다양한 신뢰도 척도를 이용한 SVM 기반 발화검증 연구

권석봉(ICU), 김회린(ICU), 강점자(ETRI), 구명완(KT), 류창선(KT)

<차 례>

<Abstract>

SVM-based Utterance Verification Using Various Confidence Measures

Suk-bong Kwon, Hoirin Kim, Jeomja Kang,
Myong-Wan Koo, Chang-Sun Ryu

In this paper, we present several confidence measures (CM) for speech recognition systems to evaluate the reliability of recognition results. We propose heuristic CMs such as mean log-likelihood score, N-best word log-likelihood ratio, likelihood sequence fluctuation and likelihood ratio testing(LRT)-based CMs using several types of anti-models. Furthermore, we propose new algorithms to add weighting terms on phone-level log-likelihood ratio to merge word-level log-likelihood ratios. These weighting terms are computed from the distance between acoustic models and knowledge-based phoneme classifications. LRT-based CMs show better performance than heuristic CMs excessively, and LRT-based CMs using phonetic information show that the relative reduction in equal error rate ranges between 8 ~ 13% compared to the baseline LRT-based CMs. We use the support vector machine to fuse several CMs and improve the performance of utterance verification. From our experiments, we know that selection of CMs with low correlation is more effective than CMs with high correlation.

* Keywords: Utterance verification, Confidence measure, Support vector machine.

1. 서 론

현재 음성인식 기술은 적용될 응용분야에서 환경에 적합한 충분한 음성데이터를 가지고 있으면 훈련을 통해 일정 수준의 인식성능을 갖는 시스템을 구성할 수가 있다. 하지만 사용 환경에 적합한 충분한 훈련데이터로 훈련된 최적의 음성인식 시스템일지라도 실제 환경에서 사용될 때에는 심각한 문제점들에 직면하게 된다. 아무리 사용 환경을 최대한 고려한다고 해도 실제 환경에서는 다양한 주변 잡음, 화자의 변화, 채널의 변화 등으로 인해 음성인식 과정에서 여러 가지 문제점들이 발생한다. 게다가 어떠한 음성인식기라도 인식과정에서 발생하는 오인식을 피할 수 없다. 특히 실제 환경에서는 오인식이 빈번히 발생한다. 따라서 실제 환경에서 사용될 인식기에서는 인식된 결과가 맞는지 또는 어느 정도 신뢰할 수 있는지에 대한 판단이 필요하다. 이와 같이 발화된 음성으로부터 얻은 인식 결과에 대한 신뢰도를 결정하는 기술을 발화검증(utterance verification)이라 한다. 즉, 발화검증을 통해 비인식대상어휘(out-of-vocabulary)를 거절하고, 인식대상어휘(in-vocabulary)라도 신뢰도가 떨어지는 인식결과를 거절한다.

발화검증 기술은 다양한 응용분야에서 적용될 수 있고 음성인식 시스템의 가치를 높여준다. 홈오토메이션에 음성인식 기술을 적용할 때는 높은 수준의 비인식대상어휘에 대한 거절기능이 요구되고, 인식대상어휘라도 오인식을 충분히 방지해 주어야 한다. 이러한 기술은 PDA, 휴대폰, 로봇 등 실제 환경에 노출되어 있는 단말기에 적용이 된다. 그 외에 핵심어 인식 시스템, 발화된 음성을 인식하고 자동적으로 재훈련을 하거나, 발화한 화자에 대해 자동적으로 모델을 적응하는 시스템 등 다양한 응용분야에 사용될 수 있고, 수요가 증가하고 있다.

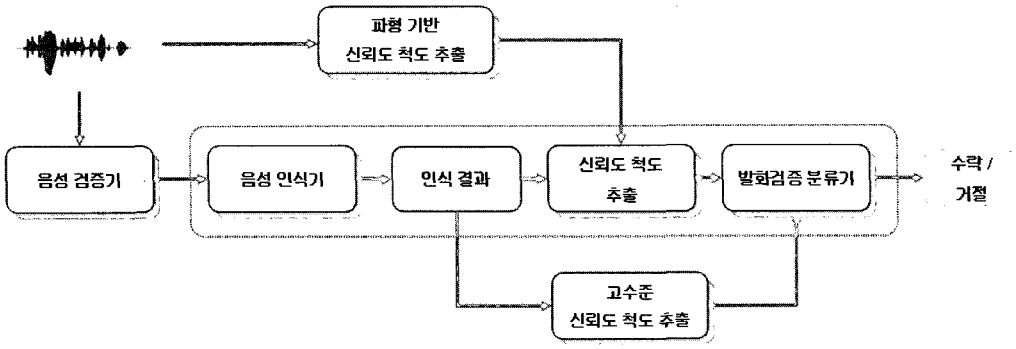
지난 수십 년 동안 많은 연구자들이 발화검증에 사용될 신뢰도 척도(confidence measure)를 계산하는 다양한 방법들을 연구해 왔다. 신뢰도 척도를 계산하는 알고리즘에 따라 크게 세 가지 범주로 분류할 수 있다. 발견법적(heuristic) 기반 신뢰도 척도[1], LRT(Likelihood Ratio Testing) 기반 신뢰도 척도[1][3][4][5], 후확률(posterior probability) 기반 신뢰도 척도[1][2]이다. 현재 발화검증이 성능이 좋은 신뢰도 척도는 LRT 기반 신뢰도 척도와 후확률 기반 신뢰도 척도이고, 현재까지 두 방식의 성능은 서로 비슷한 결과를 내고 있다. 최근에는 발화검증의 성능을 높이기 위해 FLDA(Fisher's Linear Discriment Analysis), 신경망, SVM(Support Vector Machine), 결정트리, Bayesian[1] 등과 같은 기술을 사용하여 여러 개의 신뢰도 척도들을 통합하여 사용하는 발화검증 기술들이 개발되었다.

