한국음향학회지 제22권 제3호 pp. 194~201 (2003)

음의 유사도 비율 누적 방법을 이용한 발화검증 연구

A Study on Utterance Verification Using Accumulation of Negative Log-likelihood Ratio

한 명 희*, 이 호 준*, 김 순 협*
(Myung-Hee Han*, Ho-Jun Lee*, Soon-Hyub Kim*)

*광운대학교 컴퓨터공학과

(접수일자: 2003년 2월 15일; 채택일자:2003년 3월 14일)

음성인식에서 신뢰도 측정이란 인식된 결과에 대한 신뢰 여부를 결정하는 것이다. 신뢰도는 프레임을 음소 및 단어 수준으로 통합하여 측정된다. 단어 인식의 경우, 신뢰도를 이용하여 인식 결과와 미등록 어휘를 검증한다. 따라서 이러한 후처리를 통해 이를 인식 결과로 승인하지 않음으로써 성능을 높일 수 있다. 본 논문에서는 기존의 신뢰도 측정 방법인 로그 유사도 비를 수정하여 신뢰도를 측정하였다. 제안된 방법은 프레임 수준에서 음소 수준으로 신뢰도를 통합할 때 로그 유사도 비가 음수인 것만을 누적하는 것이다. 단어 인식기의 인식 결과에 대한 검증 성능을 기존의 방법과 비교한 결과, CAR (Correct Acceptance Ratio)이 90%인 지점에서 FAR (False Acceptance Ratio)을 미등록 어휘에 대해서는 약 3.49%, 오인식에 대해서는 15.25% 감소시킬 수 있었다.

핵심용어: 신뢰도 측정, 발화검증, 후처리
투고분야: 음성처리 분야 (2.5)

In speech recognition, confidence measuring is to decide whether it can be accepted as the recognized results or not. The confidence is measured by integrating frames into phone and word level. In case of word recognition, the confidence measuring verifies the results of recognition and Out-Of-Vocabulary (OOV). Therefore, the post-processing could improve the performance of recognizer without accepting it as a recognition error. In this paper, we measure the confidence modifying log likelihood ratio (LLR) which was the previous confidence measuring. It accumulates only those which the log likelihood ratio is negative when integrating the confidence to phone level from frame level. When comparing the verification performance for the results of word recognizer with the previous method, the FAR (False Acceptance Ratio) is decreased about 3.49% for the OOV and 15.25% for the recognition error when CAR (Correct Acceptance Ratio) is about 90%.

Keywords: *Confidence measure, Utterance verification, Post-processing*
ASK subject classification: *Speech signal processing (2.5)*

I. 서 론

음성인식을 이용한 다양한 응용제품 및 서비스들이 개발되면서 순수하게 음성을 인식하는 것 이외에도 음성인식 시스템에서 해결해야 할 문제들이 많아지게 되었다. 음성을 인터페이스로 이용하는 제품 및 서비스들은 대부분의 경우 제한된 음성 입력을 갖게 된다. 따라서 인식할 수 있는 영역이 제한되어 있으며 그 외의 불필요한 음성의 입력을 처리할 수 있는 기능 역시 인식 시스템에 포함되어야 한다. 이러한 기능은 인식 시스템의 형태에 따라 미등록 어휘를 거절하거나 연속 문장의 입력에서 핵심어의 검출을 가능하게 해 준다.

인식 시스템이 단어 인식기인 경우 미등록 어휘에 대한 거절 기능으로 앞에서 언급한 기능이 이루어지게 된다. 미등록 어휘란 인식기에서 인식 대상 어휘가 아닌 것을 말한다. 좀 더 넓게 비인식 대상을 언급하자면 미등록 어휘뿐만 아니라 숨소리나 주변 소리 같은 비음성 역시 포함된다. 이러한 비인식 대상을 거절하기 위해서는 신뢰도라는 것을 측정하여 그 신뢰성 여부를 결정하게 된다. 핵심

교신저자: 한명희 (angma@kw.ac.kr)
139-701 서울시 노원구 월계1동 447-1
광운대학교 컴퓨터공학과 정보통신연구실 506호
(전화: 02-940-5123; 팩스: 02-941-8919)

어 검출기의 경우 입력은 연속된 음성의 형태이다. 이 중에서 핵심어 구간만을 추출하는 방법 역시 단어 인식기에서 사용된 신뢰도를 기반으로 핵심어와 비핵심어를 구분하는 것이다. 또한 핵심어 검출 이후에도 핵심어 구간이라고 판단된 부분 역시 다시 측정하여 신뢰성을 판별할 수 있다.

인식된 결과에 대한 검증은 시스템의 형태에 따라 그 역할이 조금씩 다르지만 기본적으로 신뢰성의 여부를 판단하는 기준이 된다. 본 연구에서는 인식된 결과를 검증하는 후처리에 대해 다룰 것이다. 인식된 결과를 거절해야 하는 경우는 앞에서 언급한 것처럼 미등록 어휘를 인식한 경우와 인식기의 인식 오류이다. 음성인식의 처리 특성상 오류는 항상 존재하기 마련이며 이러한 오류를 없애는 것이 음성인식의 궁극적인 목표가 될 수 있다. 이런 오류에 대한 처리를 다른 방법으로 접근하자면 인식 결과에 대한 신뢰성을 측정하여 오인식을 검출하는 것이다. 이때의 오인식은 사용자가 인식 대상을 발성한 경우를 음성인식 시스템이 오인식한 경우를 나타낸다.

