었다. 따라서 ACHARF 알고리즘을 사용한 ANC가 음성처리 시스템의 잡음처리 방법에 적용할 수 있는 가능성을 확인하였다

참고문헌

- [1] Widrow et al. "Adaptive noise canceling: principles and application" *Proc. of IEEE*, vol. 1, pp. 1692-1716, Dec. 1975
- [2] 윤병우, 신윤기 "적용 보상기를 채용한 초안정성 적응순환필터," *전자공학회 논문지*, 제 27권, 제3호, 1990년 3월.
- [3] G. Long, D. Shwed, and D. D. Falconer, "Study of a Pole-Zero adaptive echo canceller," *IEEE Circuits Syst.*, vol. CAS-34, Special Issue on Adaptive Systems and Applications. July 1987.
- [4] S. Boll, and D. pulsipher, "Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancellation", *IEEE Tr. ASSP-28*, pp. 752-753, 19
- [5] Gish, H. and Schmidt. M., "Text-independent speaker identification", *IEEE Signal Processing Magazine*, pp. 18-32, Oct. 1994.