본 논문에서는 기존에 개발된 발견법적 기반 신뢰도 척도와 LRT 기반 신뢰도 척도에 log-likeliood 수열을 사용한 발견법적 기반 신뢰도 척도와 음소모델간의 거리 및 음소 종류에 따른 가중치를 적용한 LRT 기반 신뢰도 척도를 제안하고 추가해서 SVM을 통한 인식결과의 수락/거절을 결정하는 발화검증 시스템을 구성하였

다. 더불어 환경에 강인한 새로운 신뢰도 척도를 개발하기 위해 단어별, 음소별 신뢰도 척도의 분석, 사용된 신뢰도 척도간의 상관관계, 상관관계에 따른 SVM의 성능을 다양한 실험을 통해 분석하였다. 본 논문의 구성은 다음과 같다. 2장에서는 발화검증 시스템을 설명하고, 3장에서는 본 논문에서 사용된 신뢰도 척도들에 대해 기술하고, 4장에서는 ETRI에서 제공된 발화검증용 음성데이터를 사용한 실험과 결과를 기술하고, 5장에서는 결론과 더불어 신뢰도 척도 개발에 대한 향후 방향과 과제에 대한 제안을 한다.

2. 발화검증 시스템

발화검증은 발화된 음성으로부터 인식된 결과가 어느 정도의 신뢰도를 갖고 있는가에 대한 척도를 계산하는 방법이다. 이를 사용하여 비인식대상어휘를 거절하고, 오인식된 인식대상어휘에 대해서도 신뢰도에 따라 거절한다. 최근에 사용되고 있는 발화검증 시스템은 신뢰도 척도 하나만을 사용하지 않고 다양한 신뢰도 척도를 통합하여 사용한다. <그림 1>은 전체적인 발화검증 시스템을 보여 주고 있다. 발화검증을 위한 신뢰도 척도들은 입력 음성파형으로부터 직접 구하거나 음성인식기의 인식결과부터 구할 수 있고, 인식결과로부터 구할 수 있는 신뢰도 척도에는 문맥정보, 대화정보와 같은 고수준의 정보를 이용한 high-level 신뢰도 척도와 발견법적 기반 신뢰도 척도, LRT 기반 신뢰도 척도, 후확률 기반 신뢰도 척도와 같은 low-level 신뢰도 척도가 있다. 추출된 다양한 신뢰도 척도들로부터 FLDA, 신경망, SVM, 결정트리, Bayesian를 사용해서 최종적으로 발화검증을 한다. 본 논문에서는 전체적인 발화검증 시스템보다는 low-level 신뢰도 척도의 개발에 중점을 두었기 때문에 <그림 1>에서 네모상자 안에 있는 시스템만을 구현하였고, 발화검증 분류기에 사용한 알고리즘은 SVM이다.



<그림 1> 발화검증 시스템

2.1 음성인식기

본 논문에서 구현된 발화검증 시스템은 기본적으로 HMM(Hidden Markov Model) 모델을 사용하여 Viterbi 탐색을 하는 음성인식기를 통해 얻은 인식결과를 바탕으로 다양한 신뢰도 척도를 계산한다. 본 논문에서 사용된 음성인식기는 ECHOS-1.0[1][8]이다. 일반적으로 음성인식기는 구성된 인식 네트워크 안에서 최고의 확률을 갖는 최상의 경로만을 찾는다. 하지만 본 논문에서는 단어의 다양한 신뢰도 척도를 구하기 위해 네트워크 구조를 플랫렉시콘으로 사용하고, 1-best / N-best 인식결과를 모두 사용한다. 그리고 음소단위의 인식정보를 얻어서 음소단위의 신뢰도 척도를 계산할 수 있도록 하고, 또한 음소단위의 인식결과로부터 다시 인식된 음소별로 재인식하여 각 프레임마다 최대의 likelihood를 갖은 state와 likelihood 수열을 얻는다. 음성인식기에 사용되는 음향모델은 MFCC (Mel-Frequency Cepstral Coefficients) 12차에 log-에너지를 더한 13차에 delta 13차, delta-delta 13차로 만들어진 39차의 특징벡터를 사용하여 훈련된 HMM을 사용한다. 입력 음성의 규격은 16 kHz, 16 bits, PCM 데이터이다.

2.2 Support Vector Machine

다양하게 구해진 신뢰도 척도들을 통합하는 방식은 여러 가지가 있지만, 본 논문에서는 SVM을 사용하여 신뢰도 척도들을 통합하였다. 사용되고 있는 SVM 툴은 LIBSVM[2][9]이다. SVM 모델은 음향모델 훈련에 사용되지 않고, 테스트 셋에도 포함되지 않는 인식대상어휘에 해당되는 음성 데이터베이스뿐만 아니라 비인식대상어휘로 수집된 음성 데이터베이스로 훈련이 된다. SVM에 사용되는 kernel은 $\exp(-\gamma(u-v)^2)$(radial basis function)이다. SVM은 <그림 1>에서 발화검증 분류기 부분에서 사용된다. 미리 SVM 훈련 셋으로 훈련된 SVM모델을 가지고 있고, 입력 음성으로부터 구해진 다양한 신뢰도 척도들을 입력으로 받아 SVM모델을 사용하여 최종 수락/거절 결과를 출력한다.