기존의 미등록 어휘의 거절이나 핵심어를 검출하는 것에 쓰인 신뢰도를 바탕으로 오인식된 결과를 검출해 내는 것은 비인식 대상을 검출하는 것보다 그 성능이 떨어짐을 실험을 통해 알 수 있었다. 이러한 오인식을 검출하는데 그 성능이 떨어지는 것은 당연한 결과일 수도 있다. 왜냐하면 현재의 음성인식 시스템의 결과는 상대적으로 확률이 높은 것을 선택하도록 처리되므로 이는 인식 모델 자체의 문제일 수도 있기 때문이다. 또한 인식 결과에 대해 어떤 처리를 하는 후처리는 인식 결과를 바탕으로 하는 부가적인 것이기 때문에 어느 정도 제약이 따르게 된다. 후처리 방법 중의 한 가지는 인식 결과를 여러 개의 후보로 구성하여 이 후보들에 언어적인 정보를 이용하여 최종적인 인식 결과를 결정하는 것이다.

본 논문에서는 신뢰도 측정의 관점에서 이러한 문제를 해결하고자 하였다. 신뢰도를 측정하여 인식 결과를 거절하는 것은 인식률을 직접적으로 올리는 것은 아니다. 하지만 음성인식 시스템의 신뢰성을 유지함으로써 오인식으로 인한 사용자의 불편을 최소화하여 서비스의 질을 향상시킬 수 있게 된다. 본 논문에서는 기존에 사용된 신뢰도 측정 방법을 수정하여 비인식 대상에 대한 변별 능력을 유지하면서 오인식된 결과에 대한 거절을 높여 인식 시스템의 성능을 높일 수 있었다. 2장에서는 신뢰도 측정을 어떻게 해왔는지에 대해 알아보고 제안한 수정된 방법 역시 언급할 것이다. 3장에서는 신뢰도 측정 가능이 인식 시스템의 후처리 시스템으로 구성되는 구조를 살펴보고 4장에서는 단어 인식기에서 기존의 신뢰도 측정 방법들과 제안된 방법의 성능을 실험을 통해 비교하여 성능의 유효성을 확인할 것이다.

II. 신뢰도 측정

신뢰도는 기본적으로 통계적인 가설을 검증하는 방법으로 측정된다[1]. 이것은 다음 절에서 다룰 유사도의 비를 이용하여 구할 수 있으며 인식된 모델과 반모델 (anti-model)의 비를 말한다. 신뢰도는 프레임 수준의 신뢰도를 음소 수준으로 통합한 뒤, 음소 수준의 것을 단어 수준으로 통합하여 최종적으로 측정하게 된다. 먼저 유사도 비와 반모델을 정의하고 이것들을 이용하여 기존의 방법들이 어떻게 신뢰도를 측정하는지 살펴본 뒤, 제안한 방법을 소개한다.

2.1. 유사도 비

일반적으로 신뢰도 측정은 통계적인 가설 테스트의 관점에서 이루어진다. 참고문헌[1]의 정의에 의하면 다음과 같다.

우선 어떤 O를 실제 음성의 관측 세그먼트 (segment)라 하면 음성인식 과정에서 O가 입력되었을 때는 크게 두 가지의 가정이 가능하다. 즉, 그 O가 실제 어떤 음성 세그먼트 k로부터 발화되었을 것이라 가정하는 것이 가능한데 이를 영가설 (null hypothesis)라 하고 H_0으로 표현한다. 반면, 그 O가 실제 음성 세그먼트 k가 아닌 다른 유사한 음성에서 발화되었을 것이라 가정할 수 있는데 이를 대립가설 (alternative hypothesis)라고 H_1로 표현한다. 그러면 주어진 테스트 세그먼트 O에 대해 발화 검증 과정은 영가설에 대한 확률과 대립가설의 확률을 비교하여 영가설에 대한 확률이 크면 이를 인식하고 아니면 거절하는 것이다.

$$P(O|H_0) > P(O|H_1) \tag{1}$$

위 식을 Bayes rule에 의해 다시 쓰면

$$P(H_0|O)P(H_0) > P(H_1|O)P(H_1) \tag{2}$$

$$\frac{P(H_0|O)}{P(H_1|O)} > \frac{P(H_1)}{P(H_0)} \tag{3}$$

이 된다. 여기서 $P(H_0|O)$는 은닉 마코프 모델 (HMM) λ_k에서 O가 관측될 확률이고, $P(H_1|O)$는 그와는 다른 모델에서 O가 관측될 확률이다.

