
연속 음성 인식 향상을 위해 LMS 알고리즘을 이용한 CHMM 모델링

안찬식*, 오상엽**

CHMM Modeling using LMS Algorithm for Continuous Speech Recognition Improvement

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

요약 본 논문은 반향 제거 평균 예측 LMS 알고리즘을 이용하여 반향 잡음에 강인한 연속 음성 인식 모델인 CHMM 모델을 구성하는 방법을 제안하였다. 변화하는 반향 잡음에 적응하고 연속 음성 인식 성능 향상을 위한 반향 잡음 제거 평균 예측 LMS 알고리즘을 이용하여 CHMM 모델을 구성하였다. 제안한 알고리즘에 의해 구성된 CHMM 모델에 대하여 연속 인식 성능을 평가하였다. 실험 결과 변화하는 환경 잡음을 제거하여 얻은 음성의 SNR은 평균 1.93dB이 향상되었고 연속 음성의 인식률은 2.1% 향상되었다.

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

Abstract In this paper, the echo noise robust CHMM learning model using echo cancellation average estimator LMS algorithm is proposed. To be able to adapt to the changing echo noise. For improving the performance of a continuous speech recognition, CHMM models were constructed using echo noise cancellation average estimator LMS algorithm. As a results, SNR of speech obtained by removing Changing environment noise is improved as average 1.93dB, recognition rate improved as 2.1%.

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

1. 서론

음성 인식 시스템은 하드웨어와 신호 처리 기술, 알고리즘의 발전으로 인하여 시스템의 성능이 향상되고 적용 분야도 확대되고 있다. 하지만 인식 시스템의 실용적인 측면을 저해시키는 요인으로 잡음의 환경적인 변화를 들 수 있으며 변화하는 잡음에 적용하지 못하여 인식 시스템의 성능 저하를 야기 시키고 있다.[1][3]

환경 잡음을 제거하고 성능을 향상시키기 위하여 잡음의 특성을 고려한 필터를 설계하는 방법들이 제안되었고 적응 필터를 이용한 LMS(Least Mean Square) 방법 [8]이 사용되었으며 반향 잡음 제거를 위해 평균 예측 LMS(Least Mean Square) 필터를 이용한 반향 잡음 제

거 알고리즘[6]이 사용되었다.

FIR 필터를 기반으로 사용하는 LMS 필터는 복잡한 필터 설계 과정과 채널 잡음이 포함되어 적응 계수 값의 증가로 인해 적응 시간이 길어져 잡음 제거율이 저하되는 원인이 된다. 또한 연속 음성 인식인 문장을 입력하여 인식할 경우 어휘와 어휘 사이의 잡음이 포함되어 혼합된 음성으로 입력되면 인식률을 저하시키는 원인이 되며 연속 음성 인식을 위한 잡음이 제거된 모델 구축이 중요한 요인으로 작용한다.[10][13]

따라서 본 논문에서는 연속 음성 인식 시스템을 위하여 반향 잡음 제거 평균 예측 LMS 알고리즘을 이용한 CHMM(Continuous Hidden Markov Model) 모델을 구축하고 인식 실험을 수행하였다.

* 본 논문은 2012년 가천대학교의 교내연구비 지원에 의한 결과임(GCU-2012-R203).

*광운대학교 컴퓨터공학과 박사과정

**가천대학교 IT대학 인터랙티브미디어학과 교수(교신저자)

논문접수: 2012년 10월 28일, 1차 수정을 거쳐, 심사완료: 2012년 11월 20일

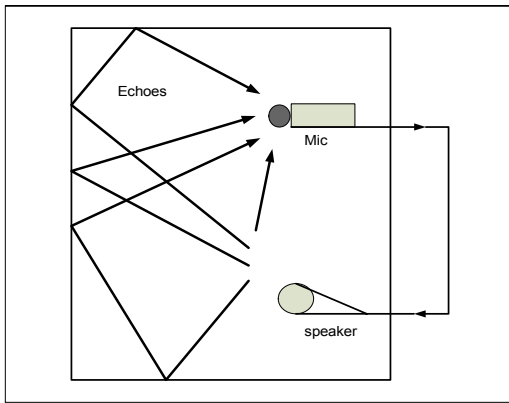
변화하는 반향 잡음에 강인하도록 평균 예측 LMS 알고리즘을 적용하고 CHMM 모델을 구성하여 실험한 결과 음성의 SNR은 평균 1.93dB이 향상되었고 연속 음성의 인식률은 2.1% 향상되었다.

