
평균 예측 LMS 알고리즘을 이용한 반향 잡음에 강인한 HMM 학습 모델

안찬식*, 오상엽**

Echo Noise Robust HMM Learning Model using Average Estimator LMS Algorithm

Chan-Shik Ahn*, Sang-Yeob Oh**

요 약 음성 인식 시스템은 다양하게 변화하는 환경 잡음에 빠르게 적응할 수 없어서 인식 성능을 저하시키는 요인이 된다. 본 논문에서는 평균 예측 LMS 알고리즘을 이용하여 반향 잡음에 강인하게 하는 방법으로 HMM 학습 모델을 구성하는 방법을 제안하였으며, 변화하는 반향 잡음에 적응하도록 HMM 학습 모델을 구성하여 인식 성능을 평가하였다. 실험 결과 변화하는 환경 잡음을 제거하여 얻은 음성의 SNR은 평균 3.1dB이 향상되었고 인식률은 3.9% 향상되었다.

주제어 : 반향 잡음 제거, 평균 예측, LMS 필터, 적응 필터, MHH 모델

Abstract The speech recognition system can not quickly adapt to varied environmental noise factors that degrade the performance of recognition. In this paper, the echo noise robust HMM learning model using average estimator LMS algorithm is proposed. To be able to adapt to the changing echo noise HMM learning model consists of the recognition performance is evaluated. As a results, SNR of speech obtained by removing Changing environment noise is improved as average 3.1dB, recognition rate improved as 3.9%.

Key Words : Echo Noise Cancellation, Average Estimator, LMS(Least Mean Square) filter, adaptive filter, HMM(Hidden Markov Model) Model

1. 서론

컴퓨터 하드웨어의 발전과 신호 처리 기술 및 알고리즘의 발전에 힘입어 실험적인 결과보다는 실용적인 측면을 부각시킨 인식 시스템의 결과들을 활용하는 연구가 활발하게 진행되고 있다. 그러나 음성 인식의 실용적인 측면을 저하시키는 요인은 환경적인 잡음과 채널 잡음이다.[1] 이를 극복하기 위해 잡음의 특성을 고려한 필터를 설계하는 방법들이 제안되고 있다. 잡음 제거를 위해 가장 널리 사용되는 필터로는 LMS(Least Mean Square) 적응 유한 임펄스 응답(FIR) 필터가 있다.[2]

LMS 필터는 FIR 필터의 성능으로 인하여 좋은 성능을 나타내기 위해 중요한 역할을 하지만 필터 설계가 복잡하고 변화하는 잡음 환경에 간섭을 받는 채널 잡음이 포함되어 적응 필터 계수 값이 커지게 된다. 또한 입력 신호와 예측이 가능한 확률 값의 상관도가 높은 학습 모델이 구성되어 있을 때, 모델 상호간의 상관 계수가 높은 패턴 모델의 계수로 인한 인식 성능이 저하된다.[4]

따라서 본 논문에서는 변화하는 반향 잡음을 평균 예측량을 이용한 LMS 검출 알고리즘을 이용하여 반향 잡음을 제거하고 인식 성능을 향상시키기 위해 반향 잡음에 강인한 HMM 학습 모델을 구성하고 인식 실험을 수

*본 논문은 2012년 가천대학교의 교내연구비 지원에 의한 결과임(GCU-2012-R183).