이를 시간에 대해 정규화한 유사도 비 (likelihood ratio)로 나타낸 식은 다음과 같다[2].

$$\frac{Score_{keyword} \Big|_{StartTime_{keyword}}^{EndTime_{keyword}}}{EndTime_{keyword} - StartTime_{keyword}} \quad (4)$$

일반적으로 식 (4)에 로그를 취한 형태로 쓰며 이를 로그 유사도 비 (LLR: log likelihood ratio)이라 하고 식 (5)와 같이 쓴다.

$$Score_{keyword} - Score_{BackgroundModel} \Big|_{StartTime_{keyword}}^{EndTime_{keyword}} \quad (5)$$

이때의 $P(H_1|O)$의 H_1나 $Score_{BackgroundModel}$은 반모델의 유사도가 된다.

2.2. 반모델

반모델 (anti-model)은 앞 절에서 잠깐 언급하였듯이 인식된 모델을 검증하기 위해 유사도 비를 구할 때 쓰이게 된다. 일반적으로 신뢰도를 측정할 때는 신뢰도가 인식된 모델과 인식되지 않은 모델들의 상대적인 유사도이기 때문에 각 모델에 대해 가장 혼돈하기 쉬운 유사한 것들을 찾아 그에 대한 모델을 구성하며 이를 반모델이라 한다. 음소 단위로 인식하는 경우 모델들은 음소 모델이며 이때의 반모델은 반음소 모델 (Anti-Phoneme Model)이 된다.

반음소 모델을 구성하는 방법에는 음성학적 지식을 이용한 분류 방법과 유사한 음소들의 집합을 생성하는 것이 있다. 일반적으로 음성학적 분류에 의한 방법보다 유사한 음소들의 집합을 생성하는 것이 LLR이 높은 것으로 알려져 있다. 유사 음소 집합을 구성하는 모델들이 많을수록 반음소 모델이 잘 모델링 되지만, 집합의 크기가 커질수록 훈련 데이터의 양이 많아야하는 단점이 있다[3].

또 다른 방법으로는 인식 단어를 훈련하는 과정 중에 생성되는 단일 음소 모델로 반음소 모델을 생성하는 것이다. 이렇게 함으로써 유사 음소 집합을 만들기 위해 따로 훈련할 필요가 없어지게 된다[3]. 그림 1은 훈련 과정 중에 어느 시점에서 반음소 모델이 생성되는지를 나타내고 있다. 특징을 가지고 인식 훈련을 하는 순서는 먼저 단음소 모델을 만든 뒤 전후 문맥을 고려한 트라이폰 모델 (tri-phone model)로 확장하게 된다. 이때 반음소 모델은 인식을 위한 모델을 만드는 중간 과정에서 생성됨을 알 수 있다.

본 논문에서는 훈련 과정 중에 생성된 음소 모델을 반모델로 삼았으며 이때 모델의 출력확률 역시 인식 모델과 동일한 연속분포 은닉 마코프 모델 (CHMM)의 형태이다.

2.3. 단어 수준 신뢰도 측정

일반적으로 신뢰도는 프레임 수준의 것을 통합하여 음소 수준으로, 음소 수준을 통합하여 단어 수준의 신뢰도를 측정하게 된다[4]. 먼저 프레임 수준 (frame level)을 보면 다음과 같다.

$$LLR_i(o_t) = \log \frac{P(o_t|\lambda_i)}{P(o_t|\lambda_i^a)} = \log \frac{l_i(o_t)}{l_i^a(o_t)} \quad (6)$$

λ_i : 상태 i에서 끝나는 음소 단위의 HMM

λ_i^a : 반모델

l_i : 상태 i에서 HMM의 프레임 유사도

l_i^a : 반모델의 프레임 유사도

프레임 수준 신뢰도를 음소 수준 (phone level)의 신뢰도로 통합한 것은 식 (7)과 같다.

$$PLLR_n(o_t) = \frac{1}{\tau} \sum_{t-\tau \le i \le t} LLR_i(o_t) \quad (7)$$

τ는 음소 n의 프레임 지속시간이며 $LLR_i(o_t)$는 상태 i에 대응되는 HMM 수준의 한 프레임에 대한 LLR이다.

$$\log PCM_n(o_t) = \log \frac{1}{1+\exp(-\alpha \cdot (PLLR_n(o_t) - \beta))} \quad (8)$$

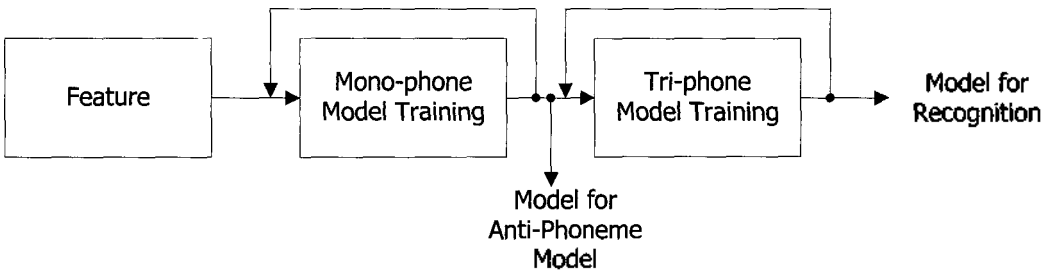

그림 1. 인식 모델과 반음소 모델 구축 방법
Fig. 1. Method of Modeling Recognition Models and Anti-Phoneme Models.