3. 신뢰도 척도

신뢰도 척도를 나누는 기준에 따라 여러 가지 방법으로 분류할 수 있다. 본 논문에서는 신뢰도 척도를 구하는 알고리즘에 따라 발견법적 기반 신뢰도 척도,

1) ECHOS(Easy Compact Hangeul Object-oriented Speech recognizer): SiTEC 용역과제로 ICU, KAIST, 충북대에서 공동 개발된 한국어 음성인식 플랫폼.
2) LIBSVM: http://www.csie.ntu.edu.tw/~cjlin/libsvm의 공개 SVM 툴.

LRT 기반 신뢰도 척도, 그리고 후확률 기반 신뢰도 척도로 나눈다. 본 논문에서는 후확률 기반 신뢰도 척도는 구현되지 않았다.

3.1 발견법적 기반 신뢰도 척도

발견법적 기반 신뢰도 척도는 baseline 인식기로부터 얻어지는 일반적인 정보로부터 얻어진다. 예를 들면, 인식된 likelihood을 정규화하거나, N-best 리스트의 인식단어의 개수, 또는 상위 몇 개의 likelihood의 평균, 또는 인식된 단어, 음소, state 지속시간 정보등을 통해 얻을 수 있는 신뢰도 척도를 말한다. 본 논문에서는 발견법적 신뢰도 척도로 인식된 결과의 log-likelihood를 정규화한 값과 N-best 인식단어의 log-likelihood 비율, 그리고 프레임별 log-likelihood 수열로부터 얻을 수 있는 log-likelihood 요동률을 신뢰도 척도로 사용하였다. log-likelihood를 정규화한 값과 N-best 인식단어의 log-likelihood 비율은 기존의 많은 발화검증 시스템에서 많이 사용되고 있고, log-likelihood 요동률은 정상적인 인식이 이루어 졌을 때 log-likelihood의 변화가 전체적으로 매끄러운 곡선을 가지고 그렇지 않을 경우 다른 음향모델과 log-likelihood를 계산해야 되기 때문에 log-likelihood값이 심하게 요동을 친다는 점을 이용하여 본 논문에서 제안하였다.

1-best 인식단어의 log-likelihood의 정규화를 구하는 식은 다음과 같다.

$$MLL(w) = \frac{1}{\tau(w)} logP(X|\lambda_w) = \frac{1}{\tau(w)} \sum_{t=\tau_s(w)}^{\tau_e(w)} \log P(x_t|s_t) \tag{1}$$

여기서 $\tau(w)$는 인식된 단어의 프레임 개수, λ_w는 단어 w의 HMM 모델, X는 입력음성의 특징벡터열을 나타낸다. $\tau_s(w)$는 인식단어의 시작 프레임, $\tau_e(w)$는 인식단어의 마지막 프레임을 나타낸다. MLL로부터 간단히 얻을 수 있는 신뢰도 척도는 다음과 같다.

$$C_{MLL}(w) = MLL(w) \tag{2}$$

식 (2)의 신뢰도 척도는 매우 단순하며 발화검증 성능이 좋지 않고 환경에 매우 민감하지만 log-likelihood로 가장 간단히 빨리 판단할 수 있는 장점을 지니고 있다. N-best 인식결과로부터 얻을 수 있는 신뢰도 척도는 N-best word log-likelihood 비율이고 다음과 같이 구해진다.

$$C_{NBD}(w) = \frac{1}{N-1}\sum_{n=2}^{N}(MLL(w^{(1)}) - MLL(w^{(n)})) \tag{3}$$

여기서 $w^{(n)}$는 n-번째로 높은 확률을 가지고 인식된 단어를 말한다. 정상적으로 음성인식이 이루어질 경우 1-best 인식단어와 n-best 인식단어 사이의 거리는 오인식 되었을 때 보다 크다는 점으로 이용한 신뢰도 척도이다. log-likelihood 요동률은 다음과 같이 구해진다.

$$C_{LLF}(ph) = \frac{1}{\tau(ph)}\sum_{t=\tau_s(ph)}^{\tau_e(ph)}(\log P(x_t|s_t) - MLL(ph))^2 \tag{4}$$

$$C_{LLF}(w) = \frac{1}{n_p(w)}\sum_{j=1}^{n_p(w)} sigmoid(C_{LLF}(ph_j)) \tag{5}$$

식 (5)에서 사용된 sigmoid 함수는 $1/(1+\exp(\alpha C_{LLF}(ph_j)-\beta))$이고, $\alpha = 1.0$, β는 훈련된 모든 음소의 log-likelihood 요동률의 평균값을 사용하였다. 발견법적 기반 신뢰도 척도는 LRT 기반 신뢰도 척도나 후확률 기반 신뢰도 척도보다 발화검증 성능이 많이 떨어진다. 그래서 일반적으로 발견법적 기반 신뢰도 척도는 후자의 신뢰도 척도들의 보조 수단으로 많이 사용된다. 본 논문에서 추가적으로 N-best monophone 반모델을 사용하여 LRT를 구할 때 인식된 likelihood보다 큰 likelihood를 갖는 monophone의 개수를 나타내는 신뢰도 척도를 제안하였는데 LRT 기반 신뢰도 척도와 유사한 개념을 사용하고 있지만 likelihood 비율을 사용하지 않았기 때문에 발견법적 기반 신뢰도 척도로 분류하였다. C_{VAC}는 정상적으로 인식을 되었을 경우, 각 음소에서 인식된 likelihood가 반음소모델의 likelihood보다 대체적으로 크다는 사실과 오인식 되었을 경우 인식된 음소의 likelihood보다 큰 likelihood를 갖는 반음소모델의 개수가 많아진다는 점을 이용한 신뢰도 척도이다. C_{VAC}를 구하는 식은 다음과 같다.

