

전화망을 위한 어구 종속 화자 확인 시스템

김 유 진(金 儒 辰), 정 재 호(鄭 在 皓)

인하대학교 전자공학과 DSP lab.

전화 : (032) 860-7420 / 팩스 : (032) 868-3654

Text-dependent Speaker Verification System Over Telephone Lines

Yu-Jin Kim, Jac-Ho Chung

Digital Signal Processing Lab., Dept., of Electronic Engr. Inha University

E-mail : g1982832@inhavision.inha.ac.kr, jhchung@inha.ac.kr

Abstract

In this paper, we review the conventional speaker verification algorithm and present the text-dependent speaker verification system for application over telephone lines and its result of experiments.

We apply blind-segmentation algorithm which segments speech into sub-word unit without linguistic information to the speaker verification system for training speaker model effectively with limited enrollment data. And the World-model that is created from PBW DB for score normalization is used.

The experiments are presented in implemented system using database, which were constructed to simulate field test, and are shown 3.3% EER.

I. 서론

개인의 정보 욕구가 증가하고 유무선 통신망, 인터넷과 같은 통신 인프라가 급속도로 보급되면서 다양한 망을 통한 음성 인터페이스 기술의 실용화 요구가 증가하고 있다. 특히 앞으로는 음성 인식 기술과 더불어 다양한 망을 통한 데이터나 시스템의 접근을 쉽고 정확하게 사용하기 위한 기술로서 화자 인식 기술에 대한 관심이 높아질 것으로 예상된다.

이러한 화자 인식 기술 중에서 특히 청구된 화자의 음성과 등록된 동일 화자의 성문 모델을 비교하여 본인임을 확인하는 화자 확인 기술은 개인 전자 문서의 관리로부터 전화 또는 인터넷을 통한 은행 업무 및 전자 상거래에 이르기까지 광범위하게 응용될 수 있으리라 사료된다.

특히 음성을 이용한 화자 확인 기술이 선호되는 이유는 카드, 도장, 서명 그리고 신분증 등의 물리적인 수단이 가진 도난이나 위조의 문제점이 전혀 없으며, 다른 생체 계측 수단인 지문 또는 망막을 입력받기 위해서 고가의 스캐너 장비가 필요한 반면 음성은 상대적으로 저가인 마이크 또는 유무선 전화를 통해 원

거리에서 쉽게 처리될 수 있기 때문이다.

화자 인식 기술에 대한 연구는 이미 1960년대부터 음성 인식 기술과 함께 연구되어 왔으며 1970년대 중반부터 Texas Instruments의 소규모 화자 인식 시스템과 AT&T Bell Lab.의 시스템들을 통해 실용화에 대한 연구 결과가 선보이기 시작했다. 최근에는 음성 인식 기술성과에 힘입어 실험실 결과로서 EER(Equal Error Rate)이 약 1% 미만인 기술 수준이 발표되고 있으며 최근에는 유무선 전화망을 이용한 화자 확인 기술 및 실용화 연구가 한창 진행되고 있다.[1, 2] 특히 미국의 경우 다수 고객 서비스에 응용된 Calling Card Service시스템이 미국의 Sprint 통신 회사에 의해 1995년부터 이미 선보이고 있다.

화자 확인 시스템은 사용하는 어구의 종류에 따라 어구 종속(Text dependent), 어구 독립(Text independent) 그리고 어구 종속 시스템의 변형으로서 어구 지시(Text prompted) 시스템으로 구분할 수 있다. [3]

어구 독립 시스템(TI SV) 시스템은 등록과 확인 과정에서 자연스런 발성을 통해 본인임을 확인한다. 즉, 어구의 의미와는 무관한 음성의 특징을 추출하여 화자의 성문을 생성한다. 이러한 특징으로는 피치(pitch), 톤(tone) 그리고 음색 등이 있다. TI SV 시스템의 장점은 화자가 등록에 사용된 패스워드 또는 문장 등을 기억할 필요가 없다는 점이다. 하지만 녹음에 의한 사칭자를 막을 수 없는 치명적인 단점이 있다.

어구 종속(TD SV) 시스템은 등록과 확인 과정에서 동일하게 발성된 패스워드, 카드 번호 또는 PIN(Personel Identity Number)번호와 같은 특별한 어구로서 본인임을 확인한다. 이 시스템은 화자가 패스워드를 기억해야 하는 단점이 있지만 화자가 자유롭게 선택한 패스워드의 시간적인 정보 및 스펙트럼 정보 불 이용한 성문 모델을 생성할 수 있으며 TI SV 시스템에 비해서 뛰어난 성능을 가질 수 있다.