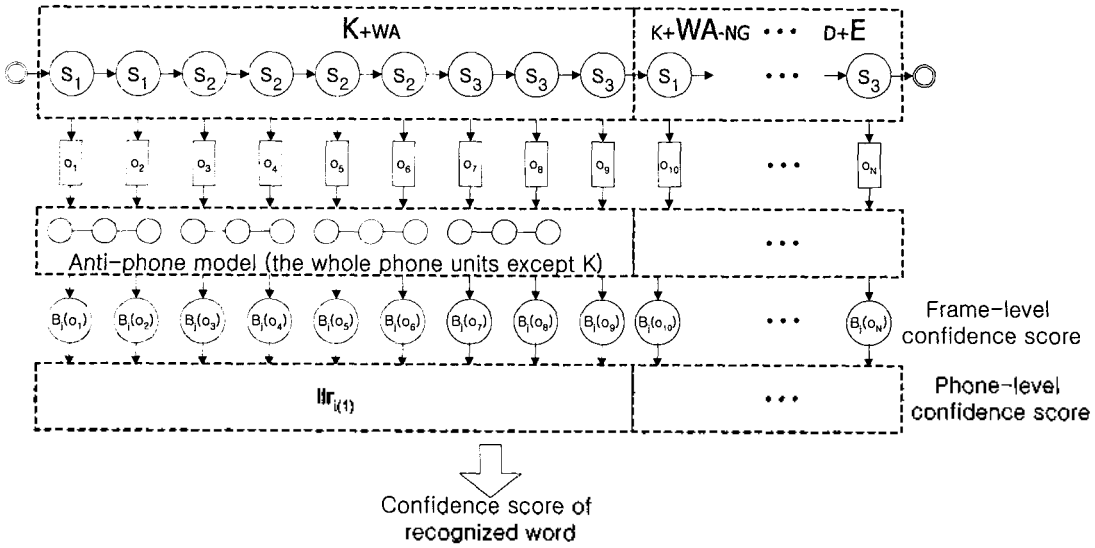

그림 2. 음소 및 단어 단위의 신뢰도
Fig. 2. Confidence Measure of the Phone and Word.

식 (8)의 시그모이드 함수 (sigmoid function)는 LLR의 값을 일정 범위 이내로 제한하기 위해 사용한다. $PCM_n(o_t)$는 프레임 t에서 끝나는 음소 n의 음소 수준 신뢰도이다. 각 단어의 끝에서 계산된 신뢰도인 단어 수준 신뢰도는 식 (9)와 같다.

$$\log WCM_w = \frac{1}{N}\sum_n \log PCM_n(o_t) \qquad (9)$$

N은 단어 w 내의 음소 수이며 t는 단어 w 내에서 음소 n이 끝나는 프레임이다.

단어 수준 신뢰도를 측정하는 것을 그림 2에 나타내었다. 그림 2를 보면 프레임 수준의 신뢰도를 음소 수준으로, 음소 수준을 단어 수준으로 통합하는 것을 알 수 있다.

2.4. 제안한 단어 수준 신뢰도 측정

기존의 LLR을 살펴보면 프레임 수준의 신뢰도를 음소 수준으로 통합할 때 각 프레임의 LLR을 계속 누적시킨 뒤 프레임 수로 정규화를 시켜준다. 등록 어휘와 비슷한 단어가 들어오는 경우나 오인식된 경우 프레임 신뢰도의 평균을 취하면 반모델의 확률이 더 큰 것이 다른 것에 의해 평활화되어 값이 작아지게 된다. 예를 들면 "가운데"라는 발성은 "광운대"와 뒤의 두 "운대"라는 발음이 같게 된다. 따라서 기존의 LLR은 LLR이 양수인 것은 인식된 결과에 대한 신뢰도를 증가시키고 음수인 것은 신뢰도를 감소시키므로 "가"와 "광"의 신뢰도보다 "운대"의 신뢰도에 더 큰 영향을 받아 그 변별력이 떨어지게 된다.

본 논문에서는 이러한 점을 감안하여 음 (negative)의 LLR만을 누적시키고 이를 음소수로 정규화하여 단어 수준 신뢰도를 측정하는 방법을 제안하였다.

그림 3을 보면 먼저 각 프레임에 해당하는 인식된 모델에 대한 반모델의 확률을 구한다. 인식된 모델에 대한 확률과 반모델의 확률을 비교하여 반모델의 확률이 더 크면 LLR은 음의 값이 되고 이를 누적시켜 음소 수준의 신뢰도를 측정한다. 이렇게 수정된 음소 수준의 LLR이 나오면 식 (9)와 같은 방법으로 단어 수준 신뢰도를 측정하여 음