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 요인에 민감하게 작용하여 많은 변화가 일어나며 환경에 따라 잡음의 형태는 매우 다양하게 나타난다. 반향에 의한 잡음 또한 환경에 따라 다양한 형태로 나타나고 있다. 이러한 반향의 경로는 사물의 움직임, 온도, 기압, 습도에 따라 반사되는 형태가 달라진다. 그림 1은 반향 잡음 경로를 다이어그램으로 나타내었다.[9][14]



[그림 1] 반향 잡음 경로 다이어그램

실제 환경의 반향 잡음을 디지털 신호로 변환하여 계산하려면 임펄스 응답을 가진 선형 시변 필터를 이용하여 모델링한다. 임펄스 응답 $r(n, l)$ 은 음성 신호를 $v(n)$ 으로 식(1)과 같이 나타낸다.[11][12]

$$v(n) = \sum_{l=-\infty}^{+\infty} r(n, l) x(n-l) \quad (1)$$

출력 신호 $y(n)$ 은 입력 신호 $x(n)$ 의 음성 신호 $v(n)$ 과 $noise(n)$ 으로 구성되며 식(2)와 같이 나타낸다.

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

식(2)에 식(1)을 대입하고 임펄스 응답 $r(l)$ 은 지수 함수로 감소하여 l 에 비해 매우 작은 값을 나타내며 $r(n, l) = 0 (l \geq n)$ 로 표현되며 식(3)과 같이 나타낸다.[11]

$$y(n) = \sum_{l=0}^{N-1} r(n, l) x(n-l) + noise(n) \quad (3)$$

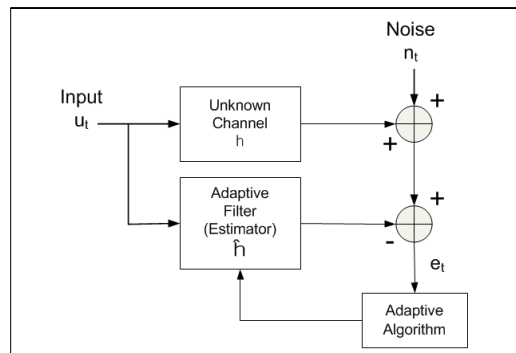
예측 신호 $\hat{v}(n)$ 은 far-end 신호 $x(n)$ 을 임펄스 응답 신호 $r(n, l)$ 의 조건에 만족하는 $vy(n, l)$ 에 필터링하여 식(4)와 같이 나타낸다.

$$\hat{v}(n) = \sum_{l=0}^{N-1} vy(n, l) xu(n-l) \quad (4)$$

예측치 신호 $\hat{v}(n)$ 과 출력 신호 $y(n)$ 은 far-end로 보내어 추적한다.[7]

$$e(n) = y(n) - ry * U \quad (5)$$

그림 2는 수식으로 설명한 일반적인 반향 잡음 제거기를 나타낸다.[9]



[그림 2] 일반적인 반향 잡음 제거기

2.2 HMM(Hidden Markov Model)

HMM을 이용한 음성의 모델링은 음성으로부터 추출

한 특징 파라미터를 관측 열로 직접 모델링하는 연속 출력 확률 분포 HMM과 추출한 특징 파라미터를 벡터 양자화(Vector Quantization: VQ) 과정을 거쳐 코드 북의 코드 워드 인덱스로 매핑한 관측 열을 출력 확률로써 모델링하는 이산 출력 분포 HMM으로 구분할 수 있다.

신호를 모델링하기 위한 방법은 두 가지로 이루어진다. 첫 번째는 숨겨진 상태 천이 확률 과정을 처리하는 과정이며 두 번째는 각 상태에서 관측 신호를 발생하는 관측 확률 처리 과정으로 이루어진다.

