Journal of Korea Multimedia Society Vol. 11, No. 11, November 2008(pp. 1471-1479)

가변 신뢰도 문턱치를 사용한 미등록어 거절 알고리즘에 대한 연구

방기덕[†], 강철호[††]

요 약

본 논문에서는 음성인식 분야에서 많이 사용되고 있는 가변어휘 단어 인식 시스템에서 미등록어에 대한 거절 성능을 향상시키는 방법을 제안한다. 거절 기능을 구현하는 방식은 핵심어 검출(keyword spotting)방식과 발화검증(utterance verification)으로 구분이 된다. 발화 검증 방식은 각 음소마다 이와 유사한 반음소 모델(anti-phoneme model)을 생성한 후 정상적인 음소 모델과 반음소 모델의 유사도를 비교하여 결정하는 방식이다. 본 눈문에서는 화자가 발성할 때마다 구해지는 화자확인 확률값을 신뢰도 문턱치를 결정할 때 적용하는 방법에 대하여 제안하였다. 제안한 방법을 사용하였을 때, 사무실 환경에서 CA(Correctly Accepted for keyword)가 94.23%, CR(Correctly Rejected for out-of-vocabulary)이 95.11%로 나타났고, 잡음 환경에서는 CA가 91.14%, CR이 92.74%로 나타나서 성능이 향상됨을 확인할 수 있었다.

A Study on Out-of-Vocabulary Rejection Algorithms using Variable Confidence Thresholds

Ki-Duck Bhang[†], Chul-Ho Kang[††]

ABSTRACT

In this paper, we propose a technique to improve Out-Of-Vocabulary(OOV) rejection algorithms in variable vocabulary recognition system which is much used in ASR(Automatic Speech Recognition). The rejection system can be classified into two categories by their implementation method, keyword spotting method and utterance verification method. The utterance verification method uses the likelihood ratio of each phoneme Viterbi score relative to anti-phoneme score for deciding OOV. In this paper, we add speaker verification system before utterance verification and calculate an speaker verification probability. The obtained speaker verification probability is applied for determining the proposed variable-confidence threshold. Using the proposed method, we achieve the significant performance improvement; CA(Correctly Accepted for keyword) 94.23%, CR(Correctly Rejected for out-of-vocabulary) 95.11% in office environment, and CA 91.14%, CR 92.74% in noisy environment.

Key words: Out-of-vocabulary rejection(미등록어 거절), variable confidence threshold(가변 신뢰도 문턱치)

1. 서 론

사람과 기계 상호간에 가장 편리한 인터페이스는 물리적인 접촉이 없이 의사전달이 가능한 음성이며 많은 곳에서 보다 나은 성능을 위한 연구가 진행되고 있다[1]. 현재 한국의 음성인식 시장은 홈오토메이

※ 교신저자(Corresponding Author) : 방기덕, 주소 : 서울시 노원구 월계동 447-1(139-701), 전화 : 02)940-5136, FAX : 02)940-5136, E-mail : carrot1110@hotmail.com
접수일 : 2008년 6월 13일, 완료일 : 2008년 8월 27일
[†] 광운대학교 전자통신공학과

[††] 정회원, 광운대학교 전자통신공학과 정교수
(E-mail : chkang5136@kw.ac.kr)
※ 본 논문은 2006년도 광운대학교 교내학술연구비 지원에 의해 수행되었습니다.

션, 네비게이션, 소형 로보트 등을 통하여 시장이 형성되어 지고 있으며 제품들이 출시되어 지고 있다. 특히, 교통 안전과 직격되는 차량용 네비게이션 제품에 음성인식 기능에 대한 요구가 많았으며 최근에 음성인식기능이 탑재된 기기들이 출시되고 있다. 이러한 음성인식시스템 중에서 가변어휘 인식 시스템이 많이 사용 중인데 인식기에 등록이 되지 않은 단어를 발성하게 되면 처리할 수 없다는 단점을 가지고 있다. 따라서 사용자는 미리 정해진 등록어를 알고 있어야 하고 등록어 만을 사용해야하는 문제가 있었다. 이런 문제점을 극복하는 방법으로 인식대상으로 등록된 단어에 대해서만 인식을 하고 그 외의 단어에 대해서는 인식을 거절(rejection)하여 시스템의 인식성능을 향상시키는 미등록어 거절(out-of-vocabulary rejection)기능이 개발되어 성능향상을 위한 연구가 진행되고 있다.

미등록어 거절방식은 구현 방식에 따라 발화검증(utterance verification)방식과 핵심어 검출(keyword spotting) 방식으로 구분할 수 있다. 우선 핵심어 검출 방식은 문법을 설계할 때 핵심어만 고려하고 이외의 단어는 가비지(garbage)모델을 사용하여 불필요한 단어를 제거하는 방법이며, 이 방법은 가비지 모델의 우도비값이 인식대상 핵심어의 우도비값보다 클 경우 제거하는 방법이다[2]. 발화검증방식은 인식결과를 확인하는 과정이 추가되며 이때 필러(filler) 모델을 이용하는 방법이 사용되었다. 하지만, 필러모델은 그 구성방식이 단어기반이므로 가변어휘 단어인식시스템을 위한 발화 검증 구현을 위해서는 매 음소단위의 검증기능이 있어야 하는데 이를 위해서 반음소 모델(anti-phoneme model)을 사용하는 방식이 제안되었다[2]. 가변어휘 단어 인식기에서 인식된 단어의 등록여부를 판별하는 것이 발화검증의 역할이다. 일반적으로 유사도 비를 사용한 테스트를 많이 사용하는데, 입력단어가 등록어라고 가정하는 영가설(Null Hypothesis)과 미등록어라고 가정하는 대립가설(Alternative Hypothesis)의 비를 이용하는 통계적인 가설 테스트를 음성인식의 많은 분야에서 사용하고 있다[2,3].