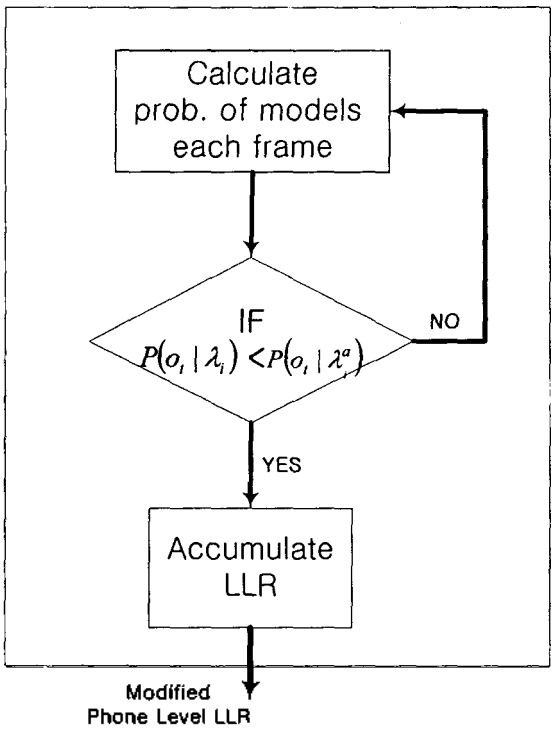

그림 3. 수정된 음소 수준 로그 유사도 비
Fig. 3. Modified Phone Level LLR.

소 수에 따른 정규화를 한다.

III. 음성인식 시스템의 구성

이번 장에서는 음성인식 시스템에서 검증 시스템이 어떻게 연결되고 어떤 정보를 입출력으로 하는가를 살펴본다. 검증 시스템은 발화를 검증 (Utterance Verification) 하는 기능으로 인식 후의 결과를 이용하여 검증을 하여 후처리 시스템 (Post-processing System)이라고도 한다. 먼저 검증 시스템의 구성을 자세히 살펴보고 음성인식 시스템과의 연결을 보기로 하자.

3.. 검증 시스템의 구성

그림 4를 보면 비터비 디코더를 통해 인식된 결과의 정렬 (align)된 정보가 검증 시스템의 입력으로 들어온다. 인식 결과의 형태는 음소 수준으로 프레임을 묶는 것으로 여길 수 있는데 실제로는 각 프레임이 음소 수준 모델의 번호를 가지고 있다. 그림 4에서 신뢰도 측정은 하나의 처리기로 처리할 수 있음을 알 수 있다.

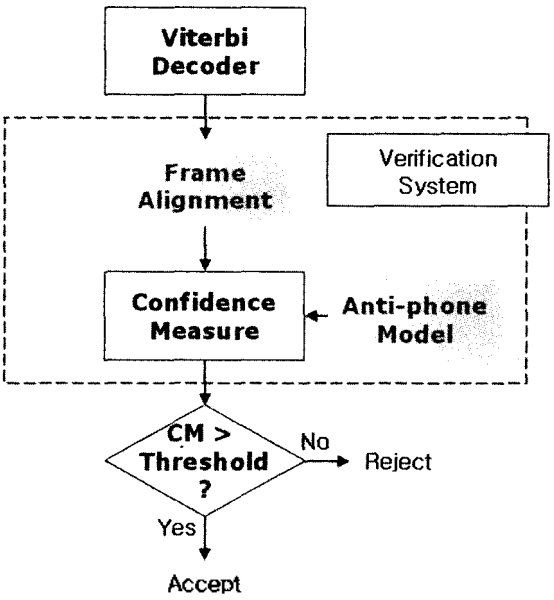

그림 4. 검증 시스템
Fig. 4. Verificarion System.