HMM에서는 관측할 수 없는 신호의 통계적인 특성을 관측 가능한 벡터 열을 통해 추정함으로써 음성 신호의 통계적인 변이성을 반영한다. 따라서 음성 구간의 변이에 의한 상태 천이 확률 A_i 와 각 상태에서의 관측 심볼의 출력 확률 B_j , 초기 확률 π 로 표현할 수 있으며 다음 식 (6)과 같이 정리된다.[3]

$$\lambda = (A, B, \pi) \tag{6}$$

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

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

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

3. CHMM(Continuous Hidden Markov Model) 모델링

3.1 반향 잡음 제거 평균 예측 LMS 알고리즘

반향 잡음 제거 알고리즘은 잡음 제거 필터로 구성되어 예측치 신호 $\hat{v}(n)$ 과 출력 신호 $y(n)$ 의 최소 차이 값을 만족해야 하며 식(7)과 같이 나타낸다.[5]

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

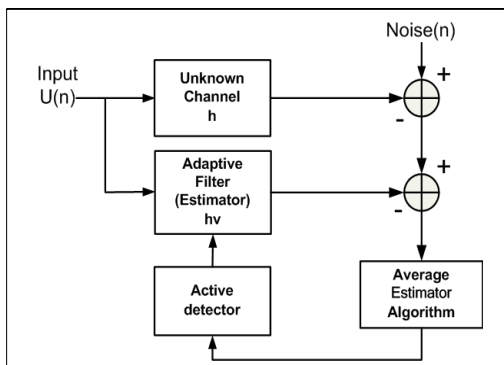
잡음 제거 필터 $ry(n, l)$ 와 최적 잡음 제거 필터 $ry_{opt}(n, l)$ 의 l 번째 필터 계수이며 ry 와 ry_{opt} 는 $N \times 1$ 벡터로 나타낸다. 잡음 제거 필터 $ry(n, l)$ 는 적응 필터로 데이터가 입력되면 자동으로 필터의 계수를 갱신한다.

LMS 알고리즘은 선형 시 불변 시스템으로 구성되며 평균값에 의해 안정성이 나타나며 평균값이 커지면 안정성이 감소하고 평균값이 작아지면 안정성은 증가하게 된다.[2] 따라서 반향 잡음 제거 평균 예측 LMS 알고리즘

은 평균값을 작게 유지하며 안정된 필터로 구성하고 식 (8)과 같이 나타낸다.

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

반향 잡음 제거 평균 예측 LMS 알고리즘은 시간 n 의 $e[n]$ 에 대하여 $L-1$ 개의 입력 값으로 인한 평균값을 계산하여 출력한다.



[그림 3] 반향 잡음 제거 평균 예측 LMS 알고리즘

그림 3에서는 제안한 반향 잡음 제거 평균 예측 LMS 알고리즘을 나타낸다.

3.2 CHMM 모델링

파라미터를 벡터 양자화 과정을 거쳐 코드북 벡터로 부호화한다. CHMM[4]을 이용해서 음성 인식을 수행하려면 미리 필요한 카테고리 수만큼 모델을 준비해 두고 미지의 입력 패턴에 대해 발생 확률을 최대로 하는 카테고리리를 인식 결과로 나타낸다. 상태 i 에서 j 로의 천이 확률 a_{ij} , 심벌 k 의 발생 확률 $b_{ij}(k)$ 를 학습하기 위한 알고리즘으로서 Baum-Welch 알고리즘을 사용한다. 상태수를 N , T 를 심벌 열 길이로 한다.

전향 확률을 $\alpha(i, t) (i = 1, 2, \dots, N; t = 0, 1, \dots, T)$, 후향 확률을 $\beta(j, t) (j = 1, 2, \dots, N; t = T, T-1, \dots, 0)$, 그리고 모델 M 이 관측열 $O = o_1, o_2, \dots, o_T$ 를 출력하는 확률을 $P(O|M)$ 를 사용해서 식(9)와 같이 나타낸다.

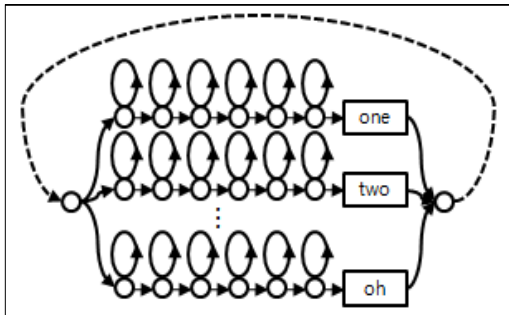
$$v_i(i, j) = \frac{\alpha(i, t-1)a_{ij}b_{ij}(o_t)\beta(j, t)}{P(O|M)} \tag{9}$$

천이 확률 a_{ij} 와 출력 확률 $b_{ij}(k)$ 의 재추정식은 식(10)과 식(11)에 각각 나타내었다.