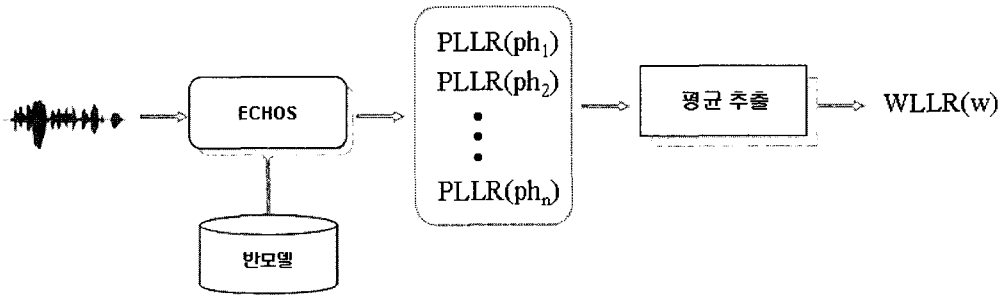
$$C_{VAC}(ph) = \sum_{m \neq mono(ph)}(MLL(m) > MLL(ph)\,?\,1:0) \tag{6}$$

$$C_{VAC}(w) = \frac{1}{n_p(w)}\sum_{j=1}^{n_p(w)} C_{VAC}(ph_j) \tag{7}$$

$mono(ph)$는 triphone ph의 base monophone을 나타내고, m은 반음소모델의 monophone 이름을 나타낸다.

3.2 LRT 기반 신뢰도 척도

LRT 기반 신뢰도 척도는 통계적 가설 검증 방식[1]에 기반을 두고 있다. 기본 음성인식기로부터 얻어진 음소단위의 인식결과로부터 반모델을 사용하여 log-likelihood 비율을 구한다. <그림 2>은 LRT 기반 신뢰도 척도를 계산하는 알고리즘을 나타내고 있다. 반모델은 생성하는 방법에 따라 All Mixtures 반모델[7], Cohort 반모델[7], N-best monophone 반모델[4], 적응 반모델[7] 등이 있다. 본 논문에서는 앞의 세 가지 반모델을 사용한다. 일반적으로 LRT 기반 신뢰도 척도는 인식된 각 음소에 대해 음소 단위 log-likelihood 비율(PLLR)을 구한 다음 산술, 기하, 조화 평균을 하여 단어 단위 log-likelihood 비율(WLLR)을 구한다.



<그림 2> LRT 기반 신뢰도 척도 계산 알고리즘

PLLR을 구하는 방식은 다음과 같다.

$$PLLR(ph) = \frac{\log P(X_{ph}|\lambda_{ph}) - \log P(X_{ph}|\overline{\lambda_{ph}})}{\tau(ph)} \tag{8}$$

여기서 X_{ph}는 인식된 음소에 해당되는 입력 특징벡터열, λ_{ph}는 인식된 음소의 음향모델, $\overline{\lambda_{ph}}$는 인식된 음소에 해당되는 반모델, $\tau(ph)$는 인식된 음소의 프레임 수를 나타낸다. 실제로 사용하게 될 PLLR은 다음과 같은 형태의 음소 단위의 신뢰도 척도를 사용한다.

$$C_{LLR}(ph) = \frac{1}{1+\exp(-\alpha PLLR(ph) - \beta)} \tag{9}$$

α와 β는 실험을 통해 정해진다. 각 음소에 대해 구해진 $C_{LLR}(ph)$를 산술, 기하, 조화 평균을 취하여 단위 단위의 신뢰도 척도를 구한다. 기하평균은 PLLR의 차이가 큰 단어에 대해 산술평균 보다 덜 민감하도록 하는 효과가 있고, 조화평균

은 PLLR의 차이에 가장 덜 민감한 평균을 가진다. 단어 단위의 LRT 기반 신뢰도 척도를 계산하는 식은 다음과 같다.

$$\text{산술평균} : C_{LLR}^{A}(w) = \frac{1}{n_p(w)} \sum_{j=1}^{n_p(w)} C_{LLR}(ph_j) \tag{10}$$

$$\text{기하평균} : C_{LLR}^{G}(w) = \exp\left(\frac{1}{n_p(w)} \sum_{j=1}^{n_p(w)} \log(C_{LLR}(ph_j)+1) \right) \tag{11}$$

$$\text{조화평균} : C_{LLR}^{H}(w) = \frac{n_p(w)}{\sum_{j=1}^{n_p(w)} C_{LLR}(ph_j)} \tag{12}$$

여기서 $n_p(w)$는 단어 w를 이루는 음소의 개수를 나타낸다.

