

잡음환경에 강인한 HMM기반 화자 확인 시스템에 관한 연구

Speaker Verification System Based on HMM Robust to Noise Environments

위진우*, 강철호*
(Jinwoo Wee*, Chui-Ho Kang*)

*광운대학교 전자통신공학과
(접수일자: 2001년 7월 16일; 채택일자: 2001년 8월 18일)

화자 확인에서 화자내 변이, 잡음 환경, 그리고 학습 환경과 인식 환경의 불일치는 화자 확인 시스템이 실용화될 수 없는 가장 큰 원인이다. 본 연구에서는, 실제 환경에 강인한 화자 확인 시스템의 구현에 초점을 맞추어 음성 전처리 과정인 잡음환경에 강인한 끝점추출 알고리즘, 잡음제거 및 마이크특성 보상기법, LPC (Linear Predictive Coefficient) 켈프스트럼 가중치에 의한 화자간 변별력 향상 기법을 제안한다. 실험 결과, LPC 잔차신호 (residue)를 이용한 끝점추출 알고리즘을 사용한 경우 약 17.65% 가량의 끝점 추출 어려움을 향상시켰으며, 제안한 잡음제거 및 마이크특성 보상기법을 사용한 경우 다른 마이크 환경에서 화자 오인식율이 약 36.93% 가량 개선되었다. 또한, 제안한 LPC 켈프스트럼 가중치에 의한 화자간 변별력 향상 기법은 평균 화자 오인식율을 약 6.515% 향상시켰다.

핵심용어: 잡음 제거, 화자내 변이, 마이크특성 보상기법

투고분야: 음성처리 분야 (2.5)

Intra-speaker variation, noise environments, and mismatch between training and test conditions are the major reasons for the speaker verification system unable to use it practically. In this study, we propose robust end-point detection algorithm, noise cancelling with the microphone property compensation technique, and inter-speaker discriminate technique by weighting cepstrum for robust speaker verification system. Simulation results show that the average speaker verification rate is improved in the rate of 17.65% with proposed end-point detection algorithm using LPC residue and is improved in the rate of 36.93% with proposed noise cancelling and microphone property compensation algorithm. The proposed weighting function for discriminating inter-speaker variations also improves the average speaker verification rate in the rate of 6.515%.

Keywords: Noise cancelling, Intra-speaker variation, Microphone property compensation technique

ASK subject classification: Speech signal processing (2,5)

I. 서론

화자확인론은 발성화자의 음성에 내포된 개인성을 이용하여 다른 화자와의 변별을 통해 본인임을 확인하는 것이다. 일반적으로 이러한 개인성에는 음성을 이용하여

책임저자: 위진우 (jwwee@unitel.co.kr)
139-701 서울시 노원구 월계동 447-1
광운대학교 전자통신공학과
(전화: 02-940-5136; 팩스: 02-940-5467)

발성화자의 성도를 모델링한 LPC (Linear Predictive Coefficient)와 발성음성의 주기성을 내포하는 Pitch나 Formant 등이 있으며 후천적으로 학습된 개인성으로는 발성패턴이나 억양 등과 같은 것들이 존재한다[1]. 그러나 이러한 특징만으로는 발성화자 자신의 변이를 정확히 추적하지 못하거나 주변의 잡음과 같은 환경적인 요인에 의해 인식율이 현격히 저하된다[2]. 따라서 본 연구에서는 화자인식시스템의 성능을 저하시키는 요인을 분석하

고 이를 제거하여 화자확인 시스템의 오인식율을 줄이고, 잡음환경에서 강인한 특징 파라미터를 추출하여 실생활에 이용될 수 있도록 시뮬레이션하였다.

본 논문의 구성은 다음과 같다. II 장에서는 화자 확인 시스템에서의 성능저하 요인에 대해 간략히 설명한 뒤, III 장에서는 제안하는 잡음환경에 강인한 끝점추출 알고리즘, 잡음제거 및 마이크특성 보상기법, LPC 캡스트럼 가중치에 의한 화자간 변별력 향상 기법을 기술한다. 또한, IV 장에서는 실험 환경 및 기존의 방식과 제안한 방식의 성능 평가 결과를 보여주고 마지막으로 V 장에서는 결론 및 고찰에 대하여 기술한다.