3.2. 음성인식 시스템 내의 검증 시스템의 구성

검증 시스템을 추가한 음성인식 시스템은 그림 5와 같다. 그림 5의 위 부분은 반모델의 생성 방법을 나타내고

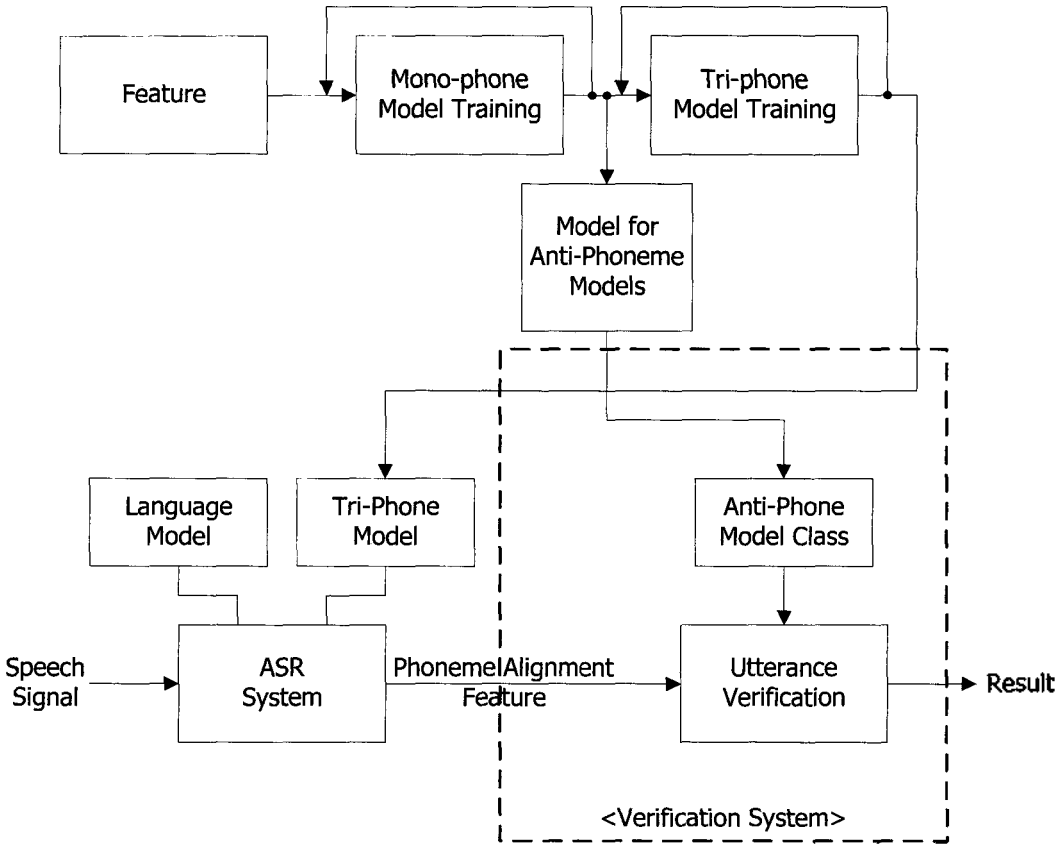

그림 5. 검증 시스템이 추가된 음성인식 시스템
Fig. 5. Added Verification System on Speech Recognition System.

있으며 이는 인식 모델의 생성 과정에서 발생함을 앞에서 언급하였다. 그 기능은 3.1절의 그림 4에 나타나있다.

IV. 실험 및 결과

4.1. 실험 환경

등록 어휘 500개인 단어 인식기로 등록 어휘에 대한 모델을 위한 음성 데이터베이스는 화자 30명분의 양으로 조용한 사무실 환경에서 수집하였다. 음성 데이터베이스는 8 KHz 샘플링 주파수로 이산화되어 16 bit로 양자화된다. 이 과정은 PC 사운드 카드를 통해 이루어진다. 마이크는 일반 마이크를 사용하였다.

기본 인식률을 구하기 위한 테스트 데이터는 모델 생성에 참여하지 않은 화자 10명이 각 500 단어씩 발성한 것으로 구성하였으며 기본적인 인식률은 91%이다. 테스트 데이터의 구성은 표 1과 같다.

테스트 데이터를 등록 어휘와 미등록 어휘로 구성한 것은 미등록 어휘에 대한 성능을 유지하면서 등록 어휘 중 오인식에 대한 성능을 보여주기 위해서이다.

표 1. 테스트 데이터의 구성
Table 1. Configuration of Test Data.

Type	Composition	Number of Words
In-Vocabulary	Correctly recognized	2275
	Incorrectly recognized	225
Out-Of-Vocabulary	PBW445	890

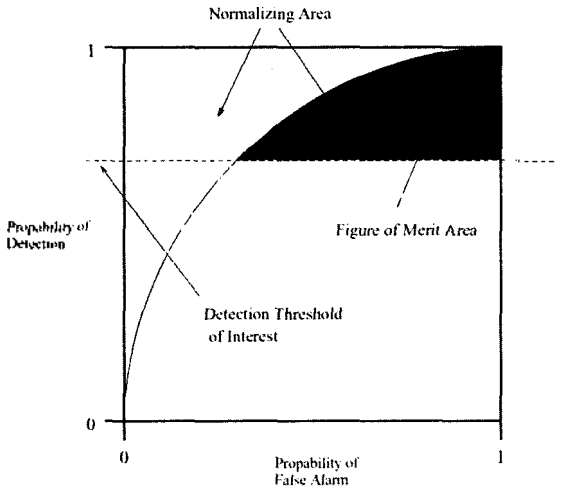

그림 6. 신뢰도 성능을 나타내기 위한 ROC[5]
Fig. 6. Illustration of the ROC as an indicator of confidence score performance[5].

4.2. 성능 평가 방법

신뢰도 측정 방법의 평가를 위해서 주로 쓰이는 것이 수신자 동작 특성 (ROC: Receiver Operating Characteristic) 곡선이다[5]. x축은 False Acceptance Ratio이고 y축은 Correct Acceptance Ratio이다. 일정한 문턱치에서 성능지수 (FOM: Figure Of Merit) 영역이 더 넓은 것이 더 나은 성능을 갖는다고 판단할 수 있다.

본 논문에서는 그림 6과 같이 CAR과 FAR을 가지고 제안한 방법과 기존 방법들의 성능 차이를 살펴볼 것이다. 미등록 어휘에 대한 성능과 오인식에 대한 성능의 시각적인 비교를 위해 올바르게 인식한 것의 임계값 변화에 따른 각각의 성능을 보기로 하자.

4.3. 성능 평가 및 결과

실험에서 각각의 비교 대상은 다음과 같다.