3.3. 음소정보를 이용한 LRT 기반 신뢰도 척도

앞 절에 제안된 LRT 기반 신뢰도 척도는 각 음소별 PLLR을 단순히 산술, 기하, 조화 평균을 구하여 단어 단위의 신뢰도 척도를 구하였다. 하지만 음향모델을 훈련하는 과정에서 보면 어떤 음소는 훈련 음성 데이터가 충분하고 발성이 잘 이루어지기 때문에 매우 정교한 음향모델이 생성되기도 하고, 어떤 음소는 훈련 음성 데이터의 부족으로 덜 정교한 음향모델이 생성되기도 한다. 그리고 어떤 음소는 대부분의 사람이 발성하는 방법이 비슷하고, 어떤 음소는 발성하는 사람에 따라 변동이 심한 음소가 있다. 또한 음성인식기를 통해 인식되는 과정으로 보면 어떤 음소의 음향모델에서는 비터비 탐색이 매끄럽게 이루지기도 하지만, 어떤 음소의 음향모델에서는 비터비 탐색이 혼란이 심해 likelihood의 변화가 심한 경우가 발생하기도 한다. 이러한 현상은 반모델에서도 같은 경향으로 일어난다. 따라서 같은 가중치를 두고 PLLR로부터 WLLR을 계산하는 방법에는 부족한 점이 있다. 이러한 문제점을 해결하기 위해 본 논문에서는 음소정보를 이용한 LRT 기반 신뢰도 척도를 계산하는 방법을 제안한다.

첫 번째로 음소간의 거리를 이용한 WLLR을 구하는 방법이다. 여기서 사용되는 음소간의 거리는 두 음소에 해당되는 음향모델로부터 서로의 likelihood를 평균해서 구한다. 훈련이 잘 되거나 발성간의 변이가 적은 음향모델은 자기 자신과의 음향모델간의 거리를 계산해 보면 높은 likelihood를 가지는 것을 알 수 있고, 그렇지 않는 경우는 낮는 likelihood를 가지는 것을 알 수 있다. 대체적으로 유성음의 경우 높은 likelihood를 가진다는 것을 훈련된 음향모델로부터 확인할 수 있었다. 음소간 거리를 사용한 산술평균의 WLLR를 구하는 식은 다음과 같다.

$$C_{LLR}^{DA}(w) = \frac{1}{n_p(w)} \sum_{j=1}^{n_p(w)} \frac{C_{LLR}(ph_j)}{D(mono(ph_j))} \tag{13}$$

여기서 음소간의 거리는 다음과 같이 구해진다.

$$D(mono(ph_j)) = \sum_{k=1}^{K} m_k \log P(o_k|\lambda_{mono(ph_j)}) \tag{14}$$

본 논문에서는 인식된 결과인 ph_j는 triphone이다. 여기서 $mono(ph_j)$는 triphone ph_j의 base monophone을 나타낸다. o_k는 $mono(ph_j)$ 음향모델의 k번째 mixture의 mean 특징벡터를 나타내고, m_k는 k번째 mixture의 가중치이다. K는 음향모델의 mixture 개수이다.

두 번째로 지식기반의 음소정보를 이용하여 WLLR를 구하는 방법이다. 일반적으로 유성음의 경우에는 인식될 때 likelihood가 위로 볼록한 형태를 가지며 무성음보다 상대적으로 높은 값을 갖는다. 그리고 유성음 중에서도 장모음과 이중모음이 보다 또렷이 이러한 경향을 나타낸다. 무성음의 경우에는 likelihood이 변화가 심하고 예측하기 힘들 뿐만 아니라 상대적으로 낮은 값을 갖는다. 음소간 거리정보는 훈련을 통해 얻어지므로 테스트 입력음성에 대해 불일치가 발생한다. 이를 보상하기 위한 가중치가 필요한데 지식기반의 음소분류에 의한 가중치를 제안하였다. 음소 분류는 지식기반에 의해 수동적으로 여섯 클래스로 분류하여 각 클래스마다 다른 가중치를 적용하여 WLLR을 구하였다. 여섯 클래스와 각각의 가중치는 다음과 같다. 장모음(1.5), 단모음(0.8), 이중모음(1.3), 비음(1.3), 받침 'o'(1.6), 자음(1.0). 위의 두 가지 음소정보를 이용한 산술평균 WLLR를 구하는 식은 다음과 같다.

$$C_{LLR}^{WDA}(w) = \frac{1}{n_p(w)} \sum_{j=1}^{n_p(w)} \frac{C_{LLR}(ph_j)}{\gamma(mono(ph_j))D(mono(ph_j))} \tag{15}$$

여기서 $\gamma(mono(ph_j))$는 ph_j의 base monophone에 해당되는 가중치를 나타낸다.

4. 실험 및 결과

4.1. 데이터베이스

본 연구에서는 한국전자통신연구원에서 수집된 발화검증용 음성 데이터베이스

를 사용하였다. 인식용 음향 모델을 훈련하기 위해 조용한 사무실 환경에서 수집된 800명의 화자로부터 발성된 79,948개의 고립단어 음성 데이터베이스를 사용하였다. 음성 특징 파라미터는 MFCC 12차에 로그 에너지를 더한 13차, delta 13차, delta-delta 13차를 합친 39차를 사용하였고, CMN(Ceptral Mean Normazliation)을 적용하였다. 훈련된 음향모델은 3개의 state와 7개의 가아시안 분포를 갖는다. 발화검증 실험을 위해 훈련에 사용되지 않은 1,000단어 규모의 실험용 고립단어 음성 데이터베이스를 사용하였다. SVM모델을 생성하기 위해 사용한 데이터베이스와 발화검증 성능 실험에 사용된 데이터베이스는 훈련에 사용된 데이터베이스와 다른 저가의 마이크를 사용하여 수집이 되었다. 전체 데이터베이스에 대한 설명은 <표 1>과 같다. 음성인식기(ECHOS-1.0)의 테스트 셋에 대한 인식성능은 95.90%이다.