가변어휘 인식 시스템은 인식 대상 어휘가 바뀌어도 인식할 수 있는 시스템으로 인식 대상 어휘가 추가되어도 훈련과정을 새로 거치지 않고 기존의 훈련된 정보를 바탕으로 인식하는 시스템이다. 만일, 인식대상이 바뀌게 되면 인식대상이 되는 변경된 단어에 대한 PLU 단위의 정보는 미리 모델링된 상태이므로 단어 단위의 인식 결과 만들어 주는 과정만 변경해 주면 된다. 따라서, 추가적인 음성 훈련이 없이도 단어독립 음성인식이 가능하게 된다[4].

본 논문에서는 발화검증 단계에서 미등록어의 거절기능을 향상시킬 수 있는 방법을 제안하였다. 우선 한국어에 존재하는 모든 음소를 다양한 환경에서 모델링해야 한다. 또 이런 다양성을 잘 수용할 수 있는 음소 모델 구조를 만들어야 한다. 본 논문에서는 이러한 가변어휘 단어 인식기의 요구사항을 충족시키기 위한 방법으로 네비게이션 기기나 홈 네트워크 시스템 등을 호출할 때 사용하는 호출 키워드(call_keyword)에 대하여 화자확인 방식을 적용하는 방법을 제안하였다. 인식된 호출키워드에서 유도된 가중치를 사용하여 인식대상이 되는 단어들을 발화검증하는 방법을 제안하였으며 실험결과 기존의 등록된 단어 외에 새로운 단어가 추가되어도 패널티 조정 등의 추가적인 변화가 없이 거절율과 인식률이 사무실환경이나 잡음 환경에서도 모두 개선됨을 확인하였다. 또, 화자확인시 등록자에 대하여 적용되는 가중치로 인해 시스템의 인식성능이 잡음환경 하에서도 증가하였는데, 이는 등록된 화자에 대해서는 새로운 환경에 대한 모델링의 요구가 줄어들기 때문에 사용 환경이 계속적으로 변하는 자동차 환경에서는 더 큰 장점이 될 수 있다.

2장에서는 기존의 발화검증 시스템을 사용한 가변어휘 단어인식시스템에 대하여 설명하였고, 3장에서는 본 논문에서 제안한 가변 신뢰도 문턱치를 사용한 가변어휘 단어인식 시스템에 대하여, 4장에서는 실험 방법과 결과에 대하여 각각 기술하였다.

2. 발화검증시스템을 사용한 가변어휘 단어인식시스템

2.1 기본적인 시스템의 구성

기본적인 시스템은 음성인식 기능과 검증기능이 동시에 검색이 되도록 하는 One-pass 시스템과 인식기의 후처리 방식으로 검증기능을 구현하는 Two-pass방식이 있다, Two-pass 방식은 기존 시스템의 수정없이 검증 과정을 추가한 것으로 구현이 쉽다는 장점을 가지고 있다[5]. 발화 검증 시스템을

설계할 때 첫째, 미등록어와 잘못 인식된 단어를 잘 선별할 수 있는 검증 모델에 기반한 적정한 신뢰도(confidence measure)를 정의해야 하고, 둘째 훈련 데이터에서 검증 오류를 최소화할 수 있도록 검증모델을 적응시키는 훈련과정을 선택해야 하며, 셋째 유사도의 변화와 검증 문턱치의 변화, 훈련과 테스트 상태의 변화에 강해야 한다[6].

그림 1에서는 인식과 검증으로 구성된 2단계 시스템의 기본 구조를 보여주고 있다. 1단계에서 인식 모델을 사용해서 비터비(viterbi) 탐색 알고리즘에 의한 인식과정을 수행한다. 음소 모델들은 ML (Maximum Likelihood)를 이용하여 HMM의 파라미터를 최적화시켰다. 인식 과정 동안 각 단어의 발화는 음소 가설로 분할되며, 그 결과를 발화 검증 시스템으로 전달한다. 두 번째 단계인 발화 검증 과정은 인식된 후보 단어의 음소열에 대해 반음소 모델과의 신뢰도를 구해 그 단어의 신뢰도 값을 결정한다. 이 신뢰도 값이 미리 정해둔 문턱치보다 크면 인식단어로 인식이 되고 아니면 거절된다[4,6].