1. LLR을 프레임 수 및 음소 수로 정규화한 것 (LLR)
2. LLR을 시그모이드 함수를 적용하여 LLR의 범위를 한정지은 것 (LLR_SIG)
3. 음의 LLR만을 누적하여 음소 수로 정규화한 것 (NEG_LLR)

본 논문에서 실험의 기준이 되는 축은 CAR 축이다. 본래는 미등록 어휘의 인식의 경우 CRR (Correct Rejection Ratio)에 대한 성능을 비교하게 된다. 하지만 이럴 경우 본 논문에서 중점을 두고 있는 등록 어휘의 오인식된 결과에 대한 거절률을 직접적으로 비교하기가 어렵게 된다. 왜냐하면 이 경우에는 FAR을 보고 그 성능을 평가하기 때문이다. 따라서 본 논문에서는 오인식된 결과에 대한 신뢰도 측정 성능을 비교하기 위하여 CAR을 중심으로 두 경우를 살펴보고자 한다.

미등록 어휘에 대한 ROC 곡선은 등록 어휘 중 올바르게 인식된 것의 신뢰도 임계값 변화에 따른 미등록 어휘의 잘못된 승인에 대한 비율이다. 즉 올바르게 인식된 것의 신뢰도 임계값을 변화시키면 동일하게 미등록 어휘의 인식에 대한 신뢰도에 적용되고 이때의 CAR과 FAR을 보고 성능을 판단하게 된다. 미등록 어휘 (OOV: Out-of-Vocabulary)에 대한 실험의 결과를 보면 그림 7과 같다.

그림 7을 보면 제안한 방법 (NEG_LLR)이 기존의 방법 (LLR, LLR_SIG)과 미등록 어휘에 대한 성능이 거의 비슷함을 알 수 있다. 4.2절에서도 잠깐 언급하였듯이 올바르게 인식하는 것의 승인률을 유지함을 보이기 위해 미등록 어휘인 경우에도 CAR의 성능을 기본으로 하였다. 표 2를 보면 높은 CAR인 경우의 각 방법에 대한 성능이 나와 있다. 일반적으로 인식 결과에 대한 신뢰성이 높은 경우에

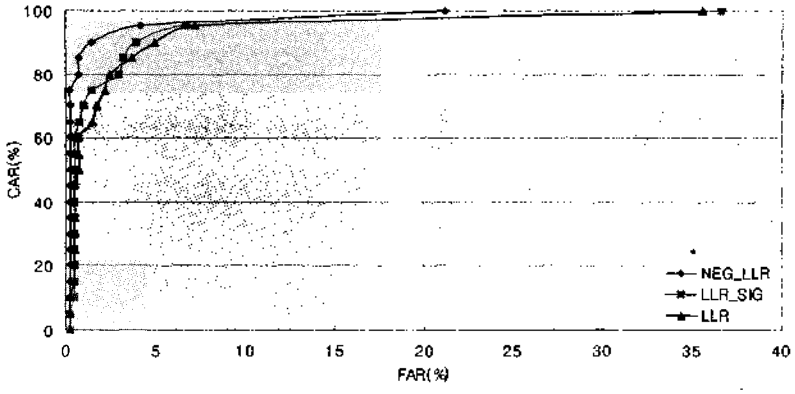

그림 7. 미등록 어휘에 대한 성능
Fig. 7. Performance of Confidence Measures for OOV.

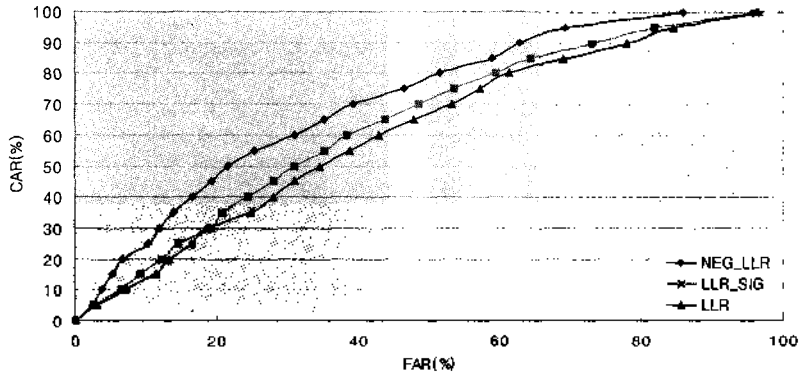

그림 8. 오인식된 결과에 대한 성능
Fig. 8. Performance of Confidence Measures for Recognition Error.

표 2. 미등록 어휘에 대한 상위 비율 성능
Table 2. High Position Ratio Performance for OOV.

Correct Acceptance (%)	False Acceptance (%)		
	LLR	LLR_SIG	NEG_LLR
95	7.27	6.72	4.23
90	4.98	3.98	1.49
85	3.73	3.23	0.75

표 3. 오인식된 결과에 대한 상위 비율 성능
Table 3. High Position Ratio Performance for Recognition Error.

Correct Acceptance (%)	False Acceptance (%)		
	LLR	LLR_SIG	NEG_LLR
95	84.53	81.84	69.28
90	78.03	73.09	62.78
85	69.06	64.35	58.97

는 CAR의 상위 지점에 대한 비교를 하게 된다[5]. 예를 들면 인식 환경의 잡음 정도에 따라 인식 결과에 대한 신뢰도가 시스템 외적인 요소로 달라지는 경우를 들 수 있다. 본 논문에서의 실험 환경은 조용한 사무실이므로 인식 시스템에 대한 신뢰성을 90% 지점으로 하여 표 2를 통해 나타내었다.