<표 1> 발화검증용 고립단어음성 데이터베이스

구 분		내 용	생성물
훈련 셋	발화 형태	고립 단어	음향모델, 반모델
	단어 수	10,280	
	화자 수	800 명	
	발화 수	79,948 발화	
SVM 훈련 셋	발화 형태	고립 단어	SVM 모델
	단어 수	인식대상어휘 : 999 비인식대상어휘 : 400	
	화자 수	인식대상어휘 : 80 명 비인식대상어휘 : 60 명	
	발화 수	인식대상어휘 : 7,995 발화 비인식대상어휘 : 3,200 발화	
테스트 셋	발화 형태	고립 단어	발화검증 성능
	단어 수	인식대상어휘 : 999 비인식대상어휘 : 400	
	화자 수	인식대상어휘 : 20 명 비인식대상어휘 : 8 명	
	발화 수	인식대상어휘 : 1,998 발화 비인식대상어휘 : 800 발화	

4.2. 실험 방법

본 논문에서는 발견법적 기반 신뢰도 척도 4가지와 반모델에 따른 LRT 기반 신뢰도 척도 9가지에 대해 발화검증 성능 실험을 하였고, 제안된 음소정보를 이용한 LRT 기반 신뢰도 척도에 대한 실험은 N-best monophone 반모델에 대해 성능검증을 하였다. 모든 발화검증 결과를 단어별, 음소별로 분석하고 신뢰도 척도 사이의 상관관계에 대한 조사도 하였다. 그리고 신뢰도 척도의 상관관계에 따른 SVM의 발화검증 성능 실험도 하였다. 발화검증 성능 결과는 FA(False Alarm)과

FR(False Rejection)은 같은 EER(Equal Error Rate)로 측정하였다. FA는 오인식된 것이 정상적으로 인식된 것으로 나타내고, FR는 정상적으로 인식된 것이 신뢰도가 낮아 거절된 경우를 말한다.

4.3. 실험 결과

발견법적 기반 신뢰도 척도에 대한 발화검증 성능은 <표 2>와 같다. <표 2>에서 보듯이 발견법적 기반 신뢰도 척도의 발화검증 성능은 좋지 않다. 그렇기 때문에 발견법적 신뢰도 척도는 발화검증 시스템에서 보조적인 수단으로 많이 사용된다.

<표 2> 발견법적 기반 신뢰도 척도에 대한 발화검증 성능

신뢰도 척도	EER(%)	FR(%)	FA(%)
C_{MLL}	30.42	30.33	30.50
C_{NBD}	33.18	33.23	33.13
C_{LLF}	39.16	38.94	39.38
C_{VAC}	12.76	12.76	12.75

LRT 기반 신뢰도 척도에 대한 발화검증 성능은 <표 3>과 같다. 식 (9)에서 PLLR을 구할 때 사용된 함수에서 $\alpha=-1$, β는 모든 비인식대상어휘의 모든 음소에 대한 PLLR의 평균값을 사용하였다. 본 실험에서 β는 -2.53이였다. LRT 기반 신뢰도 척도의 발화검증 성능은 반모델이 달라도 성능차이는 크게 보이고 있지 않지만 Cohort 반모델을 사용한 것이 가장 좋은 성능을 보이고 있다. 발견법적 기반 신뢰도 척도보다는 훨씬 우수한 성능을 보이고 있다. 평균을 취하는 방식에 따는 성능차이도 크게 나타나지 않지만, PLLR의 차이에 민감한 산술평균이 대체적으로 좋은 성능을 보이고 있음을 알 수 있다. 기하평균이 산술평균보다 항상 작은 값을 갖기 때문에 높은 PLLR를 갖는 음소가 있을 경우 기하평균에서는 WLLR에 큰 영향을 주지 않지만 산술평균에서는 WLLR에 많이 반영이 된다. 음소별로 PLLR을 비교 분석한 결과 음소마다 평균 PLLR이 차이를 보이고 있다. 예를 들면, "E-N" 음소의 평균 PLLR은 1.04이고, "H-E+N" 평균 PLLR은 4.63이였다. 가령 "H-E+N"이 높은 값을 가지고 있어도 기하평균의 WLLR이 산술평균의 WLLR보다 작기 때문에 오인식될 확률이 좀 더 크다는 것을 알 수 있다. 오인식의 경우 반대의 경우도 발생하기 때문에 항상 산술평균이 좋다고는 할 수 없지만 음소에 따라 PLLR이 WLLR에서 다른 비중을 가지고 있음을 유추해 낼 수 있다.