2.2 기존의 발화 검증 방법[7]

핵심어 검출방법은 등록어 모델인 핵심어 모델과 미등록어 모델인 필러모델을 사용하는 연결단어 인식 알고리즘을 기반으로 하며, 기존의 발화검증방법은 음소 필러 모델을 사용한 핵심어 검출방식(keyword spotting method)을 이용하여 미등록어를 거절시키는 방식이다. 여기서 필러모델들은 핵심어에 해당하지 않는 음성구간들, 즉 비핵심어들과 묵음과 배경잡음 등의 비 음성구간들을 표현하는데 사용된다. 기존의 미등록어 거절 방법은 그림 2와 같이 가변어휘단어 인식 시스템에서 비터비 탐색시 사용되는 네트워크 망을 구성한다. 네트워크 망에서 인식된 결과는 등록어들과 음소들의 열로 나타난다. "묵음+(등록어 및 음소들의 열)+묵음"과 같은 형태가 된다. 만일 단어 패널티를 잘 조정하였을 때, 입력된 음성이 등록어면 인식된 결과는 "묵음+(등록어 및 약간의 음소들의 열)+묵음"으로 나타나게 되고, 미등록어면 인식된 결과는 "묵음+(등록어및 다수의 음소들의 열)+묵음" 또는 "묵음+(다수의 음소들의 열)+묵음"으로 나타나게 된다. 이 인식된 결과를 발화검증 시스템으로 넘긴다. 발화검증시스템에서는 가변어휘 단어인식시스템의 단어 패널티와 인식된 결과의 삽입된 음소들의 개수를 이용하여 미등록어를 거절시킬 수 있다. 삽입된 음소들은 필러모델을 뜻하며, 삽입된 음소가 많다는 것은 인식 결과에 핵심어가 없다는 의미가 된다. 즉, 사용자가 미등록어를 발성하게 되면, 필러모델로 인식됨을 알 수 있다. 또한, 삽입된 음소가 미리 정해둔 문턱치 이하라도 인식결과에 등록어가 포함되어 있지 않거나 2개 이상이면, 무조건 거절시킨다.

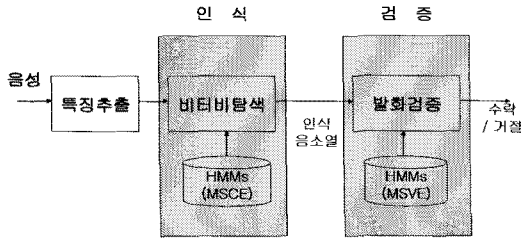


그림 1. 발화 검증을 가지는 가변어휘 단어인식시스템

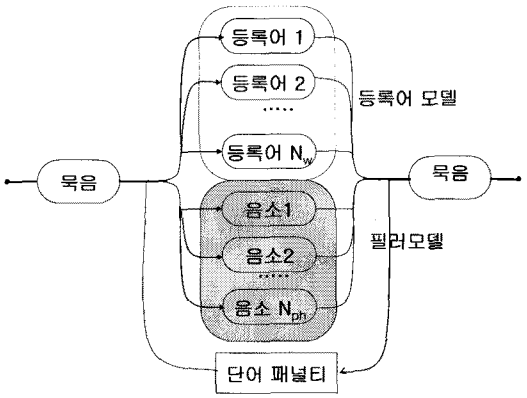


그림 2. 기존의 가변어휘 단어인식 시스템에서의 네트워크

반음소 모델은 자기 음소를 제외한 유사 음소 집합을 말하는데 일반적으로 유사 음소 집합이 많을수록 반음소가 잘 모델링되지만, 유사 음소 집합의 크기가 너무 크게 되면 훈련 데이터량이 너무 많아지는 단점이 있다[4]. 또한 인식 대상 어휘의 목록이 추가되거나 변경되면 시스템이 최적의 성능을 내도록 단어 패널티를 조절해야 하고 미등록어 거절을 위한 네트워크를 새로 구성해야하는 단점이 있는 것으로 알려져 있다.

2.3 단어 단위의 신뢰도

가변어휘 단어 인식기를 이용하여 비터비 탐색을

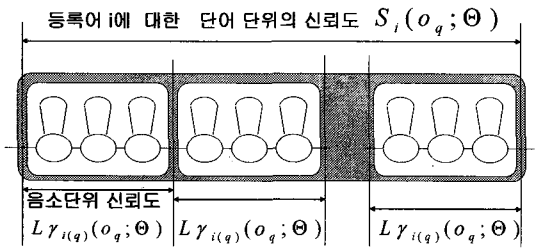


그림 3. 음소 및 단어 단위의 신뢰도 측정

하기 때문에 기본적으로 단어 단위로 인식이 되지만, 그 인식된 단어는 내부적으로 음소 단위로 인식된다. 따라서, 그림 3처럼 인식된 음소 단위들을 각각의 반음소 모델과 비교하여 신뢰도를 구하고, 음소 단위의 신뢰도를 단어 단위의 신뢰도로 환산하기 위해서 음소 단위의 신뢰도 평균을 내었다. 본 논문에서도 이와같은 방법을 사용하였다.

서로 다른 패턴들, $\Theta=\Theta_1,\ldots,\Theta_l,\ldots,\Theta_{43}$에 상응하는 발화검증모델을 사용하는 신뢰도를 선택했다. 각 패턴 l에 대하여, 음소 모델을 $\Theta_l^{(k)}$라 표시하고, anti-model인 반음소 모델을 $\Theta_l^{(a)}$, 필러(filler) 모델을 $\Theta_l^{(f)}$라 표시했다. 즉, $\Theta_l=\{\Theta_l^{(k)},\Theta_l^{(a)},\Theta_l^{(f)}\}$이다.

등록어 i는 $N(i)$원소들로 구성되어 졌으며, 음소 단위들을 평균낸 단어 단위의 신뢰도는

$$s_i(O;\Theta)=\log\{\frac{1}{N(i)}\sum_{q=1}^{N(i)}\exp\{k\cdot\omega_{i(q)}\cdot L_{i(q)}(O_q;\Theta)\}\}^{\frac{1}{k}}<\tau_s \quad (1)$$

가 되며[4], 이 신뢰도가 미리 정해둔 문턱치 τ_s 이하라면 $N(i)$원소는 거절 시키게 된다.

