

하이브리드 신뢰도를 이용한 제한 영역 핵심어 검출 성능향상

Improvement of Domain-specific Keyword Spotting Performance Using Hybrid Confidence Measure

이 경 록*, 서 현 철*, 최 승 호**, 최 승 호***, 김 진 영*
(Kyong-Rok Lee*, Hyun-Chul Seo*, Seung-Ho Choi**, Seung-Ho Choi***, Jin-Young Kim*)

* 전남대학교 전자공학과 & RRC HECS, ** 동신대학교 정보통신공학과, *** 서울산업대학교 전자공학과
(접수일자: 2002년 2월 21일; 수정일자: 2002년 8월 19일; 채택일자: 2002년 8월 26일)

본 논문에서는 기존의 RLJ-신뢰도 (RLJ-confidence measure)와 정규화 신뢰도 (normalized CM)의 단점을 보완하기 위해 ACM (Anti-filler CM)을 제안하였고, HCM (hybrid CM)을 이용하여 기존의 NCM과 제안한 ACM을 통합하였다. 제안된 ACM은 기존 신뢰도의 단점 중 하나인 오인증 (FA: false acceptance)의 원인이 반음소 모델의 구성방법에 있다고 보고 음소 인식기를 이용하여 실제 음소 수열을 추정한 다음, 이를 반음소 모델로 정의하고 신뢰도를 계산하였다. 두 가지 신뢰도의 특성을 살펴보면, 기존 NCM은 오거부 (FR: false rejection)에 좋은 성능을 보이고, 제안한 ACM은 FA에서 좋은 성능을 보여 두 신뢰도가 상보적인 특성을 가진다. 이를 이용하여 두 가지 신뢰도를 가중치 벡터 α 를 이용하여 통합하고 이를 합성 신뢰도 (HCM: Hybrid CM)라고 정의하였다. 실험결과 미검출율 (MDR: missed detection rate) 10% 부근에서, HCM 적용시에 0.219 FA/KW/HR (false alarm/keyword/hour)로서 NCM 단독사용에 비해 성능이 22% 향상되었다.

핵심용어: 합성 신뢰도, 반필러 신뢰도, 정규화 신뢰도, 후처리, 핵심어 검출
투고분야: 음성처리 분야 (2.5)

In this paper, we proposed ACM (Anti-filler confidence measure) to compensate shortcoming of conventional RLJ-CM (RLJ-CM) and NCM (normalized CM), and integrated proposed ACM and conventional NCM using HCM (hybrid CM). Proposed ACM analyzes that FA (false acceptance) happens by the construction method of anti-phone model, and presumed phoneme sequence in actuality using phoneme recognizer to compensate this. We defined this as anti-phone model and used in confidence measure calculation. Analyzing feature of two confidences measure, conventional NCM shows good performance to FR (false rejection) and proposed ACM shows good performance in FA. This shows that feature of each other are complementary. Use these feature, we integrated two confidence measures using weighting vector α . And defined this as HCM. In MDR (missed detection rate) 10% neighborhood, HCM is 0.219 FA/KW/HR (false alarm/keyword/hour). This is that Performance improves 22% than used conventional NCM individually.

Keywords: Hybrid confidence measure, Anti-filler confidence measure, Normalized confidence measure, Postprocessing, Keyword spotting

ASK subject classification: Speech signal processing (2.5)

I. 서론

연속된 음성 중 핵심어의 출현을 인식하는 핵심어 검출기는 연속 음성인식기에 비해서 검출부의 구조가 비교적 간단하고 연산비용이 저렴하다. 때문에 특정분야에서 연속 음성 인식기보다 우월한 특성을 가진다.

이러한 핵심어 검출기는 연속 입력음성 중에서 핵심어만을 검출하는 시스템의 특성상 FA (false alarm)이 다수 발생하고, 이를 처리하는 후처리가 전체 성능에 큰 영향을 미친다.