*광운대학교 컴퓨터공학과 박사과정

**가천대학교 IT대학 인터랙티브미디어학과 교수(교신저자)

논문접수: 2012년 10월 10일, 1차 수정을 거쳐, 심사완료: 2012년 11월 10일

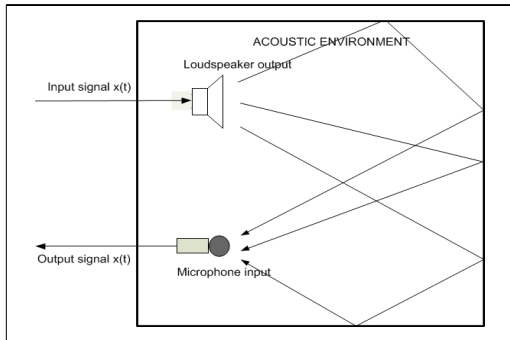
행하였다. 변화하는 반향 잡음에 강인하도록 HMM 학습 모델을 구성하고 평균 예측량을 이용한 반향 잡음 제거 알고리즘을 적용하여 실험한 결과 음성의 SNR은 평균 3.1dB이 향상되었고 인식률은 3.9% 향상되었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 언급하고 3장에서는 제안한 HMM 학습 모델 구성 방법에 대해 설명하며, 4장에서는 시스템 평가를 수행하고 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

2.1 LMS 반향 잡음 제거 필터

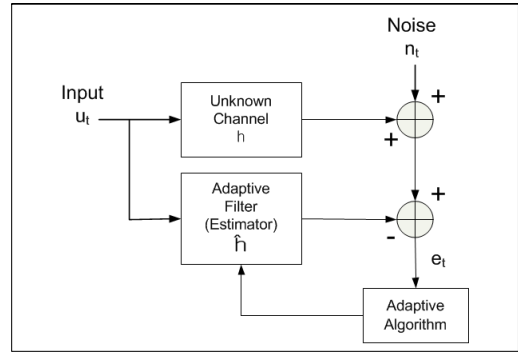
잡음은 환경적인 요인의 의해 수많은 변화를 나타내며 특정 공간에서의 다양한 잡음의 형태는 매우 다양하다. 또한 음향의 반향 경로는 매우 유동적이며 공간에서의 사물의 움직임, 온도, 기압, 습도에 따라 반사되는 형태가 달라진다.[7][9]



[그림 1] 반향 잡음 블록 다이어그램

그림 1은 반향 잡음 블록 다이어그램을 나타내었다. 시간적으로 변화하는 임펄스 응답 필터는 다양한 잡음의 변화에 적응할 수 있도록 수학적으로 효과적인 모델링이 가능하다.[6]

반향 잡음 제거기의 입력 신호 $u(n)$ 으로부터 $y(n)$ 은 음성 신호이고, $\hat{y}(n)$ 은 예측치로 출력 $v(n)$ 에서 감산된다. 출력 $v(n)$ 과 예측치 $\hat{y}(n)$ 의 차이 값이 에러 신호이며 이는 $e(n)$ 으로 표시하며 차이 값인 에러 신호가 존재하면 far-end로 보내지게 되고, 차이 값이 존재하지 않고 같은 값을 나타내면 near-end 신호만 가지게 된다. 그림 2는 일반적인 LMS 잡음 제거기의 기본 개념이다.



[그림 2] 일반적인 LMS 잡음 제거기

실제 환경에 존재하는 반향 잡음을 디지털 영역 안에서 계산하게 되므로 임펄스 응답을 가지는 선형 시변 필터를 이용하여 모델링이 가능하다. 임펄스 응답 $h(n, l)$ 은 다음과 같이 음성 신호를 $y(n)$ 으로 표현 가능하다.[8]

$$y(n) = \sum_{l=-\infty}^{+\infty} h(n, l) u(n-l) \quad (1)$$

출력 $v(n)$ 은 입력 신호 $u(n)$ 으로부터의 음성 신호 $y(n)$ 과 $noise(n)$ 인 near-end 신호로 구성되었으며 다음 식으로 표현이 가능하다.

$$v(n) = y(n) + noise(n) \quad (2)$$

식(2)에 식(1)을 대입하여 다음 식으로 표현한다.