여기서 k는 음의 값을 가지는 상수이며, $w_{i(q)}$는 등록패턴 $i(q)$의 q^h 모델에 대한 가중치이며, O_q는 $w_{i(q)}$에 상응하는 음성의 세그먼트이다. 각 음소의 반음소 모델과의 유사도비거리, $L_{i(q)}(O_q;\Theta)$는 아래 식(2)와 같이 정의되어졌다.

$$L_{i(q)}(O_q;\Theta)=g_{i(q)}(O_q)-G_{i(q)}(O_q) \quad (2)$$

패턴 l(즉, $l=i(q)$)인 일반적인 음소에 대하여

$$g_l(O_q)=\log[p(O_q|\Theta_l^{(k)})^{\frac{1}{k}}], \quad (3)$$

$$G_l(O_q)=\log[\frac{1}{2}\cdot p(O_q|\Theta_l^{(a)})^{\frac{1}{T_l}}+\frac{1}{2}\cdot p(O_q|\Theta_l^{(f)})^{\frac{1}{T_l}}] \quad (4)$$

수식(1)의 신뢰도 측정은 키워드와 미등록어 간의 더 나은 식별력을 보일뿐만 아니라 음성인식에서 근소한 오류(near-misses) 의 검출 능력이 향상되었음을 보여준다[7,8].

이론상 발화검증에서 등록어로 분류가 될 때 신뢰도 $s_i(O;\Theta)$가 문턱치 τ_s보다 크며, 미등록어로 분류될 때는 문턱치 τ_s보다 작다.

이 실험을 성공적으로 수행하기 위해서 검증모델 Θ는 미등록어에 대한 잘못된 인식을 최소화하고 등록어에 대한 인식을 최대화할 수 있는 방향으로 훈련되어져야 한다. 그러나, 기존의 시스템이 특정 환경에 성능이 최적화되었을지라도 다양한 배경잡음에 노출되게 되면 미리 정해졌던 문턱치 등을 새로 적용해야 하는 문제점이 있다. 본 논문에서는 이러한 문제점을 위해 다음과 같은 방법을 제안하였다.

3. 제안한 가변 신뢰도 문턱치를 사용한 미등록어 거절

기존의 발화검증 시스템에서는 반음소 모델과 음소모델의 차이로써 신뢰도를 계산하고 사전에 미리 정해놓은 신뢰도의 문턱치에 따라 등록어인지 미등록어인지 구별하게 된다. 그러나, 이 방법은 음소마다 신뢰도의 분포가 다르기 때문에 단어마다 신뢰도의 분포가 다르고 단어마다 거절 성능이 균일하지 않는 문제점과 또 실제 환경에서는 잡음이 많이 추가되어지면 잡음에 따른 신뢰도의 분포 또한 달라진다[8]. 잡음환경에서의 가변어휘 단어 인식 시스템의 거절기능의 향상을 위해서 화자확인 시스템을 이용한 신뢰도를 결정하는 새로운 방법을 제안하였고 그림 4와 같이 나타내었다.

기존의 가변어휘 단어인식 시스템과 발화검증 시스템사이에 호출 키워드(call_keyword)에 대한 화자확인 시스템과 제안한 가변 신뢰도 문턱치 계산 부분을 추가한 구조이다. 음성입력이 들어오게 되면 먼저

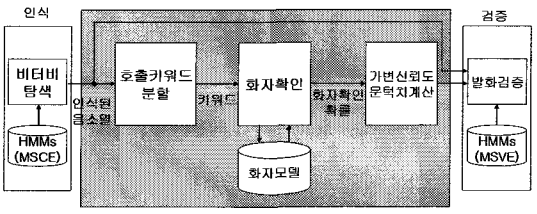


그림 4. 호출 키워드에 대한 화자확인 시스템

전처리 과정을 거치게 된다. 그 다음 가변어휘 단어 인식기를 통과한 입력 음성 파라메타에 대하여 호출 키워드 부분만을 추출한다. 추출된 호출키워드 부분에 대하여 화자확인을 실시하며 호출 키워드에 대한 화자모델을 구성한다. 이 화자모델은 인증된 화자의 새로운 호출이 있을 때마다 적응을 하게 된다. 화자확인 시스템에서 계산된 화자모델의 확률값을 이용하여 발화검증 시스템의 신뢰도 문턱치를 가변적으로 결정하게 된다. 발화 검증 시스템에서는 인식 시스템을 거친 인식 음소열과 제안한 화자확인 시스템을 거쳐 구한 가변 신뢰도 문턱치를 이용하여 발화검증을 수행하여 수락(accept)과 거절(reject)을 결정하게 된다. 그 방법은 다음 절에서 서술하였다.

3.1 화자 모델 생성 알고리즘

화자가 초기에 호출 키워드를 5회 발성하면 DHMM(Discrete Hidden Markov Model)기반 화자 모델이 생성되고, 추후 인증된 화자에 대해 화자의 변화에 적응하기 위해서 또한, 화자 모델을 좀 더 강건하게 만들기 위해서 화자 적응 기법이 사용된다. 본 논문에서는 MAP 기법에 기초하여 화자의 변화에 적응하는 적응 기법을 사용하였다[9,10].