어구 지시(TP SV) 시스템은 화자의 패스워드를 녹음한 사칭자를 막기 위해 매번 다른 패스워드를 제시하여 발생하도록 함으로써 보다 높은 확인 성능을 구

현할 수 있는 TD SV 시스템의 발전된 형태이다. 이는 미리 선정된 여러 단어 집합 중에서 임의의 단어를 요구하도록 구현되는데, 현재 2연 숫자음이 조합된 형태가 가장 일반적이다. (예, 23-48-56) TP SV 시스템은 높은 화자 확인 성능이 가능하지만 선택된 어휘의 지시를 위한 음성 합성 기술, 요구된 어휘의 올바른 발성을 검사하기 위한 음성 인식 기술 등이 조합되어야 하므로 등록 및 확인 과정의 복잡도가 증가한다는 단점을 가진다.

결론적으로 어구 독립 시스템은 보안보다는 화자 분류(Speaker Classification)등의 좀 더 완화된 보안 형태의 어플리케이션에 적용 가능하며 어구 지시 시스템은 강력한 보안이 필요한 신용 거래, 홈뱅킹 등의 분야에 적합하다. 반면 어구 종속 화자 확인 시스템은 비교적 간단한 등록, 확인 과정과 함께 비교적 정확한 확인 성능을 구현할 수 있으므로 다양한 응용 분야에 적용될 수 있다고 사료된다.

본 고에서는 기본적인 화자 확인 알고리즘을 2장에서 살펴보고 3장에서 전화망 어플리케이션에 적용하기 위해 설계된 화자 확인 알고리즘을 설명한다. 4장에서는 구현된 시스템의 성능 실험 방법 및 결과에 대해 논하고 마지막으로 5장에서 결론 및 향후 연구 방향에 대해 논한다.

II. 화자 확인 알고리즘

화자 확인 알고리즘은 일반적인 패턴 인식 시스템의 과정을 따른다. HMM을 기반으로 한 일반적인 형태의 화자 확인 알고리즘 또는 시스템은 그림1과 같이 구성될 수 있다.

2.1. 분절

화자의 음성은 패스워드로 사용될 어구뿐만 아니라 묵음 및 주변 잡음을 포함하고 있다. 정확한 패턴 모델링을 위한 패턴 구간을 찾는 것은 전체 시스템 및 알고리즘의 성능에 큰 영향을 미친다.

끝점 검출(End-point detection) 알고리즘은 음성 구간만을 전체 입력 신호에서 분절해내는 진통적인 음성 분절 알고리즘이다.[4] 주변 잡음 또는 전화선 채널의 영향에 강인한 알고리즘을 위해 현재까지도 꾸준히 연구되고 있으며 시간축이 아닌 주파수축의 정보를 이용한 알고리즘들도 제안되고 있다.[5, 6] 또한 음성 구간 검출 알고리즘은 front-end 알고리즘으로서 성능과 처리 속도의 trade-off를 선택해야 하는 어려움이 있다.

분절 알고리즘의 또 다른 목적은 패스워드로 사용될 어구 자체를 효과적으로 모델링하기 위해 사용된다. 특히 HMM을 기반으로 한 어구 지시 시스템의 경우 제한된 어휘 내에서 확인 때마다 다르게 발생된 어휘를 인식, 확인해야 하므로 하위 단위(Sub-word unit)의 HMM 모델이 반드시 필요하다. 어구 종속 시스템의 경우에도 적절한 분절 알고리즘을 통해 하위 단위 단위의 HMM 모델을 생성하는 것이 성능을 향상시킬 수 있는 것으로 나타났다.[7]