II. 화자확인 시스템에서의 성능저하 요인

화자확인 시스템에서는 음성인식과 같이 외부적인 요인 즉, 배경잡음이나 음성입력 시스템 내에서 발생하는 잡음, 에코 등에 의해 시스템의 성능이 크게 저하되며 장단기적으로는 화자 자신의 발성패턴의 변화라든가 입력 시스템의 가변성으로 인해 인식율이 크게 저하된다[3].

이를 분류하여 살펴보면 크게 다음 세 가지 형태로 나누어 볼 수 있다.

2.1. 화자 자신의 발성특성의 변화

화자 자신의 변화는 화자 스스로가 학습시와 인증시 발성 패턴이나 감정상태, 강세, 억양 등이 변화하거나 감기가 들었을 때, 발성하는 자세가 변화할 때, 아침 또는 저녁 무렵 목이 잠길 때 등에 의한 영향을 말하는 것으로 실제 시스템의 인식을 변화에 커다란 영향을 미친다.

본 연구에서는 이러한 변화요인으로부터 강인한 화자 확인용 특성 파라미터를 추출하기 위해 화자의 발성특성에 근거한 LPC 캡스트럼의 가중치를 적용하였다.

2.2. 주변잡음과 주위 다른화자로부터의 영향

실생활에 적용가능한 화자확인 시스템은 실험실 환경과 같은 조용한 곳만이 아니다. 즉, 전시장이나 차안, 길거리 등과 같은 주변잡음이 많이 존재하는 곳이나 주위의 다른 화자들로부터의 발성음들이 함께 존재하는 상황을 고려하여야 한다. 이러한 가정을 통해 실험한 결과 대부분의 화자확인에서는 커다란 오류율을 보이게 되고 특히 이는 끝점추출과 같은 전처리에 크게 영향을 받게 된다[4].

따라서 본 연구에서는 이러한 환경을 고려하여 기존의

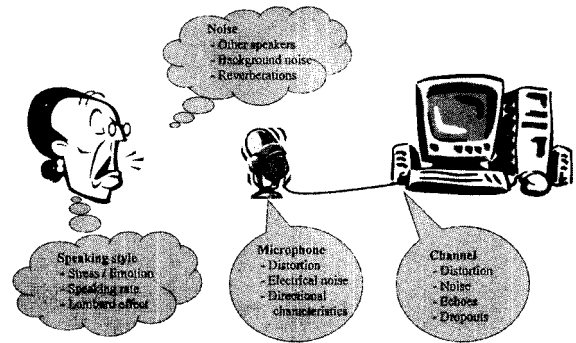


그림 1. 화자확인시스템에서의 성능저하 요인
Fig. 1. The reason of performance reduction on speaker verification system.

에너지평균과 영교차율 등을 이용한 끝점추출 알고리즘을 변형하여 새로운 끝점추출 방식을 제안하였다. 또한 잡음제거 알고리즘을 통해 보다 강인한 화자확인 시스템을 설계하였다.

2.3. 마이크 특성 및 채널 왜곡

화자확인 시스템의 경우 학습시 사용된 시스템과 동일한 시스템으로 인증시험을 할 경우 그 특성의 차이가 없기 때문에 높은 인식율을 보이게 되나 실생활에서 네트워크를 기반으로 적용하게 될 경우 이러한 예는 극히 찾아보기 힘들게 된다.

즉, 임의의 화자가 네트워크를 기반으로 임의의 시스템으로 학습하여 자신의 모델을 생성한 후 이를 인증하고자 할 때 항상 동일한 시스템으로 확인하지는 않는다. 따라서 이러한 경우 마이크와 같은 음성입력시스템이나 사운드카드와 같은 채널의 고유한 특성에 영향을 받게 되며 이로 인한 바이어스효과는 많은 오인식율을 가져오는 원인이 된다[5].

본 연구에서는 이러한 채널특성에 대한 보상을 위하여 음성합성을 통한 마이크보상 알고리즘을 제안하였으며 이를 시뮬레이션하였다.

그림 1은 화자확인시스템의 성능을 저하시키는 각종 요인을 표현한 것이다.

III. 제안하는 잡음환경에 강인한 음성 전처리 기법

3.1. 잡음환경에 강인한 끝점추출 알고리즘

일반적으로 끝점추출의 경우 프레임의 평균에너지와 영교차율을 이용하여 수행되나 배경잡음이 포함된 음성

에서는 그 정도가 심한 경우 정확한 끝점추출이 어려워진다. 이러한 끝점추출의 부정확성은 실제 시스템의 인식률을 크게 저하시킨다.