MAP(Maximum a Posteriori) 적응 기법은 학습 데이터에 포함되어 있는 선 지식 정보를 선 밀도 함수에 포함시켜 이를 적응 데이터와 최적의 방법으로 결합하여 적응하는 기법이다. MAP에서는 파라미터 λ가 어떤 분포를 갖는 랜덤 변수라 가정한다. 만약 λ가 상위모수(Hyperparameter) Φ을 갖는 선 확률밀도함수 $g(\lambda|\Phi)$와 유사도 $f(X|\Phi)$를 갖는 관측 열로부터 추정된다면 MAP 기법은 다음과 같이 λ의 posterior model로 정의된다.

선 밀도 함수 $g(\lambda|\Phi)$는 관측열이 주어지기 전에 관심 있는 파라미터에 대한 통계적 특성을 포함하여 파라미터가 어떤 제약된 값을 갖도록 한다. 일반적으로 HMM과 같이 상태와 혼합 성분이 내재된 은닉 과정을 포함하는 경우에 MAP 추정은 매우 어렵다. 그러나 HMM 파라미터의 선 밀도 함수가 완전데이터 밀도의 공액족(conjugate family)에 속한다면 EM 알고리즘에 의해 MAP추정을 쉽게 할 수 있다. MAP 추정은 ML(Maximum Likelihood)에 비해 적은 적응 데이터에 대해 더 강인하게 파라미터를 추정한다. 적응 데이터의 양이 증가함에 따라 MAP은 ML 추정치로 수렴하는 장점을 가지고 있다. 그러나 MAP은 관측된 파라미터에 대해서만 적응된다. 그러므로 수백만 개의 파라미터를 갖는 대용량 인식기의 경우 적응 속도가 매우 느린 단점을 갖는다.

하지만, 추정해야 할 파라미터가 한정되어 있는 화자 인증 시스템에서는 MAP 추정이 가장 유효한 적응 기법이다. 따라서 다음과 같은 적응 식의 변형된 형태를 사용하였다.

Step 1> 개인 코드북에 VQ 과정수행 관측열 발생

$$O_{T1} = \{o_1, o_2, \cdots, o_T\} \tag{5}$$

Step 2> 코드북 갱신

$$\mu_k = \alpha \mu_k + (1-\alpha) x_t \tag{6}$$

μ_k는 k번째 코드북중심값, $0.7 \le \alpha \le 1$를 의미한다.

Step 3> 갱신된 코드북에 VQ 과정 수행 관측열 2 발생

$$O_{T2} = \{o_1, o_2, \cdots, o_T\} \tag{7}$$

Step 4> O_{T1}과 O_{T2}를 이용하여 Baum-welch 재추정 수행 λ_n추정

$$\overline{\lambda_c} = \alpha \lambda_c + (1-\alpha) \sum \lambda_n \tag{8}$$

여기서 $0.7 \le \alpha \le 1$ 의 값이며, λ_c는 기존의 화자모델, λ_n는 인증된 관측열에 의한 화자 모델, $\overline{\lambda_c}$는 MAP 적응 기법에 의해 갱신된 화자모델을 의미한다.

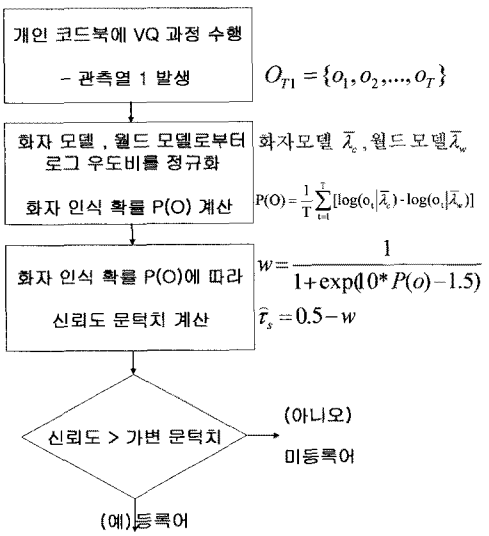


그림 5. 가변 신뢰도 문턱치 결정 방법을 사용한 미등록어 거절 방법

3.2 가변 신뢰도 문턱치 결정

음성인식을 통해 자동 분할된 호출 키워드 구간에 대한 신뢰도 측정 및 가변 문턱치 결정 기법을 제안하고 그림 5와 같이 나타내었다.

여기서, ω는 가변 신뢰도 문턱치 값을 정규화하기 위하여 시그모이드(sigmoid) 함수를 사용하여 만든 가중치이다.

제안한 가변 문턱치는 화자 인식 시스템을 음성인식 시스템과 통합 운용시킴으로써 환경변화에 적응하도록 신뢰도 문턱치를 변화시키게 되어, 미등록어 거절기능을 향상시킬 뿐 아니라, 잡음환경 하에서 등록어 임에도 불구하고 거절하는 오류까지도 감소시킬 수 있다.

4. 실험 및 결과

4.1 실험 방법

미등록어 거절기능의 성능평가를 위한 항목은 아래와 같은 기준으로 측정하였다.