이러한 분절 알고리즘으로써 유성음과 무성음을 구

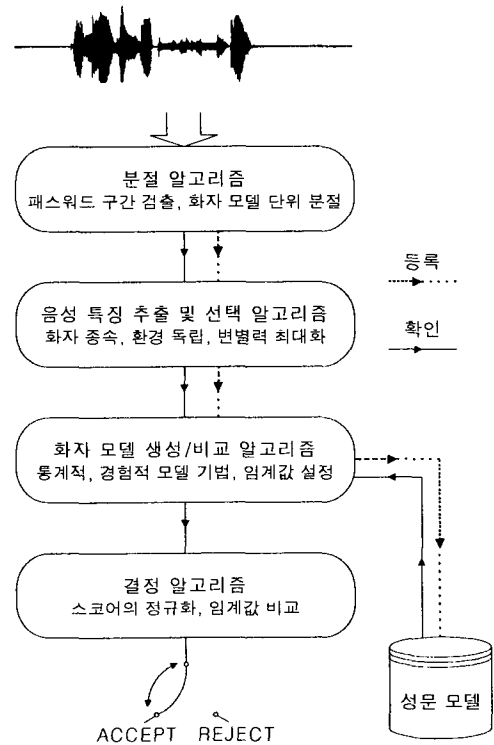


그림 1. HMM기반의 화자 확인 시스템
Figure 1. Block diagram of HMM based speaker verification system

분하여 화자의 성문 모델링에 효과적인 유성음 구간만을 선택하는 유무성음 분류 알고리즘을 예로 들 수 있다. 또한 음향학적인 전이구간을 검출하여 분절하는 알고리즘을 사용할 수도 있다. 그러나 경험적 또는 음향학적인 분류에 의한 분절 알고리즘 역시, 화자 확인 시스템의 적용 환경에 따라 그 결과 및 성능이 다르게 나타나므로 환경의 변화에 강인한 알고리즘이 되도록 설계되어야 한다.

이외에 stationary noise 환경, 전화선 채널 환경에서 효과적인 분절 또는 특징 추출을 위해 잡음 제거 및 보상 알고리즘이 추가되기도 한다.

2.2. 특징 추출 및 선택

화자 확인 알고리즘은 기본적으로 화자 고유의 발성 구조 또는 발성 습관의 특징을 모델링 함으로서 화자를 구분한다. 대표적인 화자 고유의 발성 구조 특징으로는 성도의 area function, 반사 계수 또는 pitch 등을 들 수 있다. 또한 운율과 같은 화자의 발성 습관을 화자의 고유 특징으로 사용하기도 한다.[8, 9]

특히 화자 확인 알고리즘에 HMM이 효과적으로 적용된 후로 성분 모델을 근사한 LPC 기반의 캡스트럼(LPCC)과 LPC 또는 스펙트럼을 기반으로 청각 특성을 적용한 멜 캡스트럼(MFCC)이 가장 선호되는 음성

특징이 되었다. 전화선 등의 잡음 환경에서의 화자 확인을 위한 음성 특징은 사칭자 및 타인들과 변별력이 높아야 하는 특성 외에 고유의 특징을 유지하는 강인한 특성을 가져야 한다. 기존의 LPCC, MFCC 외에 PLP(Perceptually Linear Prediction)는 스펙트럼 정보에 청각 특성을 고려한 분석을 통해 잡음 환경에서 음성 특징을 잘 나타내는 강인한 특성을 가진 것으로 알려졌다. 또한 기존의 음성 특징에서 잡음 또는 채널 효과를 제거하거나 보상하는 기법들이 연구되었다. Cepstral weighting, 특히 band-pass liftering은 잡음에 민감한 고차 영역을 상대적으로 감쇄시킴으로써 좋은 성능을 얻을 수 있다. CMS(Cepstral Mean Subtraction)는 채널 환경에서 느리게 변하는 컵스트럼 영역의 채널 왜곡 성분을 제거하는 가장 효과적인 채널 효과 정규화 방법으로 알려져 있다. 이외에 1990년 Hermansky에 의해 처음 제안된 RASTA(Relative SpecTrAl) 등의 기법은 스펙트럼을 로그 영역으로 변환한 후 필터링함으로써 채널 성분 등을 제거할 수 있다고 알려져 있다.[10]

이외에 특징 추출과 함께 추출된 특징을 분석함으로써 화자 고유의 특징을 나타내는, 즉 다른 화자와의 변별력을 유지하는 특징 구간만을 선택하거나 가중치를 부여하는 알고리즘들이 연구되기도 했다.[11]

2.3. 성분 모델 생성 / 비교

패턴 인식의 관점에서 분류기(classifier)에 해당하는 과정으로서 DTW(Dynamic Time Warping), VQ(Vector Quantization), HMM(Hidden Markov Model), ANN(Artificial Neural Network) 등의 알고리즘이 사용되며, 특징 추출/선택 과정을 거친 특징들의 시간적, 공간적 분포를 모델링하고 기존의 모델들과 입력된 음성을 비교하여 유사도를 얻는 과정이다.[11]