따라서 본 논문에서는 잡음환경에서의 정확한 끝점추출을 위해 LPC 잔차신호(residue)성분을 이용하여 끝점추출을 수행하였다.

LPC 잔차신호(residue)란 LPC분석에 의한 잔여성분으로 일반적으로 볼 때 예측오차를 말한다. LPC 분석의 특성상 이러한 예측오차는 음성내에 존재하는 비상관 특성이나 구간을 의미하는 것으로 음성의 단구간 분석시 통계적으로 정적이라는 가정에서 비롯된다[6].

본 논문에서는 주변잡음이나 배경잡음으로부터 강인한 끝점추출을 위해 다음과 같은 전처리과정을 거쳐 처리한 후 프레임 평균에너지와 레벨교차율을 통해 끝점추출하였다.

step 1) 원 입력음성 0번째 프레임의 LPC 캡스트럼 $\alpha_i^{(0)}$ 를 추출한다.

step 2) 원 입력음성 0번째 프레임의 LPC 캡스트럼 $\alpha_i^{(0)}$ 를 이용하여 입력음성 전 구간의 예측오차를 구한다.

$$e_s(n) = s(n) - \sum_{i=1}^p \alpha_i^{(0)} s(n-i)$$

step 3) 위 과정에서 얻은 $e_s(n)$ 을 새로운 입력음성 $\bar{s}(n)$ 으로 대체한다.

$$\bar{s}(n) = e_s(n)$$

step 4) 새로운 입력음성 $\bar{s}(n)$ 에 대한 2번째 프레임의 LPC 캡스트럼 $\alpha_i^{(2)}$ 을 구한다.

step 5) 2번째 프레임의 LPC 캡스트럼 $\alpha_i^{(2)}$ 를 이용하여 새로운 입력음성 $\bar{s}(n)$ 전 구간의 예측오차를 구한다.

$$e_{\bar{s}}(n) = \bar{s}(n) - \sum_{i=1}^p \alpha_i^{(2)} \bar{s}(n-i)$$

step 6) 위 과정에서 얻은 $e_{\bar{s}}(n)$ 을 새로운 입력음성 $\tilde{s}(n)$ 으로 대체한다.

$$\tilde{s}(n) = e_{\bar{s}}(n)$$

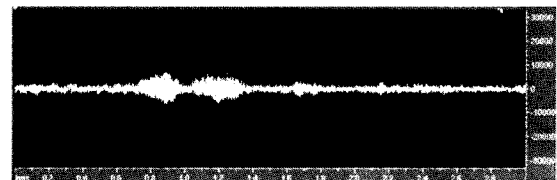
여기서, 입력된 음성의 최소 초기 3프레임은 묵음구간이라 가정하였다. 이러한 비음성구간에서 추출된 LPC 계수

를 통해 입력 음성의 전 구간을 재합성하면 비교적 정적 특성을 가진 배경잡음의 경우에는 원 파형을 예측하게 되나 음성이 존재하는 구간의 경우에는 비음성구간으로 예측된 LPC 계수에 의해 상당한 예측오차를 발생시키게 된다. 이러한 과정을 2회 반복하므로써 재구성되는 발생구간은 비음성구간에 발생하는 배경잡음이나 시스템의 전기적 특성으로 인해 발생된 잡음의 레벨을 거의 '0'으로 만들게 되고 음성구간내에서 발생된 예측오차만을 가지게 된다.

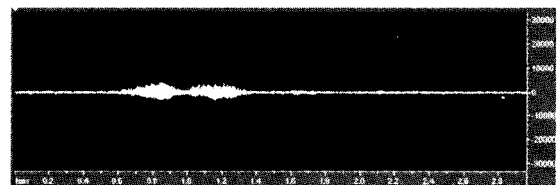
다음 그림 2에서는 잡음환경에서 입력된 발생음성의 입력파형과 정확한 끝점추출을 위해 본 논문에서 제안하고 있는 LPC 잔차신호(residue)를 이용한 변형음성의 파형을 화면캡처하여 나타내었다.