등록어:

CA(Correctly accepted for keyword)
- 등록 어휘를 제대로 승인할 경우

FAI(False accepted in-grammar word)
- 등록어휘로 승인했으나 잘못 인식한 경우

FR(False rejected for keyword)
- 등록어휘를 잘못 거절한 경우

미등록어:

CR(Correctly rejected for OOV)
- 미등록어휘를 제대로 거절한 경우

FAO(False accepted out of grammar wWord)
- 미등록 어휘를 잘못 승인한 경우

본 논문에서 사용한 가변어휘 단어 인식 시스템은 새로운 인식대상 단어목록이 발생하면 음성훈련과정을 새로 수행하지 않고 발음사전만을 새로이 교체하여 단어 모델을 구성하게 된다. 이러한 단어 인식기를 구현하기 위해서는 한국어에 존재하는 모든 음소를 충분한 음소환경에서 모델링해야 한다. 한국어의 모든 음소를 정확하게 모델링하기 위해서는 다양한 환경에서 훈련데이터를 만들어야하고, 이를 음소모델에 잘 반영하기 위한 음소 모델 구조를 가져야한다. 이를 위해 ETRI의 PBW(Phonetically Balanced Words) 445DB를 사용하였고 43개의 문맥 독립 음소 모델을 사용하였다. 본 논문에서 문맥 독립 음소 모델훈련에 사용한 PBW 445DB는 총 445개의 어휘로 구성이 되어 있고 남자 22명과 여자 19명이 445 어휘를 각각 2번씩 발성한 데이터로 구성이 되어 있다.

음성은 사무실 환경과 잡음 환경에서 이동기기 등에 내장되는 내장형 마이크를 사용하여 16Khz Mono로 녹음되고 16bit PCM 양자화를 사용하였다. 실험음성으로 300명의 성인 남성이 참가하였다. 사무실 환경은 일반 사무실보다는 좀 더 소음이 있는 상태였으며 대체로 50~55dB 정도의 소음을 나타내고 있었고, 잡음 환경은 사무실에서 오디오를 켜 놓고 근접한 위치에 마이크를 두고 65~75dB 정도의 소음환경 하에서 실험을 하였다.

본 논문의 실험을 위해서 등록어(핵심어) 모델은 트라이폰(Triphone)의 문맥 종속형 음소 모델로 확장하여 만들었다. 실제로 시스템에서 쓰이게 될 명령어들로 호출 키워드를 포함하는 200개를 "홈시스, 거실 전등 켜"의 형태로 "호출 키워드(Call-keyword) + 제어 아이템 +명령어"로 발성되어지도록 구성하여 만들었다. 미등록어 거절 기능의 알고리즘에서 문턱치를 결정하기 위해서 사용하는 테스트 데이터를 위해 200명의 데이터를 사용하였고 성능 평가를 위한 평가데이터에는 나머지 100명의 데이터를 사용하였다.

음성이 입력되면 끝점 검출기를 통해 음성 구간만을 검출하게 되고 특징 추출과정을 거치게 된다. 각 트라이폰의 모델링은 CHMM(Continuous hidden markov model)을 사용하였는데 3-state Left-to-Right 모델, 4가지 혼합밀도로 정의하였다. 훈련은 Baum-Welch 재추정(forward-backward re-estimation)을 이용하였다.

필러 모델은 화자의 발성 명령어 중에서 등록어를 제외한 부분을 흡수하기 위해 등록어를 제외한 음성부분이나 비음성 부분으로 구성하게 하였다. 필러 모델은 실제 핵심어 검출기의 성능향상에 큰 영향을 미치게 되므로 묵음(silence) 모델, 말더듬(hesitation) 모델, 핵심어 이외의 단어모델, 잡음(noise)모델, 배경잡음 모델 등이 이에 포함된다.

4.2 기존 발화 검증 방법에 의한 실험 결과

기존의 발화 검증방법에 대해서는 2장에서 자세

표 1. 기존의 발화 검증 방법

분류 / 삽입음소수	사무실내 잡음환경				
	CA	CR	FAI	FAO	FR
4개이상시 거절	78.12	69.35	5.12	30.65	16.74
3개이상시 거절	76.35	78.17	4.91	21.83	18.74
2개이상시 거절	75.81	67.42	4.26	32.58	19.93
1개이상시 거절	64.19	73.18	2.30	26.82	33.51

표 2. 기존의 발화 검증 방법

분류 / 삽입음소수	사무실 환경				
	CA	CR	FAI	FAO	FR
4개이상시 거절	91.43	51.36	4.65	48.64	3.92
3개이상시 거절	90.15	58.43	4.12	41.57	5.73
2개이상시 거절	89.62	68.29	3.81	31.71	6.58
1개이상시 거절	77.36	81.44	2.91	18.56	19.73

히 다루었다. 표 1, 2에서는 기존의 발화검증 방법을 사용하였을 때의 인식결과들을 보여주고 있다. 기존의 방법은 문턱치로 단어의 패널티와 삽입된 음소의 개수를 사용하였다. 단어의 패널티는 기존의 발화 검증 방법에서 최상의 결과가 나오도록 대략 -38.0의 값을 사용하여 실험하였고 그 결과를 표에 나타내었다. 기존의 발화 검증 방법은 사무실내 잡음 환경에서는 CA와 CR이 최대78.12%, 69.35%로 나타났고, 일반 사무실 환경에서는 91.43%, 68.29%로 대체로 제안한 방법에 비하여 인식률이 떨어지는 것을 볼 수 있었다. 기존의 방법으로도 조용한 환경에서는 어느 정도의 CA을 보장받을 수 있었으나 CR 성능은 잘 보여주지 못하였다. 또, 잡음환경에서는 성능이 급격히 떨어짐을 볼 수 있었다.