핵심어 검출에서 발생하는 여러의 패턴에는 핵심어를 잘못 검출하는 FA (false alarm)와 핵심어를 검출하지 못하는 MD (missed detection)가 있다. 이 중 후처리는 검출된 핵심어의 실제 발화여부를 검증하여 FA를 처리한다. 이러한 후처리의 목적은 MD의 증가를 최소로 억제하면서 FA의 감소를 최대로 하는 것이다.

후처리 실험에 사용된 핵심어 검출기는 핵심어 모델과 필터 모델을 사용하는 일반적인 핵심어 검출 시스템을 이용하였다[1]. 실험에 사용된 핵심어 검출기는 552개의 핵심어를 처리하며, 필터 모델로는 문맥독립 음소 모델과 비음성 모델을 사용하였다. 먼저 실험을 준비하기 위해서 여러 가지 방법을 사용하여 핵심어 검출기를 최적화하였다. 다양한 실험에 의해서 최적화된 핵심어 검출기의 성능은 미검출율 6.76%, 1.06 FA/KW/HR이다.

핵심어 검출부로부터 발생한 FA를 처리하기 위한 후처리에는 신뢰도를 이용하여 핵심어 출현여부를 검증하는 방법, 음소 인식기를 이용하여 핵심어의 음소 순열과 비교하는 방법 등이 있다. 본 논문에서는 음소 모델과 반음소 모델을 이용하는 기존의 신뢰도 기반 핵심어 검증을 채택하고, 장단점을 분석하여 이를 보완하였다.

기존의 신뢰도 기반 후처리 방법인 RLJ-신뢰도는 음소 모델과 반음소 모델간의 유사도를 이용하여 신뢰도를 계산한다[2]. 기존의 RLJ-신뢰도의 단점은 두 가지로 분석할 수 있다. 먼저 각 음소 단위 신뢰도들의 통계적 안정성이 부족하여 단어 단위 신뢰도들이 열화되는 경향이 발생한다는 것이다. 두 번째는 반음소 모델이 핵심어가 발생되지 않은 영역에서 핵심어를 검출하는 FA를 처리하는데 있어서 변별력이 떨어진다는 것이다.

이러한 두 가지 단점 중 첫 번째는 참고논문 [3]에서 NCM이라는 새로운 신뢰도를 제안하여 해결하였다. 그러나 NCM도 두 번째 단점에 대한 해결책은 되지 못한다. 두 번째 단점을 해결하기 위해서 본 논문에서는 실제 음소 수열을 추정하여 이것들을 반음소 모델로 사용하는

새로운 신뢰도 ACM을 제안하였다.

또한 본 논문에서는 NCM과 ACM이 각기 다른 영역에서 장점을 갖는 것에 착안하여, 이 두 개의 신뢰도를 가중치 벡터를 사용하여 하나로 통합한 HCM을 제안하였다.

II. 핵심어 검출 시스템

실험에 사용된 핵심어 검출 시스템은 입력음성의 핵심어부와 비핵심어부를 독립적으로 구축하여 변별력을 높이는 핵심어 모델과 필터 모델의 결합 시스템을 채택하였으며, 토큰 매칭 알고리즘을 적용하였다[4].

실험에 사용된 핵심어는 전국 대학명 및 대학교내 부서명을 기반으로 하여 552개를 선정하였고, 비핵심어는 실제 전화번호 문의패턴을 반영하기 위해 대학교내 교환실에서 수집한 전화문의 내용을 분석하여 결정하였다.

2.1. 데이터베이스

훈련 데이터베이스는 핵심어 모델과 필터 모델을 위한 음성부 훈련 데이터베이스와 주변소음, 인간의 신체잡음 등을 처리하는 비음성부 훈련 데이터베이스로 구성된다. 훈련 데이터베이스는 트라이폰 분석을 통하여 선정한 1,301문장의 발성 데이터를 이용하여 구축하였다. 비음성부 훈련 데이터베이스는 비핵심어 문장의 끝에 입술소리와 숨소리를 첨부한 53문장의 발성 데이터를 사용하였다.