$$v(n) = \sum_{l=0}^{+\infty} h(n, l) u(n-l) + noise(n) \quad (3)$$

선형 시변 필터인 임펄스 응답 $h(l)$ 은 지수 함수로 감소하는 envelope을 나타내기 때문에 l 에 비해 매우 작은 값을 나타낸다. 따라서 $h(n, l) = 0, (l \geq n)$ 로 표현되며 다음 식으로 표현 가능하다.[9]

$$v(n) = \sum_{l=0}^{N-1} h(n, l) u(n-l) + noise(n) \quad (4)$$

예측치 신호 $\hat{y}(n)$ 은 far-end 신호 $u(n)$ 을 임펄스 응답 신호 $h(n, l)$ 의 조건에 만족하는 선형 시 불변 필터 $hw(n, l)$ 에 필터링하여 나타낼 수 있으며 다음과 같이 표현한다.

$$\hat{y}(n) = \sum_{l=0}^{N-1} hv(n, l) u(n-1) \quad (5)$$

예측치 신호 $\hat{y}(n)$ 과 출력 신호 $v(n)$ 은 far-end로 보내어 추적한다.[8]

$$e(n) = v(n) - \hat{y}(n) \quad (6)$$

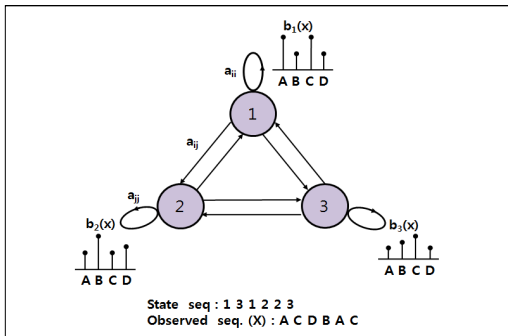
반향 잡음 제거기의 제어부에는 잡음 제거 필터로 구성되어 있기 때문에 정확한 조건에 만족하는 의미를 가지고 있으며 예측치 신호 $\hat{y}(n)$ 과 출력 신호 $v(n)$ 의 최소한의 차이 값을 보이는 조건에 만족해야 하며 일반적인 기준은 다음과 같이 표현한다.[7]

$$hv_{opt} = \arg[\min E|e(n)|^2] \quad (7)$$

hv 와 hv_{opt} 는 $N \times 1$ 벡터이며, l 번째 요소는 잡음 제거 필터 $hv(n, l)$ 과 최적 잡음 제거 필터 $hv_{opt}(n, l)$ 의 각각의 l 번째 필터 계수를 표현한다. 잡음 제거 필터 $hv(n, l)$ 은 적응 필터이므로 데이터가 입력되면 자동적으로 필터의 계수를 갱신한다.

2.2 HMM(Hidden Markov Model)

HMM은 시간에 따라서 변화하는 신호를 모델링하기 위한 방법으로 두 가지의 랜덤 프로세스로 이루어진다. 첫 번째는 숨겨진 상태 천이 확률 과정을 처리하는 과정이며 두 번째는 각 상태에서 관측 신호를 발생하는 관측 확률 처리 과정으로 이루어진다.



[그림 3] HMM 상태 모델

그림 3은 HMM의 상태 모델을 나타내며 각 상태에서

는 A, B, C, D의 관측 심볼을 출력한다. HMM의 상태 천이 확률 및 관측 확률은 시간에 독립적이며 현재 상태에만 의존한다고 가정하고 계산한다. HMM은 초기 ($t = 0$)에 상태 i 의 확률 $\pi_i = \Pr(s_0 = i)$, 상태 i 에서 j 로의 천이 확률 $a_{ij} = \Pr(s_t = i, s_{t+1} = j)$, 상태 j 에서 심볼 k 를 관측할 확률 $b_j(k) = \Pr(x_t = k | s_t = j)$ 로 표현한다.[3]

입력의 음성 특징 벡터의 관측열 $O = (o_1, o_2, \dots, o_T)$ 이 사실임을 가정할 때 주어진 N-states HMM 모델에서의 상태 열이 $q = (q_1, q_2, \dots, q_T)$ 라면 관측열의 확률은 초기 상태 $t = 1$ 에서 확률 $a_{s_0, q_1}(\pi_{s_0})$ 로 천이가 시작되며, 관측 O_1 는 출력 확률 $b_{q_1}(O_1)$ 로서 생성이 된다. 초기 상태 s_1 에서 상태 s_2 로의 천이는 천이 확률 a_{s_1, s_2} 로 이루어지며, 대응되는 상태 s_2 에서의 관측 O_2 를 생성될 확률은 $b_{s_2}(O_2)$ 가 된다.