어구 중속 알고리즘의 경우 초기 연구 결과로서 전통적인 템플릿을 기반으로 하는 알고리즘인 DTW, VQ 등이 효과적으로 적용되었다. 이는 첫째, 화자 확인의 중요한 요소로서 사용될 수 있는 패스워드내의 시간 추이에 따른 특징 분포의 정합 능력이 탁월하기 때문이며, 둘째, 적은 등록용 발성으로 비교적 정확한 화자 모델의 생성이 가능하기 때문이다. 하지만 등록 후 추가의 발성이 기존 모델을 향상시키는 위해서는 추가의 템플릿이 필요하게 되고 이는 메모리의 증가로 나타나는 단점을 가진다. 또한 동일한 화자라도 발성 환경, 시간, 감정 상태에 따른 변화를 정규화해야 하는데, 이 과정에서 통계적인 방법으로 생성된 모델이 효과적으로 사용될 수 있다. 따라서 어구 중속 알고리즘에서도 HMM 등의 통계적인 알고리즘이 효과적으로 사용되고 있다.[12, 13]

어구 독립 알고리즘의 경우 화자의 패스워드가 임의성을 가지므로 단지 화자 어구내의 특징 분포만을 통계적인 방법으로 모델링하는 방법을 사용한다. 따라서 ANN, VQ 등이 적용되어 두드러진 성능을 나타내었고 최근에는 GMM(Gaussian Mixture Modeling)이 좋은 성능을 나타내는 것으로 나타났다.[14]

어구 중속 알고리즘의 발전된 형태인 어구 지시 알

고리즘의 경우 녹음된 음성을 통한 접근이 불가능하므로 좀 더 높은 성능을 요구하는 어플리케이션을 위해 연구되고 있다. 어구 지시 알고리즘은 하위 단어 단위의 화자 모델이 필수적이므로 HMM을 이용한 알고리즘이 선호되고 있다.[15]

2.4. 결정 규칙(Decision Rule)

화자 확인 시스템의 최종 과정은 입력된 발성과 생성된 화자 모델과의 유사도를 얻고, 이 값을 미리 결정되거나 등록 과정에서 화자에 적응적으로 결정된 문턱값과 비교하는 과정이다.

HMM을 기반으로 한 어구 중속 알고리즘에서 스코어는 화자 모델에 대한 입력 음성의 우도(likelihood)로 나타나며 그 분포는 화자의 발성 환경, 시간 경과에 따른 변화뿐만 아니라 발성된 어구에 따라서도 크게 달라진다. 따라서 화자와 사칭자를 구분하는 최적의 문턱값을 결정하기 위해 변이가 적은 우도 추정 방법을 정의하고 적용하는 것이 중요하다. 일반적으로 이러한 화자간의 우도의 변이는 pseudo-imposter 모델을 이용한 정규화 방법으로 해결하며, pseudo-imposter 모델은 cohort set을 이용하는 competition based 방법과 world model을 이용하는 qualifier based 방법으로 근사 될 수 있다. 또한 이러한 추정 방법을 절충한 형태도 연구되었다.[16, 17]

한편 계산된 우도와 비교할 문턱값을 등록과정에서 화자 중속적으로 설정함으로써 최적의 화자 확인 성능을 얻기 위한 연구도 진행되고 있다.[18]

III. 하위 단어 단위를 이용한 어구 중속 화자 확인 시스템

본 고에서는 다음과 같은 목적으로 하위 단어 단위를 사용한 어구 중속 화자 확인 시스템을 설계하였다.

첫째, 화자가 선택한 어구에서 발성 습관, 문맥 정보를 포함하는 시변(time variant)정보는, 특히 제한된 발성인 경우 모델내의 각 상태에 정확하게 할당되어 모델링되지 않기 때문이다. 이는 일반적으로 EM 알고리즘으로 이루어지는 모델 내의 상태 정합(state alignment)이 충분히 많은 데이터를 바탕으로 훈련되는 것을 가정하고 있기 때문에, 제한된 데이터인 경우 상태정합은 균일하게 또는 임의로 정해진 초기 조건에 큰 영향을 받기 때문이다. 따라서 가능한 한 초기 조건의 영향을 덜 받는 짧은 상태 전이로 표현되는 모델이 보다 효과적이다.