평가용 훈련 데이터베이스는 핵심어 검출기의 성능평가를 위해서 문장 당 최대 3개까지의 핵심어가 출현하도록 허용하였다. 객관적인 성능평가를 위해서 핵심어의 출현빈도를 조정하여 각 핵심어가 최소 1회 이상 출현하도록 하였다. 평가용 훈련 데이터베이스에서 핵심어 출현 횟수는 828회이다.

데이터베이스 녹음환경은 조용한 사무실 환경이고 8 kHz, 16 bit로 A/D 변환하였다.

2.2. 핵심어 검출기

핵심어 검출 시스템은 전처리부, 핵심어 검출부, 후처리부로 구성되어 있다. 아래의 그림 1은 본 논문에서 사용된 핵심어 검출기를 나타낸 것이다.

전처리부는 입력 음향으로부터 음성구간을 검출하고 이로부터 특징 파라미터를 추출한다. 특징 파라미터로는 12차의 멜 캡스트럼과 1차의 정규화 로그 에너지, 이들과 파라미터들의 델타를 사용하였다. 핵심어 검출부는 핵심어 모델과 필터 모델로 구성된 핵심어 검출 네트워크를 사용

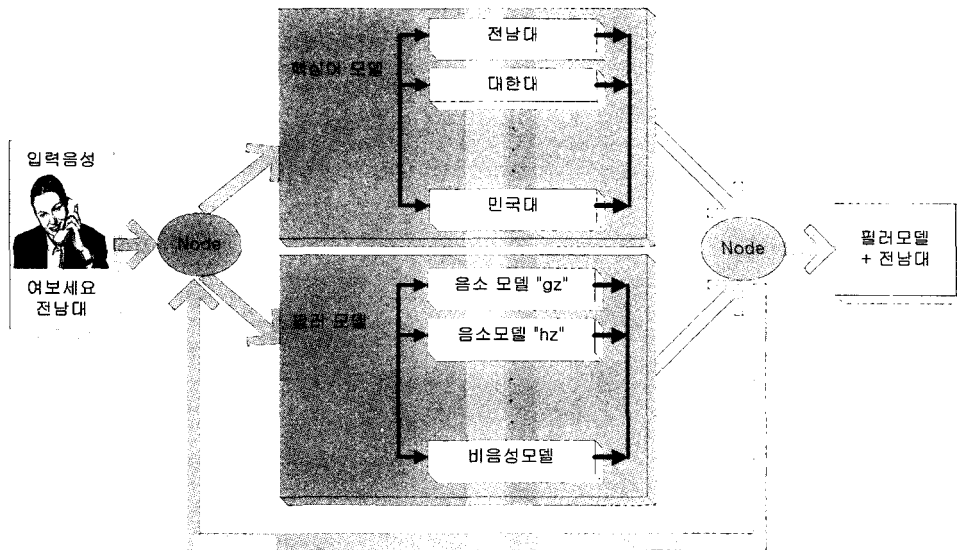


그림 1. 핵심어 검출 시스템
Fig. 1. Keyword spotting system.

하여 핵심어를 검출한다. 핵심어 검출 네트워크의 각 모델들은 GMM (Gaussian mixture model)을 이용하여 모델링되었다. 핵심어 모델과 필러 모델은 인식 네트워크에서 병렬로 연결되고 매 프레임마다 새로운 입력을 받아들인다. 실험에서는 552개의 핵심어 모델을 사용하였고, 44개의 모노폰 모델과 4개의 비음성 모델을 필러 모델로 사용하였다.

후처리는 FA를 처리하기 위해서 핵심어 검출부에 의해서 검출된 후보 핵심어의 신뢰도를 계산하고, 이를 이용하여 핵심어의 발화여부를 검증한다.

2.3. 핵심어 검출 실험결과

핵심어 검출기의 실험을 위하여 다음과 같은 평가기준을 선정하였다. 먼저 미검출율은 테스트 문장에서 출현한 핵심어를 핵심어 검출기가 제대로 검출하지 못한 경우이고, FAR로는 각 핵심어 당 FA의 출현횟수를 평가시간으로 정규화한 FA/KW/HR을 사용하였다[3].