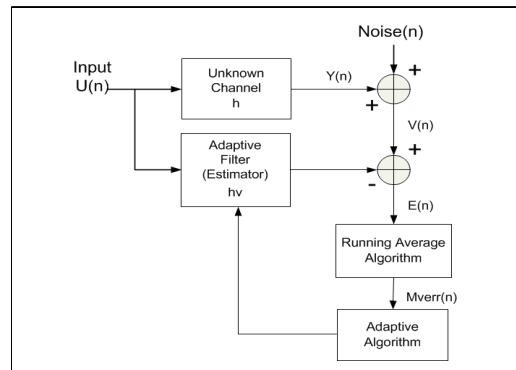
3. 잡음에 강인한 HMM 학습 모델

3.1 평균 예측량 LMS 반향 제거 필터

평균 예측량을 이용한 필터는 선형 시 불변 시스템으로 구성되며 수렴 가중치 인자인 평균값이 커지면 안정성이 감소하고, 평균값이 작아지면 안정성은 증가하게 된다. 따라서 평균 예측량을 이용한 LMS 필터는 평균값을 작게 유지하여 안정된 필터로 구성하며 다음 식과 같이 표현된다.[5]

$$AMerr[n] = \frac{1}{L} \sum_{k=0}^{L-1} e[n-k] \quad (8)$$

L -point 평균 예측 필터는 시간 n 의 $e[n]$ 의 $L-1$ 개의 입력 값으로부터 평균값을 계산하여 출력한다.



[그림 4] 평균 예측 LMS 필터를 이용한 잡음 제거기

그림 4에서는 제안한 평균 예측 LMS 필터를 이용한 반향 잡음 제거기를 나타낸다.

3.2 잡음에 강인한 HMM 학습 모델 구성

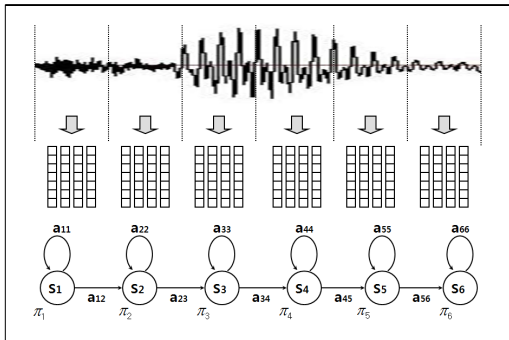
그림 4과 같이 구성된 반향 잡음 제거기를 이용하여 반향 잡음을 제거하며 반향 잡음이 제거된 깨끗한 음성 신호를 HMM을 이용하여 인식을 위한 모델로 구성한다. HMM은 관측할 수 없는 과정과 음성 신호로부터 관측해야 하는 상태로 음향학적 벡터를 연결하여 관측이 가능하게 만드는 과정으로 구성된다. HMM에서는 관측할 수 없는 신호의 통계적인 특성을 관측 가능한 벡터 열을 통해 추정함으로써 음성 신호의 통계적인 변이성을 반영한다. HMM 모델은 음성 구간의 변이에 의한 상태전이 확률 A_i 와 각 상태에서의 관측 심볼의 출력 확률 B_j , 초기 확률 π 로 표현할 수 있으며 다음 식(9)과 같이 정리된다.

$$\lambda = (A, B, \pi) \tag{9}$$

$$A = \{a_{ij}\}, \quad a_{ij} = p[q_{t+1} = j | q_t = i]$$

$$B = \{b_j(k)\}, \quad b_j(k) = p[o_t = v_k | q_t = j]$$

$$\pi = \{\pi_i\}, \quad \pi_i = p[q_1 = i]$$



[그림 5] HMM 모델링

HMM 기반 인식 시스템은 확률적 모델을 근거로 하는 전향, 후향, 바움-웰츠 재추정과 비터비로 구성된다. 학습 모델을 만드는 바움-웰츠 재추정 알고리즘은 최적의 모델을 구성하고 파라미터로 주어진 관측열의 확률을 반복적으로 증가시켜 관측열의 확률을 최대가 되도록 HMM을 그림 5와 같이 구성한다.