본 논문에서는 이러한 발생입력구간의 변형에 의해 배경잡음의 영향을 차감한 후 프레임 평균에너지와 영교차율을 통해 잡음환경에서 강인한 끝점추출 알고리즘을 구현하였다.

그림 3은 본 연구에서 제안한 끝점추출 방식의 구성도이다.



(a) 원 발생구간 입력 (“안녕하세요”) (a) Original input utterance (“Anyeonghaseyo”)



(b) 변형된 음성 (“안녕하세요”) (b) Manufactured speech (“Anyeonghaseyo”)

그림 2. 원 발생음성의 입력구간과 LPC residue를 이용하여 변형된 발생음성
Fig. 2. Original input utterance and manufactured speech using LPC residue.

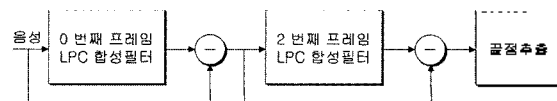


그림 3. 제안한 끝점 추출 알고리즘의 구성도
Fig. 3. Block-diagram of the proposed end-point detect algorithm.

3.2. 잡음제거 및 마이크특성 보상

본 연구에서 제안한 시스템에서는 캡스트럼 평균 차감법 (CMS: Cepstral Mean Subtraction)을 이용하여 입력 음성에 포함된 잡음을 제거하였다. 현재 잡음차감법으로 주로 사용되는 주파수 차감법과 비교할 때 캡스트럼 평균 차감법의 경우 연산량 면에서 발생하는 시스템 부하가 훨씬 적어 실시간성이 우수하고 실험결과 잡음제거시 발생하는 화자확인에서는 개인의 특성파라미터의 손실이나 왜곡이 비교적 적어 안정된 특성을 보였다[7].

다음 식 (1)은 캡스트럼 평균 차감법에 의해 추정된 캡스트럼을 나타낸 것이다.

$$\tilde{c}_n = c_n - w \cdot \bar{c}_n \quad (1)$$

여기서, \bar{c}_n 은 추정된 잡음의 평균 캡스트럼으로서 본 논문에서는 입력 전처리로부터 끝점 추출된 음성구간을 제외한 나머지 구간의 캡스트럼을 평균하였고, w 는 차감되는 잡음 캡스트럼의 가중치를 나타낸 것이다.

또한, 실제의 화자확인 시스템에서는 화자가 이용하는 제한적인 마이크의 특성으로 화자의 학습모델이 생성된다. 그러나 인식 테스트의 환경은 이와 다른 환경에서 학습시와 상이한 마이크를 이용하여 이루어질 수 있기 때문에 화자의 모델을 생성하는 학습시에는 서로 다른 시스템이나 마이크의 특성을 감안한 강인한 화자의 모델을 생성하는 것이 매우 중요한 문제가 된다.

따라서 본 연구에서는 몇가지 특성으로 대별되는 일반 탄소마이크의 특성을 이용하여 학습시 발생한 화자의 발성음을 다음 식 (2)과 같이 재구성하여 이를 토대로 화자의 특징 파라미터를 추출함으로써 마이크 특성의 변이에 강인한 화자모델을 생성하도록 하였다.

$$\overline{s(i)} = G \cdot \sum_{j=1}^n c_n(j) s(i+j) \quad (2)$$

여기서, $c_n(j)$ 는 n 번째 마이크의 목음구간에서 추출된 캡스트럼의 j 번째 차수를 의미하며 G 는 이득값이며 $\overline{s(i)}$ 는 n 번째 마이크 특성으로 합성된 음성신호를 의미한다. 즉, 마이크의 특성을 반영할 수 있는 목음구간의 캡스트럼을 이용하여 임의의 입력음성을 합성하므로써 마치 합성된 음성이 임의의 마이크로부터 입력된 음성으로 처리되는 결과를 가져오게 된다.

따라서, 본 논문에서 제안한 시스템의 전처리 특성에 대한 구성도는 다음 그림 4와 같다.