<표 3> LRT 기반 신뢰도 척도에 대한 발화검증 성능

반모델 형태	평균 형태	EER(%)	FR(%)	FA(%)
AllMixtures	산술	11.33	11.41	11.25
	기하	11.58	11.66	11.25
	조화	11.77	11.66	11.88
N-Best Monophone	산술	11.06	11.11	11.00
	기하	11.22	11.31	11.13
	조화	11.52	11.41	11.63
Cohort	산술	10.95	11.03	10.88
	기하	11.18	11.20	11.16
	조화	10.82	10.80	10.84

음소정보를 이용한 LRT 기반 신뢰도 척도에 대한 발화검증 성능은 <표 4>와 같다. 음소정보를 이용하기 때문에 음소정보가 어느 정도 반영이 되어 있는 N-best monophone 반모델에 대해 성능 검증을 하였다. AllMixture 반모델을 이용한 LRT 기반 신뢰도 척도에 음소정보를 이용하였을 때는 발화검증의 성능 향상이 미미하였다. 그 이유는 AllMixture 반모델의 경우에는 모든 음소에 대한 정보가 부분적으로 들어가 있기 때문이다. 음소간 거리 정보만을 사용하였을 때는 EER이 상대적으로 8% 정도 성능 향상을 보이고 있고, 두 가지 정보 모두를 사용하였을 경우에는 상대적으로 13% 정도의 성능 향상을 보여주고 있다. 하지만 PLLR간의 차이에 덜 민감한 조화평균의 경우에는 오히려 성능이 떨어지는 것으로 알 수 있다. 즉 음소간 정보가 PLLR을 구할 때 반영이 되어도 조화평균을 취하는 순간 그 장점이 사라지기 때문으로 분석되어 진다.

<표 4> 음소정보를 이용한 LRT 기반 신뢰도 척도에 대한 발화검증 성능

반모델 형태	평균 형태	EER(%)	FR(%)	FA(%)
N-Best Monophone	산술	11.06	11.11	11.00
	기하	11.22	11.31	11.13
	조화	11.52	11.41	11.63
N-Best Monophone (음소간 거리만 사용)	산술	10.13	10.01	10.25
	기하	10.31	10.36	10.25
	조화	11.88	11.76	12.00
N-Best Monophone (모든 음소정보 사용)	산술	9.74	9.86	9.63
	기하	9.77	9.66	9.88
	조화	11.63	11.51	11.75

최근의 발화검증 시스템은 하나의 신뢰도 척도를 가지고 발화검증을 하지 않고 여러 종류의 신뢰도 척도를 통합하여 발화검증의 성능을 높이고 있다. 본 논문

에서는 SVM을 사용하여 앞에서 언급된 신뢰도 척도를 통합하여 성능을 검증하였다. 모든 신뢰도 척도를 통합해서 사용하는 것이 바람직하나 SVM 모델 훈련 시간과 SVM 모델에 사용될 메모리가 증가함에도 불구하고 일부 신뢰도 척도를 사용한 것보다 잇점이 없을 경우가 발생한다. 이럴 경우 필요한 신뢰도 척도를 선택할 수 있는 방법이 필요하다. 신뢰도 척도의 상관관계에 따른 SVM 성능을 비교 분석함으로써 SVM에 사용될 신뢰도 척도를 효과적으로 선택할 방법을 얻을 수 있다.

<표 5>는 신뢰도 척도간의 상관관계를 나타내고 있다. 발견법적 기반 신뢰도 척도와 LRT 기반 신뢰도 척도의 상관관계가 작음을 알 수 있고, 평균하는 방식에 따른 신뢰도 척도간의 상관관계는 매우 높다는 것을 알 수 있다.

<표 5> 신뢰도 척도간의 상관관계

신뢰도 척도	C_{MLL}	C_{NBD}	C_{VAC}	C_{LLR}^{A}	C_{LLR}^{G}	C_{LLR}^{H}
C_{MLL}	1.000	0.673	0.841	0.836	0.835	0.840
C_{NBD}		1.000	0.792	0.797	0.793	0.803
C_{VAC}			1.000	0.971	0.971	0.957
C_{LLR}^{A}				1.000	0.995	0.972
C_{LLR}^{G}					1.000	0.968
C_{LLR}^{H}						1.000

<표 6>은 SVM에 사용된 신뢰도 척도에 따른 발화검증 성능을 보여주고 있다. SVM에 사용된 신뢰도 척도들의 조합은 상관관계에 따라 분류 실험하였고, 단일 신뢰도 척도로써 가장 성능이 좋은 C_{LLR}^{A}를 기준으로 조합을 하였다. 확실히 모든 신뢰도 척도들을 사용하는 것이 가장 좋은 성능을 보이고 있고, 서로의 상관관계가 큰 신뢰도 척도들을 사용할 경우 발화검증의 성능이 크게 향상되고 있지 않음을 보여주고 있다. 하지만 상관관계가 작은 신뢰도 척도들을 소수만 사용하여도 발화검증의 성능이 상관관계가 큰 신뢰도 척도들을 많이 사용한 것보다 좋은 성능을 보여주고 있다. 따라서 적은 신뢰도 척도들을 사용해야 될 경우 SVM에 사용될 신뢰도 척도의 조합은 먼저 가장 성능이 좋은 신뢰도 척도를 기준으로 서로의 상관관계가 작은 신뢰도 척도를 선택 조합하는 것이 가장 바람직하다는 것을 알 수 있다.