$$\hat{a}_{ij} = \frac{\sum_t v_t(i, j)}{\sum_t \sum_j v_t(i, j)} \quad (10)$$

$$\hat{b}_{ij}(k) = \frac{\sum_{t: o_t=k} \gamma_t(i, j)}{\sum_t \gamma_t(i, j)} \quad (11)$$

$\alpha(i, t)$ 는 o_1, o_2, \dots, o_t 를 생성해서 상태 i 에 도달하는 확률이고 $\beta(j, t)$ 는 상태 j 에서 시작되는 상태 천이에 의해 $o_{t+1}, o_{t+2}, \dots, o_T$ 가 생성되는 확률이다. 파라미터들은 반복해서 재추정됨으로 국부적으로 최적인 값에 수렴한다.



[그림 4] CHMM 모델링

반향 잡음 제거 평균 예측 LMS 알고리즘을 이용하여 반향 잡음을 제거하며 반향 잡음이 제거된 깨끗한 음성 신호를 CHMM을 이용하여 인식을 위한 모델로 구성하며 그림 4에 나타내었다.

4. 실험 결과 및 분석

본 논문에서 제안한 평균 예측 LMS 알고리즘을 이용한 반향 잡음에 강인한 CHMM 모델을 구성하여 인식 성능을 실험하였다.

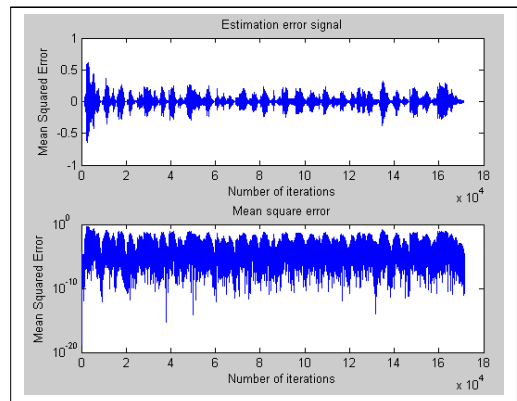
제안된 방법을 실험하기 위해 사용된 잡음은 핑크 노이즈를 반향 잡음으로 사용하였다. 5명의 화자의 임의의 음성을 각각 10개의 음성에 SNR이 5dB가 되도록 혼합하여 반향 잡음 제거 실험을 수행하였으며 이를 평가하기 위해 식 (12)를 이용하여 SNR의 향상 정도를 평가하

였다.[5]

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^N x^2(n)}{\sum_{n=1}^N [x(n) - \hat{x}(n)]^2} \quad (12)$$

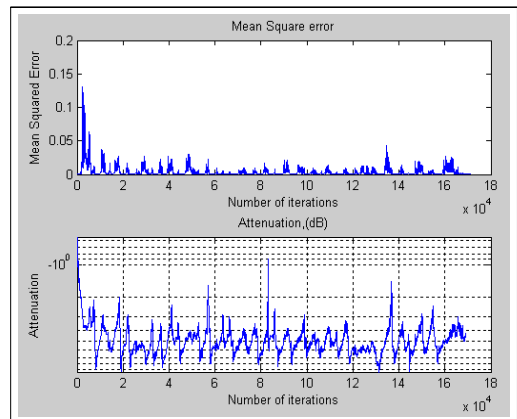
$x(n)$ 은 잡음이 혼합된 음성을 의미하고 $\hat{x}(n)$ 은 잡음이 제거된 음성을 의미하며 n 은 시간의 인덱스를 나타낸다.

그림 5에서는 잡음이 합성된 신호를 나타내었으며 입력 신호로 사용하였다.



[그림 5] 잡음이 합성된 신호

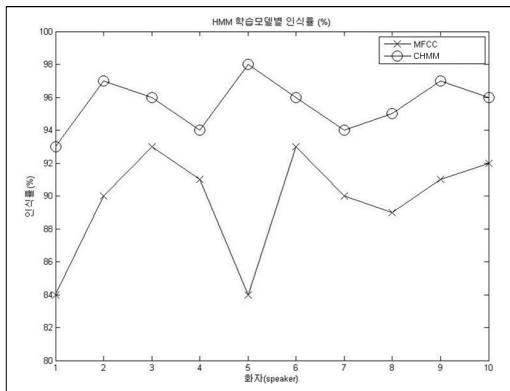
그림 6은 이동 평균 LMS 알고리즘을 수행하여 얻은 출력 신호이다. 원 음성 신호에 가깝게 잡음이 제거된 것을 확인할 수 있었으며 평균 1.93dB이 향상되었다.