4.3 제안한 발화 검증 방법에 의한 실험 결과

제안한 방법은 호출 키워드의 정보를 바탕으로 식(12)을 이용한 가변 신뢰도 문턱치를 사용하였다. 제안한 발화 검증 방법에 대한 인식결과는 호출 키워드를 포함하여 조사였을 때와 화자인식 부분에서 사용한 호출 키워드에서 화자에 대한 정보만 추출하고 인식은 그 후의 명령어만을 사용하였을 때와 비교했을 때 차이가 나지 않았기에 처리속도 향상을 위해서 호출 키워드를 제외한 명령어 부분만을 발화 검증시스템에서 사용하였다. 실험환경에 따른 화자확인 확률값은 표 3와 같이 나타났다. 이에 따른 문턱치의 분포는 등록된 화자일 경우 대체로 약 -0.5정도의 값을 가지며 미등록 화자로 판별된 경우 약 -0.48 ~ 0.5의 분포를 나타내었다.

표 3 실험환경에 따른 화자확인 확률값 (표시간격 : 0.2)

실험 환경	등록화자	미등록화자
사무실	0.5 ~ 0.9	1.1 ~ 2.5
잡 음	0.7 ~ 1.3	1.3 ~ 2.5

본 논문에서 제안한 발화검증 방법을 사무실 환경에서 실험한 결과를 표4에 나타냈으며, 등록된 화자의 인식률과 미등록 화자의 인식률의 차이를 확인할 수 있다. 등록된 화자의 확률값에 대하여 문턱치가 거의 일치하였으며 미등록 화자의 경우 문턱치의 변화가 등록화자에 비해 큰 것을 확인할 수 있다. 문턱치가 낮을수록 등록어휘를 제대로 승인할 확률인 CA가 높고 미등록어에 대한 거절률인 CR의 값이 낮음을 확인할 수 있고, 문턱치가 높아질수록 CA는 낮아지고 CR은 높아짐을 확인할 수 있다.

또, 표 5에는 사무실 내에서 인위적으로 잡음환경을 조성하여 실험한 결과는 나타내었다. 잡음 환경에서도 문턱치가 낮을수록 등록어휘를 제대로 승인할 확률인 CA가 높고 미등록어에 대한 거절률인 CR의 값이 낮음을 확인할 수 있고, 문턱치가 높아질수록 CA는 낮아지고 CR은 높아짐을 확인할 수 있다.

그림 6의 실험결과 차트에서 화자인식 시스템의 화자 확률값이 발화검증에 어떻게 영향을 미치는지

표 4. 제안한 발화 검증 방법(사무실환경)

화자확인 확률값	Weight	문턱치	CA	CR
0.5	0.999955	-0.499955	94.23	88.23
0.7	0.999665	-0.499665	94.02	88.51
0.9	0.997527	-0.497527	93.18	88.99
1.1	0.982014	-0.482014	91.54	89.38
1.3	0.880797	-0.380797	89.46	89.56
1.5	0.500000	0.000000	88.78	90.43
1.7	0.119203	0.380797	86.51	92.07
1.9	0.017986	0.482014	86.16	93.74
2.1	0.002473	0.497527	85.83	94.81
2.3	0.000335	0.499665	85.62	94.89
2.5	0.000045	0.499955	85.10	95.11

표 5. 제안한 발화 검증 방법(잡음환경)

화자확인 확률값	Weight	문턱치	CA	CR
0.7	0.999665	-0.499665	91.14	84.47
0.9	0.997527	-0.497527	91.02	84.78
1.1	0.982014	-0.482014	90.86	84.75
1.3	0.880797	-0.380797	89.51	85.51
1.5	0.500000	0.000000	86.36	85.99
1.7	0.119203	0.380797	84.73	88.38
1.9	0.017986	0.482014	83.62	89.56
2.1	0.002473	0.497527	83.14	91.43
2.3	0.000335	0.499665	83.10	92.07
2.5	0.000045	0.499955	83.02	92.74

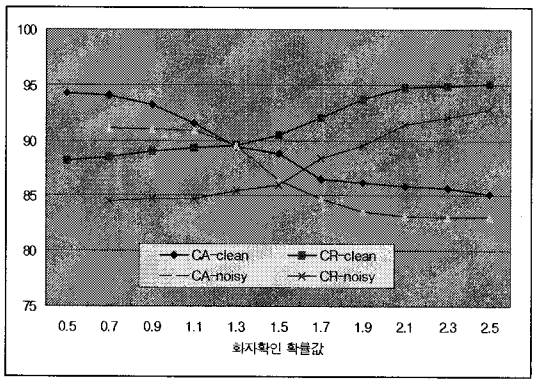


그림 6. 화자확인 확률값에 따른 인식률의 변화

확인할 수 있었다. 등록화자일 경우 조용한 환경에서의 확률값이 0.5 ~ 0.9정도로 나타나고 이 값이 발화검증의 문턱치에 반영이 된다. 잡음 환경에서는 등록화자의 확률값이 0.7 ~ 1.3정도로 높아지게 되며 발화검증의 문턱치에 반영이 되며 문턱치가 높아져서 CA는 감소하지만 CR이 높아져 미등록어가 등록어로 오인식될 확률이 낮아지게 된다.