다음의 식은 실험에 사용된 미검출율과 FA/KW/HR을 정의한 것이다. 실험에서는 552개의 핵심어가 총 828회 출현하였고, 실험에 사용된 평가용 녹음 데이터베이스의 길이는 총 1.6시간이다.

$$MOR = \frac{\text{missed detection 수}}{\text{핵심어 총 출현횟수}} \quad (1)$$

$$FA/KW/HR = \frac{\text{false alarm 수}}{\text{총 핵심어수} \times \text{평가DB의 녹음시간}} \quad (2)$$

핵심어 검출기는 표 1과 같은 조건을 이용하여 최적화

표 1. 최적화된 핵심어 검출부의 사양
Table 1. Specification of optimized keyword spotting model.

구 분	내 용
n-gram	tri-gram, bi-gram
핵심어 모델	3 state 3 mixture
필러 모델	3 state 9 mixture

하였다. 첫째, 음소 단위 n-gram을 사용하여 필러 모델의 토큰 분배에 대하여 문법 정보를 적용하였다. 둘째, 핵심어 모델과 필러 모델의 변별력을 향상시키기 위해서 각 모델의 분포 (mixture)를 조절하여 실험하였다. 셋째, 일반인들의 핵심어 발화패턴을 참고하여 핵심어 모델을 수정하였다. 최적화된 핵심어 검출부의 사양은 표 1과 같으며, 성능은 미검출율 6.76%, 1.06 FA/KW/HR이다.

III. 후처리 시스템

후처리는 핵심어 검출부에서 출력된 결과를 발화검증하여 시스템의 성능을 향상시키는 것을 말한다. 후처리 방법 중 일반적으로 사용되는 것이 신뢰도를 기반으로 한 것이다. 이것은 각 핵심어가 발생되었다는 가정에 대한 신뢰도를 계산하여 이 수치가 문턱치 이상일 경우에만 실제로 핵심어가 발화되었다고 인정한다.

후처리 적용 전의 핵심어 검출부에서 발생할 수 있는 에러를 분석한 결과 핵심어가 발생되지 않은 영역에서 핵심어를 오검출하는 오인증 (FA: false acceptance)과

유효한 핵심어를 거부하는 오거부 (FR: false rejection) 이 가장 많이 발생하였다. 그러므로 비중이 높은 두 가지 에러에 대한 적절한 후처리를 선택하여 적용하는 것이 특정영역에서의 핵심어 검출을 위한 후처리의 성능향상에 있어서 중요한 역할을 차지한다.

3.1. 기존의 신뢰도 기반 후처리 (RLJ-CM)

참고논문 [2]에서 제안되어 각 저자의 머릿글자의 조합으로 명명한 기존의 RLJ-신뢰도의 경우에는 각 음소 단위의 신뢰도를 음소 모델과 반음소 모델간의 유사도를 이용하여 계산하고 이를 단어 단위로 확장하여 검출된 핵심어의 신뢰도를 계산하였다[2,5].

$$\log pr_a = \frac{1}{M} \sum_{i=0}^{M-1} \log pr_{a_i} \quad (3)$$

$$cm_p = \frac{\log pr_p - \log pr_a}{|\log pr_p|} \quad (4)$$

$$CM = \frac{1}{f_{cm}} \log \left(\frac{\sum_{p=0}^{n_p-1} \exp(f_{cm} \cdot cm_p)}{n_p} \right) \quad (5)$$

위의 수식에서 $\log pr_a$ 는 반음소 모델의 평균 로그확률이고, $\log pr_p$ 는 핵심어 구성 음소모델의 로그확률이다. n_p 는 핵심어의 구성 음소의 수이고, M 은 반음소 모델의 수이다. cm_p 는 음소단위 신뢰도이고, f_{cm} 은 음의 가중치이다.