[그림 6] 제안 알고리즘에 의해 잡음이 제거된 신호

또한 인식 실험을 위해서 음성 DB는 ETRI에서 제작한 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 CHMM 모델을 구성한 후 인식 성능을 평가하였다.

어휘 인식을 비교 평가를 위해 MFCC[8]로 모델을 구성한 기존 음성 인식 시스템과 본 논문에서 제안한 방법의 인식률을 평가하였다. 음성 어휘는 445DB 내에 있는 단어를 무작위로 10개의 단어를 발성하여 테스트 하였으며 CHMM으로 모델링하여 인식 실험을 수행한 결과 MFCC로 모델링하여 인식 실험한 인식률보다 평균 2.1% 향상되었으며 그림 7에 인식 성능 비교를 나타내었다



[그림 7] MFCC와 CHMM 모델링의 인식 성능 비교

5. 결론

본 논문은 반향 잡음 제거 평균 예측 LMS 알고리즘을 이용한 CHMM 모델을 구축하고 반향 잡음 제거와 인식 성능을 평가하였다.

FIR 필터를 기반으로 사용하는 LMS 필터는 복잡한 필터 설계 과정과 채널 잡음이 포함되어 적응 계수 값의 증가로 인해 적응 시간이 길어져 잡음 제거율이 저하되는 원인이 된다. 또한 연속 음성 인식인 문장을 입력하여 인식할 경우 어휘와 어휘 사이의 잡음이 포함되어 혼합된 음성으로 입력되면 인식률을 저하시키는 원인이 되며 연속 음성 인식을 위한 잡음이 제거된 모델 구축이 중요한 요인으로 작용한다. 이러한 성능 저하 현상을 평균 예측을 이용한 LMS 검출 알고리즘을 통하여 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 CHMM 학습 모델을 구성하고 인식 실험을 수행하였다. 변화하는 반향 잡음에 강인하도록 평균 예측 LMS 알고리즘을 적용하고 CHMM 모델을 구성하여 실험한 결과 음성의 SNR은 평균 1.93dB이 향상되었고 연속 음성의 인식률은 2.1% 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 적용 기법을 이용한 비선형 반향신호 제거에 관한 연구. 대한전기학회:학술대회논문집:정보및제어부문B, 882-885.
- [3] 안찬식·오상엽. (2009). MLHF 모델을 적용한 어휘 인식 탐색 최적화 시스템. 한국컴퓨터정보학회지. 14(10), 217-223.
- [4] 안찬식·오상엽. (2012). CHMM 어휘 인식에서 형상 형성 제어를 이용한 가우시안 모델 최적화. 한국디지털정책학회지. 10(7) 167-172.
- [5] 이세원·박호중. (2003). 잡음에 강인한 반향 제거기 연구. 한국음향학회지. 22(4). 261-268
- [6] 이수정·신계현·김순협. (2007). 음성강화를 위한 이동 평균 예측량 기반의 검출방법 최적화. 전자공학회논문지. 44(3), 97-104.
- [7] 황성수·윤상석·김석찬·이채동(2010). NLMS 적응 필터 기반의 음향 반향 제거기. 한국통신학회논문지. 35(4C). 343-349.
- [8] C. Y. Lee. (2010). Effect of FIR filtering and spectral tilt on speech recognition with MFCC. 한국전자통신학회논문지. 5(4) 363-371.
- [9] H. Michael. (2003). Acoustic echo cancellation digital signal processing. Bachelor of Engineering thesis, The school of electrical engineering, The university of Queensland.
- [10] J. Homer. (2000). Detection guided NLMS

estimation of sparsely parameterized channels. IEEE Transactions on Circuits and Systems U:Analog and Signal Processing. 47(12).

- [11] 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.
- [12] S. L. Gay. (2000). Acoustic Signal Processing for Telecommunication. Kluwer Academic Publishers. Dordrecht. The Netherlands.
- [13] S. Haykin. (2002). Adaptive Filter Theory. PrenticeHall. Upper Saddle River. New Jersey.
- [14] 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 컴퓨터공학과(공학석사)
- 2004년 2월: 광운대학교 대학원 컴퓨터공학과(박사수료)
- 관심분야: 음성인식, 음향 신호처리
- E-Mail: coolsahn@gmail.com

오 상 엽



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