둘째, 화자의 패스워드 내에서 변별력이 높은 하위 단어 단위에 대해 가중치를 주는 등의 후처리를 비롯한 성능 향상을 위한 유연성을 가질 수 있다.

셋째, 보다 강력한 보안을 위한 어구 지시 시스템을 위한 기본 알고리즘으로 사용될 수 있다.

3.1. 분절 알고리즘

음성 검출을 위해 로그 에너지만을 이용해 잡음에 강하도록 설계된 끝점 검출 알고리즘을 사용하였다. 그러나 끝점 검출 알고리즘만으로는 실제 환경에서 화자의 발성에 포함되는 다양한 종류의 잡음을 효과적으

"황혜은은 2월 6일 입니다."

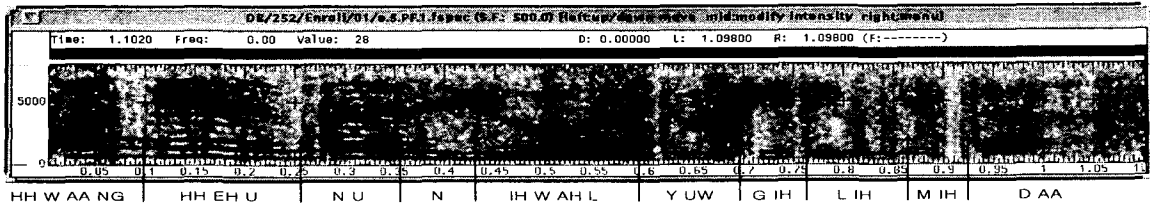


그림 2 전화선을 거친 음성에 대한 Blind-segmentation의 예[7]

Figure 2. Example of segmented speech over telephone lines by blind-segmentation[7]

로 극복할 수 없다. 따라서 자주 나타나는 잡음 모델과 전화선 환경의 묵음 모델을 생성한 후 비교 과정에서 화자 모델과의 네트워크를 통해 패스워드 구간만을 검출하는 keyword-spotting 기술을 적용하였다.

또한 하위 단어 단위의 분절을 위해 parametric filtering을 기반으로 하는 delta energy를 이용한 blind segmentation 알고리즘을 적용하였다. 이 분절 알고리즘은 이상적인 경우 사전 지식이 전혀 없이 음향학적인 변화만으로 음소 단위의 분절이 가능하다. 그러나 전화선을 거친 음성에 대해서는 음소 단위가 아닌 음절 또는 단어 단위, 즉 발성 환경에 종속적인 단위로 분절하도록 강인하게 설계되었다.(그림 2. 참고)[7]

3.2. 음성 특징

검출된 음성 구간에 대해 적용하여 관찰 벡터로서 구성한다. 관찰 벡터는 8KHz, 8bit의 μ -law형식의 음성을 10msec 마다, 20msec 길이로, 0.97 pre-emphasis 상수와 hamming 윈도우를 씌워 12차 MFCC를 추출하였다. 다시 시간 미분 계수인 Delta 계수를 덧붙여 총 24차의 음성 특징 벡터를 구성하였다.

전화망을 통한 대부분의 음성 어플리케이션에서 나타나는 채널의 불일치, 채널 환경의 영향을 최소화하기 위해 300Hz~3400Hz 대역만을 특징 추출에 사용하였다. 이외의 정보는 대부분의 채널에서 감쇄되거나 잡음에 민감하므로 정확한 정보를 얻을 수 없기 때문이다. 또한 일반적인 채널 정규화 방법인 CMS를 적용하였다.

3.3. 스코어 정규화

본 알고리즘에서는 world-model을 사용하는 스코어 정규화 방법을 적용하였는데, 이는 일반적으로 빠른 응답을 요구하는 확인 과정에서, 연산량, 계산 과정의 복잡도 그리고 저장 공간 등을 줄일 수 있기 때문이다. 결국 화자 S의 음성 패스워드 에 의해 화자 성분 모델 $M(S, W)$ 이 등록되고 청구된 화자의 음성으로부터 추출된 관찰벡터를 O라고 할 때 얻어지는 정규화된 스코어, 즉 우도율(likelihood ratio)은 world 모델 $M(S_{world}, W)$ 에 의해 다음과 같이 정의된다.[16, 17, 19]

$$R_{world} = \frac{\log[P(O|M(S, W))] - \log[P(O|M(S_{world}, W))]}{\log[P(O|M(S_{world}, W))]}$$

일반적으로 world-model의 생성을 위해 화자 확인

에 사용되는 어구에 종속적이면서 화자에는 비 종속적인 DB를 사용한다. 본 시스템과 같이 사용되는 어구가 제한되지 않은 경우, 연속 음성 인식의 음소 단위 모델 생성을 위한 DB를 사용하는 것이 타당하다고 판단하여 한국 전자 통신 연구소가 구축한 PBW(Phoneme Balanced Word) DB를 사용하였다.