5. 결 론

본 논문에서는 가변어휘 단어 인식기의 미등록어 거절 성능을 향상시키기 위한 방법으로 입력 발성 중 호출키워드 부분을 추출하여 화자 인식 확률값을 추출하고 이를 바탕으로 발화검증시스템에서의 신뢰도 문턱치를 가변적으로 적용하는 방법을 제안하였다.

호출 키워드를 사용하는 시스템에서는 호출 키워드를 통하여 명령의 처리여부를 먼저 결정하게 되므로 상당부분 미등록어에 대한 사전 검증을 한다고 볼 수 있다. 즉, 화자가 호출 키워드를 발성하지 않은 상태에서 제어 명령어를 발성하여도 이는 인식되지 않게 되며 이 때문에 뜻하지 않는 상황에서의 오작동이나 미등록어에 대한 잘못된 인식을 방지할 수 있다. 호출 키워드를 통하여 화자 확인 확률값을 추출하여 발화 검증을 하게 되므로 한정된 인원이 사용하게 되는 홈 네트워크 시스템이나 네비게이션 시스템에서 인식 성능의 향상이 있음을 확인할 수 있었다. 기존의 발화 검증 시스템과 비교하였을 때 일반적인 사무실 환경과 사무실내 잡음환경 하에서 CA와 CR이 각각 향상되었다. 등록된 화자로 판별이 되었을 경우 낮은 화자확인 확률값으로 인해 가변 신뢰도 문턱치가 낮아져 CA가 높게 나타나서 등록된 단어에 대한 인식률이 높았고, 미등록 화자로 판별이 되었을 때는 높은 화자확인 확률값으로 인해 가변 신뢰도 문턱치가 높아져 CR이 높아져서 미등록어에 대한 거절 기능이 향상됨을 확인할 수 있었다.

본 논문에서 제안한 발화 검증방법은 CA와 CR이 서로 상반되는 방향으로 인식률이 변화되어 일관성이 떨어지는 것을 확인하였다. 향후 과제로는 이러한 CA와 CR의 일관성이 결여되는 문제를 해결할 방법을 연구하고, 자연어 형태의 명령어 인식에 대한 미등록어 검출 성능 향상에 대한 연구가 이루어 져야 하겠다.

참 고 문 헌

[1] 장민석, 김성국, "음성 인터페이스의 기술 현황과 표준화 동향," *한국멀티미디어학회지*, 제10권, 1호, pp. 83-97, 2006.

[2] M. Rahim, C-H Lee, and B-H. Juang, "Robust utterance verification for connected digits recognition," *Proc. of ICASSP*, Vol.1, pp. 285-288, 1995.

[3] 김우성, 구명완, "반음소 모델링을 이용한 거절 기능에 대한 연구," *한국음향학회지*, Vol.18, No.3, pp. 3-9, 1999.

[4] R. A. Sukkar and C-H. Lee, "Vocabulary independent discriminative utterance ver-

ification for non-keyword in subword based speech recognition," *IEEE Trans. on speech and audio processing*, Vol.4, No. 6, pp. 420-429, Nov. 1996.

[5] Mazin G Rahim, Chin-Hui Lee, Biing-Hwang Juang, and Wu Chou, "Discriminative utterance verification using minimum string verification error(MSVE) training," *ICASSP*, Vol.1, pp. 3585-3588, 1996.

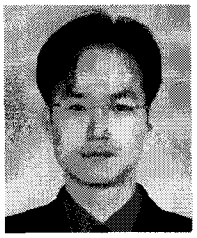
[6] Hoi-Rin Kim, SingHun Yi, and Hang-Seop Lee, "Out-of-vocabulary rejection using phone filler model in variable vocabulary word recognition," *ICSP*, Vol.1, pp. 337-339, 1999.

[7] Li Jiang and Xuedong Huang, "Vocabulary-independent word confidence measure using subword features," *ICSLP*, Vol.7, pp. 3245-3248, 1998.

[8] J. Kim, J Lee, and S Choi, "Hybrid confidence measure for domain-wpecific keyword spotting," *Proc. of IEA/AIE*, Vol.15, pp. 736-745, 2002.

[9] J.P. Neto, C. Martins, and L. B. Almeida, "An incremental speaker-adaptaion technique for hybrid HMM-MLP recognizer," *ICSLP96*, Vol.3, pp. 1293-1296, 3-6 Oct. 1996.

[10] C. -H. Lee nad J. -L. Gauvain, "Speaker adaptaion based on MAP estimaion of HMM parameters," *Proc. IEEE ICASSP*, Vol.2, pp. 558-561, 27-30 Apr. 1993.



방 기 덕

1998년 2월 안양대학교 전자통신공학과 공학사
2000년 2월 광운대학교 대학원 전자통신공학과 공학석사
2002년 2월 광운대학교 대학원 전자통신공학과 박사과정 수료
2004년 2월~현재 (주)한국파워보이스 기술연구소 연구원
관심분야 : 음성신호처리

강 철 호

1975년 2월 한양대학교 전자공학과 공학사
1979년 2월 서울대학교 대학원 전자공학과 공학석사
1988년 2월 서울대학교 대학원 전자공학과 공학박사
1977년 3월~1982년 2월 국방과학연구소 연구원
1991년 2월~1992년 1월 미국 일리노이대학교 객원교수
1983년 3월~현재 광운대학교 전자통신공학과 정교수
관심분야 : 음성신호처리, 통신신호처리