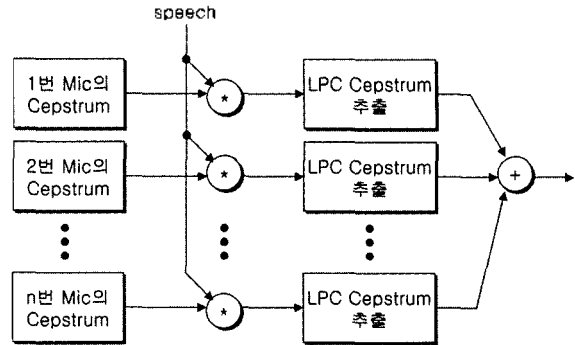


그림 4. 마이크특성 보상을 위한 입력전처리 구성
Fig. 4. Block-diagram of preprocessing for mic. characteristic compensation.

본 논문에서는 이러한 마이크보상 알고리즘을 통해 합성된 음성들과 원래의 음성을 이용하여 화자의 개인 모델을 생성하게 되고 이렇게 함으로써 시스템이나 마이크의 특성에 강인한 화자확인 모델이 생성된다.

3.3. 개선된 LPC 캡스트럼 가중치에 의한 화자간 변별력 향상

화자확인 시스템은 발성화자의 특징 파라미터와 공통 화자의 특징 파라미터간의 거리측정에 의해 승인여부를 결정하는 것이므로 무엇보다 대다수 화자들의 일반적인 특징을 담고 있는 공통화자의 특징 파라미터에 대한 특정 화자의 변별력 있는 특징 파라미터의 추출이 중요해진다. 그러나, 일반적인 성도 모델링을 이용하는 LPC 캡스트럼의 경우 Mel-캡스트럼이나 PLP와 같은 다른 특징 파라미터에 비해 비교적 화자 개개인의 개인성이 강조되기는 하지만 실험결과 많은 오수락율과 오거부율을 보이고 있다. 이것은 LPC 캡스트럼과 같은 성도 모델이 지극히 발성화자의 성도 특성에 치우쳐 있어서 학습시의 자세를 그대로 유지하지 못하거나 발성상태를 조금 달리하게 되면 본인입에도 불구하고 학습시와 많은 차이를 보이는데 그 원인이 있다. 또한, 때로 어떤 화자들간에는 그 차이가 매우 작게 나타나며 이는 오수락율의 원인이 되기도 한다.

또한, LPC 캡스트럼이나 Mel 캡스트럼의 시간적인 변화를 내포하는 동적 파라미터로써 델타 캡스트럼을 이용하게 되는데 이는 후천적인 개인성을 나타내는 습관이나 억양에 의한 화자의 발성패턴에는 적절한 성능을 보이며 본인확인시의 오거부율을 줄여주었으나 이러한 파라미터의 특성을 지나치게 강조하는 경우 성대모사에 의한 오수락율을 높여 화자확인 시스템의 성능을 저하시키는

원인이 되기도 한다.

따라서 본 논문에서는 LPC 캡스트럼과 같은 성도모델에 의한 화자개인의 특성을 내포하는 특징 파라미터에 개선된 가중치를 적용함으로써 화자간 변별력을 향상시켰다. 즉, 공통화자의 일반적인 특성에 대한 개인화자의 개인성이 두드러지게 나타나는 차수에 강조되는 형태를 가지게 함으로써 화자간의 변별력을 향상시키고 동일 화자의 변이를 적절히 수용하도록 하였다.

다음은 본 연구에서 사용된 LPC 캡스트럼의 가중치 합수를 나타낸 것이다.

$$w(i) = \frac{\sigma_s^2(i)}{\sigma_c^2(i)} \quad (3)$$

$$\begin{aligned} \sigma_s^2(i) &= \frac{1}{T} \sum_{t=0}^{T-1} (s_i(t) - m_i^{(s)})^2 \\ \sigma_c^2(i) &= \frac{1}{L} \sum_{l=0}^{L-1} (c_i(l) - m_i^{(c)})^2 \end{aligned} \quad (4)$$

식 (4)에서 나타낸 바와 같이 $\sigma_s^2(i)$ 의 경우 화자의 발생구간내의 전체 프레임에 대한 i 번째 차수에 해당하는 특징벡터의 분산을 의미하고 $\sigma_c^2(i)$ 의 경우 일반적인 화자들의 공통코드북으로부터 얻은 i 번째 차수에 해당하는 특징벡터의 분산을 의미한다.

즉, 음성의 발생구간내에서 일반적인 화자들의 LPC 캡스트럼의 차수별 변화에 대한 특정화자의 LPC 캡스트럼의 차수별 변화의 비가 곧 그 화자의 개인성을 강조하는 가중치로 작용하게 되는데 이는 발생시 나타나는 개별화자의 주파수에 대한 시간적인 변화에 민감하므로 개인별 발생특징을 잘 반영한다.