VI. 실험 및 결과

4.1. 음성 DB

알고리즘의 개발 및 평가를 위해 실제 상황을 반영한 화자 확인용 DB를 구축하였다. 남자 53명 여자 53명으로 구성된 화자들이 2가지의 문장을 나누어 발성하였다. 첫번째 문장은 worst case 테스트를 위해 "범일정보통신입니다."라는 동일한 문장을, 두번째 문장은 자신의 이름과 생년월일을 조합한 "xxx는 xx월 xx일입니다."라는 문장을 발성하였다.

각 화자 당 4~6개의 등록 세션과 10회 이상의 테스트 세션으로 나누어 등록 세션은 세션 당 6회의 발성 테스트 세션은 세션 당 4회의 발성을 하였다. 음성 수집은 약 3개월에 걸쳐 오전, 오후 그리고 저녁 시간으로 나누어 1회의 통화 당 1개 세션만을 수집하였다. 따라서 모든 세션은 모두 다른 시간과 다른 채널 상태를 나타낸다.

4.2. 등록

등록은 3회 이상의 일치된 발성을 통해 이루어졌다. 실제 어플리케이션에 적용될 수 있도록 끝점 검출 알고리즘에서 발성에 문제가 있다고 판단되는 경우 등록을 거부했다. 또한 DTW 알고리즘을 적용하여 각 발성간의 유사도를 검사함으로써 심한 잡음으로 왜곡되거나 끝점 검출의 오류로 인해 발성들이 일치하지 않는 경우 등록을 거부하거나 발성을 추가로 요구하였다.

	총 화자	등록 실패	등록율
enroll 1	106	11	90%
enroll 2	103	5	95%
enroll 3	89	11	88%
enroll 4	69	5	93%

4.3. 확인

등록 과정과 동일한 과정을 거쳐 입력된 음성의 특징 벡터를 얻고 각각 화자 모델과 world 모델에 대한

우도를 계산한 후 두 모델의 우도로 계산된 스코어를 등록 과정에서 결정된 문턱값과 비교하여 이루어진다. 평가는 본인의 테스트 세션 내 모든 발성을 테스트한 FR 테스트와 자신 외의 모든 화자의 임의의 테스트 세션의 임의의 발성으로 구성된 발성 리스트에 대한 FA 테스트 수행되었다.

	총 시도	확인 거부	거부율
enroll 1	3288	97	3%
enroll 2	3376	99	3%
enroll 3	2904	81	3%
enroll 4	2480	54	2%

4.4. 실험 결과

화자 당 4개의 성문 모델을 각각 약 100여명의 발성을 임의로 선택한 리스트 3회를 이용한 실험결과 전체 평균 약 3.30%의 EER을 나타내었다.

	test 1	test 2	test 3	
enroll 1	3.66	3.90	3.67	3.74
enroll 2	3.30	3.44	3.27	3.33
enroll 3	3.13	3.48	3.04	3.22
enroll 4	2.86	3.02	2.90	2.93
평균	3.24	3.46	3.22	3.30

특히 하위 단어 단위 모델링은 화자의 패스워드 단위 위의 whole-word 모델에 비해 시뮬레이션에서는 약 2%, 실제 알고리즘이 구현된 시스템에서는 1% 이상의 EER향상을 가져왔다.

V. 결론

본 고에서는 전화망에 적용될 수 있도록 HMM을 기반으로 하는 하위 단어 단위를 사용한 알고리즘을 제안하고 그 성능을 제시하였다.

하위 단어 단위의 화자 모델을 생성하기 위해 발성에 대한 사전 지식 없이 분절할 수 있는, parametric filtering 을 기반으로 하는 delta energy를 이용한 blind segmentation 알고리즘을 적용하였다. 또한 스코어의 정규화를 위해 PBW DB를 이용하여 생성한 world 모델을 사용하였다. 실제 상황에 가까운 전화선 DB를 수집하여 테스트한 결과 3.3%의 EER을 나타내었다.