<표 6> 신뢰도 척도의 선택에 따른 SVM 발화검증 성능

신뢰도 척도	EER(%)	FR(%)	FA(%)
$C_{LLR}^{A}+C_{LLR}^{G}+C_{LLR}^{H}$	9.11	9.09	9.13
$C_{LLR}^{A}+C_{VAC}$	9.00	9.04	8.96
$C_{LLR}^{A}+C_{MLL}$	8.76	8.80	8.72
N-Best Monophone 반모델을 사용한 모든 LRT 기반 신뢰도 척도 + 모든 발견법적 기반 신뢰도 척도	8.08	8.15	8.01

5. 결 론

본 논문에서는 발화검증을 위해 기존의 다른 연구논문에서 제안된 발견법적 신뢰도 척도들과 많이 사용되고 있는 반모델을 사용한 LRT 기반 신뢰도 척도들의 구현과 더불어, 각 프레임별 log-likelihood 수열을 이용한 신뢰도 척도를 제안하였고, 음소간의 음향모델 거리 또는 음소 형태 정보를 이용한 LRT 기반 신뢰도 척도를 제안하였다. 본 논문의 실험을 통해 음소정보를 이용한 LRT 기반 신되도 척도의 발화검증 성능이 기존의 LRT 기반 신되도 척도에 비해 상대적으로 13% 정도 성능이 향상되었음을 보여 주고 있다. SVM을 사용한 발화검증 실험에서는 확실히 하나의 신뢰도 척도를 사용하는 것보다 여러 개의 신뢰도 척도들을 통합해서 사용하는 것이 좋은 성능을 보이고 있으며, 신뢰도 척도간의 상관관계가 작은 신뢰도 척도들을 통합하는 것이 발화검증에 보다 좋은 성능이 보이고 있음을 확인할 수 있었다. 다양한 신뢰도 척도를 통한 단어별, 음소별 발화검증 성능의 분석으로 통해 음소단위에서 계산된 신뢰도 척도를 단어 단위의 신뢰도 척도로 연산하는 방식의 한계점들을 확인할 수 있었고, 인식대상 어휘 수에 따라 다른 신뢰도 척도를 적용할 필요성이 요구되었다. 따라서 향후 이러한 한계점을 극복할 신뢰도 척도의 연구와 후확률 신뢰도 척도에 대한 구현 및 연구, 그리고 환경에 강인한 신뢰도 척도에 대한 연구가 필요하다.

참 고 문 헌

[1] H. Jiang, "Confidence measure for speech recognition: a survey", *Speech Communication*, Vol. 45, Issue 4. pp. 455-470, 2005.

[2] F. Bessel, R. Schlüter et al., "Confidence measures for large vocabulary continuous speech recognition", *IEEE Transaction on Speech and Audio Processing*, Vol. 9, No. 3, pp. 288-298, 2001.

[3] M. G. Rahim, C.-H. Lee, B.-H. Juang, "Discriminative utterance verification for connected digits recognition", *IEEE Transaction on Speech and Audio Processing*, Vol. 5, No. 3, pp. 266-277, 1997.

[4] K.-S. Moon, Y.-J. Kim et al., "Out-of-vocabulary word rejection algorithm in Korean variable vocabulary word recognition", *IEEE International Symposium on Circuits and Systems*, Vol. 5, pp. 53-56, 2000.

[5] C.-H. Lee, "A tutorial on speaker and speech verification", *IEEE Nordic Signal Processing Symposium*, 1998.

[6] M.-W. Koo, S.-J. Lee, "An utterance verification system based on subword modeling for a vocabulary independent speech recognition system", *Proc. Eurospeech99*, Vol. 1, pp 287-290, 1999.

[7] 강점자, 전형배, "SVM 기반 멀티플 반모델을 사용한 발화검증", *추계한국음향학회*, 2005.

[8] 권오욱, 권석봉 외, "한국어 음성인식 플랫폼(ECHOS)의 개선 및 평가", *말소리*, 59호, pp. 53-67, 2006.

[9] LIBSVM home page, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

접수일자 : 2006년 11월 10일
게재결정 : 2006년 12월 20일

▶ 권석봉(Suk-bong Kwon)
주소: 305-732 대전광역시 유성구 문지로 119번지 한국정보통신대학교
소속: 한국정보통신대학교(ICU) 음성인식기술연구실
전화: 042) 866-6221
FAX: 042) 866-6245
E-mail: sbkwon@icu.ac.kr

▶ 김회린(Hoirin Kim)
주소: 305-732 대전광역시 유성구 문지로 119번지 한국정보통신대학교
소속: 한국정보통신대학교(ICU) 음성인식기술연구실
전화: 042) 866-6139
FAX: 042) 866-6245
E-mail: hrkim@icu.ac.kr

▶ 강점자(Jeomja Kang)
주소: 305-350 대전광역시 유성구 가정동 161번지
소속: 한국전자통신연구원 음성처리연구팀
전화: 042) 860-4880
FAX: 042) 860-4889
E-mail: jjkang@etri.re.kr

▶ 구명완(Myong-Wan Koo)
주소: 137-792 서울특별시 서초구 우면동 17 KT
소속: KT 미래기술연구소 HCI 연구담당
전화: 02) 526-5090
FAX: 02) 526-6775
E-mail: mwkoo@kt.co.kr

▶ 류창선(Chang-Sun Ryu)
주소: 137-792 서울특별시 서초구 우면동 17 KT
소속: KT 미래기술연구소 HCI연구담당 미디어처리연구부
전화: 02) 526-6759
FAX: 02) 526-6775
E-mail: csryu@kt.co.kr