다음 그림 5는 각 화자별 특징파라미터의 가중치를 파라미터의 차수별로 나타낸 것이다.

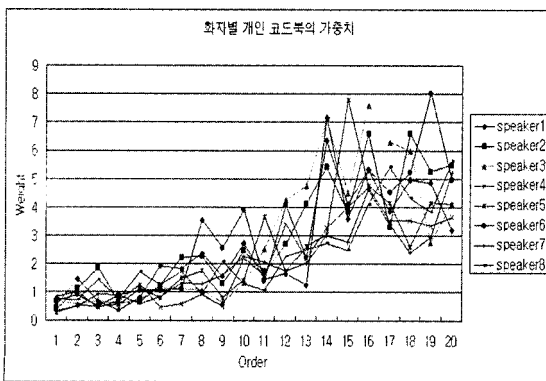


그림 5. 각 화자의 특징파라미터의 차수별 제안한 가중치
Fig. 5. The proposed personal weighting of feature parameter.

IV. 실험 환경 및 결과

본 논문에서 사용된 공통 DB는 남녀 각각 50명에 의해 각 3회씩 발생된 음성 ("안녕하세요")으로 구성되어 96차의 코드북과 world model을 생성하였다.

테스트시에는 실제환경에 적용 가능한 성능평가를 위하여 다양한 환경 (시끄러운 강의실, 조용한 연구실, 길거리 환경, PC방 등)에서 입력된 발생음성을 통해 실험되었으며 C++로 제작된 시뮬레이터를 통해 종합적으로 성능 분석 하였다.

본 논문에서는 실제환경에 적용시 사용자의 편의점을 감안하여 각 화자별 3회의 발생음성으로 화자 모델을 생성하였고, 화자적용 알고리즘을 통해 개인 코드북과 모델을 갱신하였다.

그림 6은 본 논문의 실험에서 적용된 화자확인 시스템의 전체 구성도를 나타내었고 그림 7에서는 본 연구에서 사용된 화자확인 시스템의 시뮬레이터를 화면캡쳐하였다.

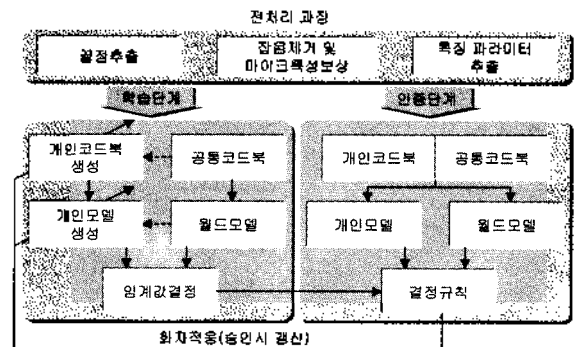


그림 6. 화자확인시스템 구성도
Fig. 6. Block-diagram of the speaker verification system.

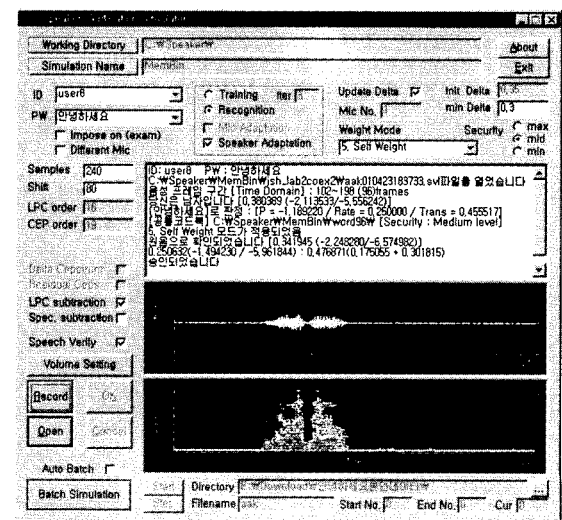


그림 7. 화자확인시스템 시뮬레이터
Fig. 7. The simulator of speaker verification system.