앞으로 화자 독립 음성 인식기를 적용한 분절 기술과 실제 전화선 발성 환경과 가까운 DB를 이용한 world 모델을 이용함으로써 성능을 향상시킬 계획이다. 또한 최적의 성능을 얻기 위해 화자 중속 문턱값 설정 방법에 대한 연구가 필요하다고 사료된다. 끝으로 보다 강력한 보안 절차를 위해 어구 지시 시스템 개발을 병행할 예정이다.

VI. 참고문헌

[1] AARON E. ROSENBERG, "Automatic Speaker Verification: A Review," Proceedings of the IEEE, vol. 64, No. 4, pp. 475-487, April 1976.
 [2] Cedric JABOULET, Johan KOOLWAAIJ, Johan LINDBERG, Jean-Benoit PIERROT, Frederic BIMBOT, "The CAVE-WP4 Generic Speaker Verification System,"

Proc. RLA2C, Speaker Recognition and its Commercial and Forensic Application, pp.202-205, 1998
 [3] K.-U. MAZEL and H.-P. FREI, "CAVE - Speaker Verification in Banking and Telecommunications," Computer Science Research at UbiLab, pp.153-162, Nov. 1996
 [4] L. R. RABINER, R. W. SCHAFER, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood cliffs, N.J., 1978
 [5] Seiichi Yamamoto, Masaki Naito and Shingo Kuroiwa, "Robust Speech Detection Method for Speech Recognition System for Telecommunication Networks and Its Field Trial," Proc. of Eurospeech97, pp. 1535-1538, 1997
 [6] 석종원, 배진성, "Wavelet 변환을 이용한 잡음음성의 끝점 검출," 제 9회 신호처리합동학술대회 논문집 제 9권 1호, pp.69-72, 1996
 [7] 김지운, 김유진, 정재호, "화자확인 시스템을 위한 Blind Segmentation 알고리즘," 제 12회 신호처리 합동 학술 대회 논문집 제 12권 1호, pp.363-366, 1999
 [8] Jayant M. Naik, "Speaker Verification: A Tutorial," IEEE Communications Magazine, pp.42-48, Jan., 1990
 [9] Bishnu S. Atal, "Automatic Recognition of Speakers from Their Voices," Proceedings of the IEEE, Vol.64, No.4, pp.460-474. April, 1976
 [10] Richard J. Mammone, Xiaoyu Zhang, Ravi P. Ramachandran, "Robust Speaker Recognition : A Feature-based Approach," IEEE Signal Processing Magazine, pp.58-71, Sep., 1996
 [11] Joseph P. Campbell, Jr., "Speaker Recognition: A Tutorial," Proceedings of The IEEE, Vol. 85, No.9, pp.1436-1462, Sep., 1997
 [12] S. Euler, R. Langlitz, J. Zinke, "Comparison of Whole Word and Subword Modeling Techniques for Speaker Verification With Limited Training Data," Proc. ICASSP-97, pp.1079-1082, 1997
 [13] Chin-Hui Lee, Frank K. Soong, Kuldeep K. Paliwal, *Automatic Speech and Speaker Recognition: Advanced Topics*, Kluwer Academic Publishers, 1996
 [14] Douglas A. Reynolds, Richard C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models," IEEE Transactions on Speech and Audio Processing, Vol.3, No.1, pp.72-83, Jan., 1995
 [15] Tomoko Matsui, Sadaaki Furui, "Concatenated Phoneme Models for Text-variable," Proc. of ICASSP, pp. II-391-394, 1993
 [16] F. Bimbot, H.-P. Hutter, C. Jaboulet, J. Koolwaaaj, J. Lindberg, J.-B. Pierrot, "An Overview Of The Project Research Activities In Speaker Verification," Proc. RLA2C-1998, pp.215-220, 1998
 [17] Yong Gu and Trevor Thomas, "A Hybrid Score Measurement For HMM-Based Speaker Verification," Proc. of ICASSP, pp.317-320, 1999
 [18] J. B. Pierrot, J. Lindberg, J. Koolwaaaj, H.-P. Hutter, D. Genoud, M. Blomberg, F. Bimbot, "A Comparison Of A Priori Threshold Setting Procedures For Speaker Verification In The CAVE Project," Proc. of ICASSP, pp.11-14, 1998
 [19] Sunil K. Gupta and Frank K. Soong, "Improved Utterance Rejection Using Length Dependent Thresholds," Proc. of ICSLP, 1998