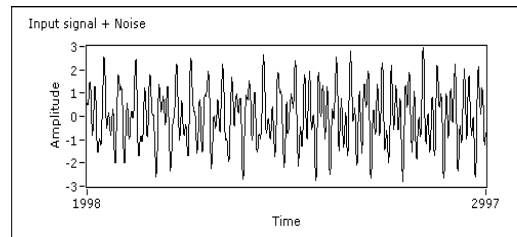
4. 실험 결과 및 분석

본 논문에서 제안한 평균 예측 LMS 알고리즘을 이용한 반향 잡음에 강인한 HMM 모델을 구성하여 인식 성능을 실험하였다. 제안된 방법을 실험하기 위해 사용된 잡음은 핑크 노이즈를 반향 잡음으로 사용하였다. 5명의 화자의 임의의 음성을 각각 10개의 음성에 SNR이 5dB가 되도록 혼합하여 반향 잡음 제거 실험을 수행하였고, 이를 평가하기 위해 식 (10)을 이용하여 SNR의 향상 정도를 평가하였다.

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^N x^2(n)}{\sum_{n=1}^N [x(n) - \hat{x}(n)]^2} \tag{10}$$

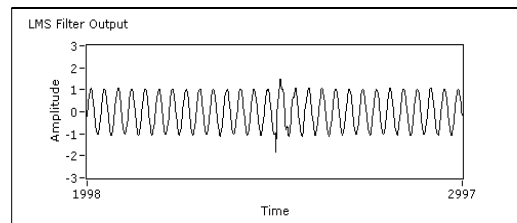
$x(n)$ 은 잡음이 혼합된 음성을 의미하고 $\hat{x}(n)$ 은 잡음이 제거된 음성을 의미하며 n 은 시간의 인덱스를 나타낸다.

그림 6에서는 잡음이 합성된 신호를 나타내었으며 입력 신호로 사용하였다.



[그림 6] 잡음이 합성된 신호

그림 7은 이동 평균 LMS 알고리즘을 수행하여 얻은 출력 신호이다. 원 음성 신호에 가깝게 잡음이 제거된 것을 확인할 수 있다.



[그림 7] 제안 알고리즘에 의해 잡음이 제거된 신호

또한, 인식 실험을 위해서 음성 DB는 ETRI에서 제작

한 445DB를 사용하였으며 16K, Mono, 8K로 실험하였다. 15명의 화자의 음성을 학습 모델로 사용하였으며, 5명의 화자의 음성을 테스트 데이터로 활용하였다. 학습 모델은 프레임별 채널에 해당하는 128차를 사용하여 HMM 모델을 구성한 후 인식 성능을 평가하였다. 비교 평가를 위해 MFCC로 모델을 구성한 기존 음성 인식 시스템과 본 논문에서 제안한 방법의 인식률을 평가하였다. 음성 어휘는 445DB 내에 있는 단어를 무작위로 10개의 단어를 발성하여 테스트 하였으며 인식률은 3.9% 향상되었다.

〈표 1〉 음성 인식률

인식	인식률(%)	
	MFCC	HMM
10개 단어	93.8	97.7

5. 결론

본 논문은 평균 예측량을 이용한 LMS 알고리즘을 이용한 반향 잡음 HMM 모델 구성을 제안하고 반향 잡음 제거와 인식 성능을 평가하였다.