실험에 사용된 음성 데이터의 샘플링 주파수는 11,025 kHz이고 240포인트 (point)의 음성 샘플을 한 프레임 (frame)으로 하고 80포인트씩 이동 (shift)하여 해밍창 (Hamming Windows)을 취한 후 20차의 LPC 케스트럼을 구하였다.

학습 알고리즘으로는 5-state left-to-right HMM을 이용하였고 인식단계에서는 연산속도를 고려하여 변형 비더비 알고리즘을 적용하였다.

다음 표 1에서는 제안한 끝점 추출 기법과 일반적인 에너지와 영교차율을 사용한 경우의 끝점 추출 어려움을

비교하였다.

실험결과 끝점 추출은 일반적인 방법에서와 같이 변형하지 않고 프레임 평균에너지와 영교차율로 끝점추출 하였을 경우 약 69.11%의 정확도를 보이거나 LPC 잔차신호를 이용하여 변형된 음성구간에 대해 끝점 추출한 결과 86.76%의 정확도를 보이므로 약 17.65% 향상됨을 알 수 있다. 본 논문의 끝점추출에 사용된 340개의 음성 데이터는 시스템 잡음 및 주위 다른 사람들로부터의 잡음을 포함한 것이며 평균 신호대잡음비가 약 7.62 dB이었으며, 실험데이터들의 신호대잡음비 범위는 최저 1,863 dB, 최대 17,254 dB까지 분포하고 있다.

다음 표 2에서는 제안한 마이크 보상 기법에 대한 성능 평가를 나타내었다.

본 실험에서는 마이크 1로 학습한 후 학습한 마이크 외의 다른 3개의 마이크로 인식 실험을 한 경우 잡음제거 및 마이크 보상 기법을 사용하지 않은 경우보다 제안한

표 1. 제안한 끝점 추출 기법의 성능 평가

Table 1. Performance of the proposed end-point detection method.

	에너지와 영교차율을 사용한 끝점 추출	LPC residue를 이용한 끝점 추출
Utterance	253/340	295/340
정확도 (%)	69.11	86.76

표 2. 제안한 마이크보상 알고리즘에 대한 성능분석

Table 2. Performance analysis for the proposed mic. adaptation algorithm.

Mic 보상 알고리즘을 적용하지 않은 경우								
조건	M1M1		M1M2		M1M3		M1M4	
	FR	FA	FR	FA	FR	FA	FR	FA
계	21/150	3/300	84/150	5/300	56/150	8/300	120/150	0/300
인식율 (%)	84.0	99.0	44.0	98.33	63.7	97.33	20.0	100.0
제안한 방식의 Mic 보상 알고리즘을 적용한 경우								
조건	M1M1		M1M2		M1M3		M1M4	
	FR	FA	FR	FA	FR	FA	FR	FA
계	17/150	4/300	17/150	3/300	12/150	13/300	15/150	2/300
인식율 (%)	88.7	98.67	88.7	99.0	92.0	95.66	90.0	99.33

표 3. 제안한 가중치에 대한 화자확인 성능분석

Table 3. Performance analysis of the speaker verification using the proposed weight.

화자	No weighting		Cepstral weighting		Proposed weighting	
	FR	FA	FR	FA	FR	FA
user1	1/40	0/300	1/40	0/300	1/40	0/300
user2	40/100	3/300	10/100	3/300	12/100	4/300
user3	1/40	0/300	0/40	0/300	0/40	0/300
user4	3/60	12/300	14/60	0/300	0/60	1/300
user5	0/50	1/300	0/50	0/300	0/50	0/300
user6	9/60	0/300	8/60	9/300	6/60	1/300
user7	6/70	0/300	5/70	0/300	2/70	0/300
user8	19/100	0/300	23/100	0/300	17/100	0/300
user9	11/60	14/300	42/60	0/300	10/60	8/300
user10	2/80	1/300	2/80	0/300	1/80	0/300
계	92/660	31/3000	105/660	12/3000	49/660	14/3000
%	13.939	1.033	15.909	0.4	7.424	0.466
인식율 (%)	86.061	98.967	84.091	99.6	92.576	99.534

기법을 사용하였을 때 매우 큰 성능 향상을 보였다.

마이크 보상 알고리즘을 적용하지 않은 경우 본인 인증에 대해 평균 52.92%의 현저히 낮은 인식율과 사칭자 거부에 대해 98.66%의 정확도를 보이나 제안한 마이크 보상 알고리즘을 적용한 경우에는 본인 인증시 89.85%의 인식율과 98.16%의 사칭자 거부에 대한 정확도를 갖는다.