표 2에서 보면 CAR이 90% 지점에서 제안한 방법의 FAR이 LLR에 비해서는 3.49%, LLR_SIG에 비해 2.49% 낮음을 알 수 있다. 이를 통해 제안한 방법이 기존의 방법보다 조금 더 나은 수준으로 성능을 보임을 확인할 수 있다.

두 번째 실험은 본 논문에서 제안한 방법을 이용하여 오인식된 결과에 대한 검증성능을 보여주고 있다. 그림 8을 살펴보면 음의 로그 유사도 비를 이용한 것이 다른 두 방법에 비해 나은 성능을 보인다. 대체적으로 미등록 어휘의 경우보다 ROC 곡선이 완만함을 통해 그 성능이 떨어지는 것을 확인할 수 있다. 이는 서론에서도 언급하였듯이 등록 어휘의 오인식 자체에 대한 변별력이 떨어지기 때문이다. 하지만 본 논문에서 제안한 방법은 다른 기

존의 방법에 비해 CAR 전체에 걸쳐 그 성능이 높음을 알 수 있다.

또한 표 3에서 살펴볼 수 있듯이 CAR 90% 지점에서 제안한 방법의 FAR이 LLR에 비해 15.25%, LLR_SIG에 비해 10.31% 낮음을 확인함으로써 성능을 향상을 확인할 수 있었다.

따라서 본 논문에서 제안한 방법이 미등록 어휘에 대한 신뢰도 측정 성능을 기존의 방법들과 비슷하게 유지하면서 등록 어휘의 오인식에 대한 신뢰도 측정에서의 성능을 향상시킬 수 있음을 실험을 통해 알 수 있다.

V. 결 론

본 논문에서는 오인식된 결과의 검출률을 높이면 응용 제품이나 음성인식 시스템의 신뢰성을 높일 수 있게 된다는 점에 기반하여 인식 결과에 대한 검증 시스템을 구현하고자 하였다. 이러한 검증 시스템을 구현하기 일반적인 신뢰도 측정 방법을 이용하여, 미등록 어휘의 제거율에는 기존의 방법과 비슷한 성능을 유지하면서 오인식된 결과의 검출률을 높이는 방법을 제안하였다. 본 논문에서 제안한 방법은 기존의 LLR이 프레임 수의 평균에 의거 신뢰도를 계산하는 것을 음수의 LLR만을 누적하여 음소 수로 정규화하는 방법을 새로이 제안하여 그 성능이 높음을 보였다. 제안한 방법이 미등록 어휘에 대해 기존의 방법보다 최대 3.49% 정도 성능을 향상시켰으며 오인식된 결과에 대해서는 최대 15.25% 정도 성능을 향상의 결과를 보임으로써 제안한 방법의 성능을 확인하였다.

단어 인식기의 후처리 시스템으로 이러한 신뢰도를 측정하는 방법은 향후 핵심어 검출기의 핵심어 구간 검출에 대한 검증을 수행하는데 이용할 수 있을 것이다.

감사의 글

본 연구는 광운대학교 2001학년도 교내 학술 연구비의 지원으로 이루어졌습니다.

참 고 문 헌

1. 김우성, 구명완, "반음소 모델링을 이용한 거절기능에 대한 연구," 한국음향학회지, **18** (3), 3-9, 1999.
2. R. Rose and D. Paul, "A Hidden Markov Model based keyword recognition system," *IEEE ICASSP*, 129-132, 1990.
3. 김기태, 문광식, 김회린, 이영직, 정재호, "가변어휘 단어 인식에서의 미등록어 거절 알고리즘 성능 비교," 한국음향학회지, **20** (2), 27-34, 2001.
4. Myoung-Wan Koo, Chin-Hui Lee and Bling-Hwang Juang, "Speech recognition and utterance verification based on a generalized confidence score," *Speech and Audio Processing, IEEE Transactions on*, **9**, Issue 8, 821-832, Nov. 2001.
5. Kampari, "Word and phone level acoustic confidence scoring for speech understanding systems," Master's thesis, MIT, 1999.

저자 약력

● 한 명 희 (Myung-Hee Han)

2002년 2월: 광운대학교 컴퓨터공학과
2002년~현재: 광운대학교 컴퓨터공학과 석사과정 재학중
※ 주관심분야: 음성인식

● 이 호 준 (Ho-Jun Lee)

2001년 2월: 관동대학교 전자공학과
2001년~현재: 광운대학교 컴퓨터공학과 석사과정 재학중
※ 주관심분야: 음성인식

● 김 순 협 (Soon-Hyub Kim)

1974년 2월: 울산대학교 전자공학과 학사졸업
1976년 2월: 연세대학교 석사과정 졸업
1983년 2월: 연세대학교 공학박사
1979년 3월~현재: 광운대학교 컴퓨터공학과 교수
※ 주관심분야: 음성인식