환경에 의해 변화하는 반향 잡음 환경에 간섭을 받는 채널 잡음이 포함되어 적응 필터 계수 값이 커지게 된다. 또한 입력 신호와 예측이 가능한 확률 값의 상관도가 높은 학습 모델이 구성되어 있을 때, 모델 상호간의 상관 계수가 높은 패턴 모델의 계수로 인한 인식 성능이 저하된다. 이러한 성능 저하 현상을 평균 예측량을 이용한 LMS 알고리즘을 이용하여 반향 잡음을 제거하였다.

본 논문에서는 변화하는 반향 잡음을 평균 예측량을 이용한 LMS 검출 알고리즘을 이용하여 반향 잡음을 제거하고 인식 성능을 향상시키기 위해 반향 잡음에 강인한 HMM 학습 모델을 구성하고 인식 실험을 수행하였다. 변화하는 반향 잡음에 강인하도록 HMM 학습 모델을 구성하고 평균 예측 양을 이용한 반향 잡음 제거 알고리즘을 적용하여 실험한 결과 음성의 SNR은 평균 3.1dB 이 향상되었고 인식률은 3.9% 향상되었다.

참고 문헌

- [1] 박윤식·장준혁. (2009). 새로운 잡음전력 추정 기법을 적용한 음향학적 반향 및 배경잡음 제거 통합시스템. 한국음향학회지. 28(7), 680-685.
- [2] 박홍원·안규영·송진영·남상원. (2003). 상관 LMS 적응 기법을 이용한 비선형 반향신호 제거에 관한 연구. 대한전기학회:학술대회논문집:정보및제어부분B, 882-885.
- [3] 안찬식·오상엽. (2012). CHMM 어휘 인식에서 형상 형성 제어를 이용한 가우시안 모델 최적화. 한국디지털 정책연구 논문지. 14(7), 167-172.
- [4] 이수정·신계현·김순협. (2007). 음성강화를 위한 이동 평균 예측량 기반의 검출방법 최적화. 전자공학회논문지. 44(3), 97-104.
- [5] 이세원·박호중. (2003). 잡음에 강인한 반향 제거기 연구. 한국음향학회지. 22(4). 261-268
- [6] 황성수·윤상석·김석찬·이채동(2010). NLMS 적응 필터 기반의 음향 반향 제거기. 한국통신학회논문지. 35(4C). 343-349.
- [7] H. Michael. (2003). Acoustic echo cancellation digital signal processing. Bachelor of Engineering thesis, The school of electrical engineering, The university of Queensland.
- [8] J. Homer. (2000). Detection guided NLMS estimation of sparsely parameterized channels. IEEE Transactions on Circuits and Systems U:Analog and Signal Processing. 47(12).
- [9] J. Homer & I. Mareels. (2004), LS detection guided NLMS estimation of sparse system. Proceedings of the IEEE 2004 International Conference on Acoustic. Speech, and Signal Processing(ICASSP). Montreal, Quebec, Canada.
- [10] S. Haykin. (2002). Adaptive Filter Theory. PrenticeHall. Upper Saddle River. New Jersey.
- [11] S. L. Gay. (2000). Acoustic Signal Processing for Telecommunication. Kluwer Academic Publishers. Dordrecht. The Netherlands.
- [12] V. Edward. (1999). Signal prewhitening schemes for detection-guided LMS estimates. Department of Electrical and Computer Engineering, University of Queensland, Brisbane, Australia.

안 찬 식



- 2002년 2월 : 광운대학교 대학원 컴
퓨터공학과 (공학석사)
- 2004년 2월 : 광운대학교 대학원 컴
퓨터공학과 (박사수료)
- 관심분야: 음성인식, 음향 신호처리
- E-Mail : coolsahn@gmail.com

오 상 엽



- 1991년 2월 : 광운대학교 대학원 전
자계산학과 (이학석사)
- 1999년 2월 : 광운대학교 대학원 전
자계산학과 (이학박사)
- 2007년 2월 ~ 현재 : 가천대학교 IT
대학 인터랙티브미디어학과 교수
- 관심분야: 버전관리, 형상관리, 음성/음향 신호 처리, 차량
통신
- E-Mail : syoh@gachon.ac.kr