다음 표 3은 제안한 가중치에 대한 성능 분석표이다. 본 실험에서는 가중치를 주지 않은 경우와 음성인식에서 흔히 적용되는 사인파형태 (sinusoidal)캡스트럼 가중치를 적용한 경우에 대해 실험해 보았고 본 논문에서 제안한 발성프레임 변화에 근거한 화자별 가중치를 적용한 경우와 비교분석하였다.

표 3에서 보여주듯이 가중치를 적용하지 않은 LPC 캡스트럼의 경우 본인인증에 대한 평균인식율이 86.061%이고 사칭자에 대한 거부율이 98.967%를 나타냈고, 사인파형태 (sinusoidal)캡스트럼 가중치를 적용한 경우 본인 인증에 대한 평균 인식율이 84.091%이고 이때 사칭자에 대한 거부율이 99.6%를 보인 반면, 제안한 화자별 가중치를 적용한 경우 본인 인증율이 평균 92.58%, 사칭자 거부율이 99.53%로 향상됨을 확인하였다.

V. 결론 및 고찰

본 연구에서는 실제 환경에서 일어나는 화자 확인 오류율을 최소화하고 잡음 환경 및 화자내 변이에 강인한 화자 확인 시스템을 위한 전처리 과정을 제안하였다. 연구실 환경이 아닌 실제 환경에서 취득한 음성 데이터의 경우 화자내 변이나 잡음의 종류가 다양하였으므로 화자 확인 성능은 시간대와 환경별로 매우 좋지 않았다.

본 연구에서는 실제 환경의 화자 확인 시스템의 성능에 큰 영향을 미치는 끝점 추출 알고리즘을 개선하였고, 학습 환경과 인식 환경의 불일치로 일어나는 성능 저하를 막기 위해 잡음 제거 및 마이크 특성 보상기법을 제안하였다.

또한, 화자간의 변별력을 더욱 높이기 위하여 화자별 발성음성의 프레임 변화에 근거한 가중치 기법을 제안하였다.

실험결과 잡음환경에서 전반적인 화자 인식율의 상당한 향상을 가져왔으나 화자 자신의 변이와 사칭자에 대한 보안성을 더욱 강조한다면 화자확인 시스템과 같은 생체 인증기술이 실생활에 적용되기 위하여 보다 많은 연구가 계속되어야 할 것이다.

더 나아가 화자의 특징을 효율적으로 표현하여 화자간 변별력을 좀더 향상시킬 수 있는 화자 특징 파라미터에 대한 연구가 이루어진다면 화자확인 시스템의 실용화에 많은 도움이 되리라 생각된다.

감사의 글

이 논문은 2001년 광운대학교 교내학술연구비 지원에 의해 연구되었음.

참고 문헌

1. J. M. Naik, "Speaker verification: a tutorial," *IEEE Communication Magazine*, January 1990.
2. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE, ASSP-27*, no. 2, 1979.
3. A. E. Rosenberg and F. K. Soong, "Recent research in automatic speaker recognition," *Advances in Speech Signal Processing*, pp. 701-738, 1992.
4. R. M. Stern, F-H Liu, Y. Ohshima, T. M. Sullival and A. Acero, "Multiple approaches to robust speech recognition," *5th Darpa Workshop on Speech and Natural Language*, Arden Conf. Center, Harriman, N. Y., Feb. 1992.
5. M. Omologo, M. Malassoni, P. Svaizer and D. Giuliani, "Microphone array based speech recognition with different talker-array positions," *Proc. ICASSP97*, pp. 227-230, 1997.
6. J. He, L. Liu and G. Palm, "On the use of residual cepstrum in speech recognition," *Speech Communications*, Vol. 17(1-2), pp. 5-8, 1995.
7. 진세훈, 이재희, 강철호, "화자 확인 시스템을 위한 적응적 모델 갱신과 사전 문턱치 결정에 관한 연구," *한국음향학회지 제19권 제5호*, pp. 20-26, 7, 2000.

저자 약력

- 위진우 (Jinwoo Wee)
한국음향학회지 제20권 제5호 참조
- 강철호 (Chul-Ho Kang)
한국음향학회지 제9권 제4호 참조