3.2. 정규화를 이용한 정규화 신뢰도

먼저 유효한 핵심어를 거부하는 FR의 경우, 참고논문 [3]에서는 그 원인을 각 음소 단위 신뢰도들의 통계적 특성의 불안정으로 분석하였다. 그리고 이것을 해결하기 위해서 각 음소 단위 신뢰도들의 평균과 표준편차를 이용하여 정규화한 NCM (normalized confidence measure)을 제안하였다.

참고논문 [3]에서 제안된 NCM은 기존의 RLJ-CM의 통계적 분포가 불안정한 현상을 해결하기 위해서 정규화라는 방법을 선택하였다. 먼저 각 음소 단위 신뢰도들이 정규분포를 가진다고 가정하고 평균과 표준편차를 조사하였다. 사전에 계산된 음소 단위 신뢰도의 평균과 표준편차를 이용하여 각 음소 단위 신뢰도들을 표준 정규분포로 정규화하였다.

$$ncm_p = \frac{cm_p - TriP_{q,mean}}{TriP_{q,sd}} + \alpha \quad (6)$$

$$NCM = \frac{1}{f_{ncm}} \log \left(\frac{\sum_{p=0}^{n_p-1} \exp(f_{ncm} \cdot ncm_p)}{n_p} \right) \quad (7)$$

f_{ncm} 은 음소 구성 신뢰도들의 가중치이다. cm_p 는 기존의 음소단위 신뢰도이고, ncm_p 는 정규화된 음소 단위 로그확률이다. 식 (6)에서의 $TriP_{q,mean}$, $TriP_{q,sd}$ 은 해당 트라이폰 (음소단위)의 신뢰도 평균과 표준편차를 말한다. α 는 음소 단위 신뢰도의 정규화에 사용되는 가중치로서 각 음소 단위 신뢰도가 양의 영역에 속할 수 있도록 실험에서는 2로 정하였다.

NCM을 적용하기 전의 기존 RLJ-CM의 경우에는 각 음소 단위 신뢰도들의 표준편차의 차이가 컸지만, 적용한 후에는 신뢰도들의 표준편차가 1의 값을 갖게 되어 보다 안정된 신뢰도 특성을 보장할 수 있게 되었다. 또한 음소 단위 신뢰도의 표준편차를 보상할 수 있는 가중치 α 를 사용하여 음소 단위 신뢰도를 항상 양의 값으로 안정시켰다.

이렇게 안정된 음소 단위 신뢰도는 단어 단위 신뢰도에 직접적인 영향을 주어서 유효한 핵심어가 불안정한 신뢰도에 의해서 거부되는 것을 막아준다.

3.3. 반떨러 모델을 사용한 Anti-filler CM

기존의 RLJ-CM에서 발생하는 주요 에러 두 번째는 핵심어가 발생되지 않은 영역에서 핵심어를 오검출하는 FA이다. 이를 제어하기 위해서 본 논문에서는 새로운 신뢰도를 제안한다.

FA가 발생하는 주된 원인은 반음소 모델이 역할을 제대로 수행하지 못하기 때문이다. 핵심어의 구성 음소 모델에 대한 반대가설로서 정의되는 반음소 모델은 일반적으로 음소 정의들 중에서 자신을 제외한 나머지를 사용한다. 이러한 반음소 모델의 확률은 각 반음소 모델들의 평균 확률로 정의되는데 여기서 문제가 발생한다. 만약 핵심어가 오검출되었다면 반음소 모델 중에서 실제 발생된 어떤 음소가 존재하게 된다. 이러한 음소는 실제로 높은 확률을 보이더라도 나머지 동료 반음소들의 확률값에 의해서 열화가 된다. 따라서 반음소 모델이 자신의 특성을 제대로 발휘하지 못하게 된다.

이를 제어하기 위해서 N-best 기반 반음소 모델 등을 사용해서 실제로 발생될 가능성이 높은 것을 반음소 모델로 사용하는 방법이 있다. 하지만 이 방법을 사용하였을 경우에는 유효한 핵심어에 대한 비교시에 반음소 모델에 실제 발생된 음소와 가장 잘 혼동되는, 즉 가장 분류하기

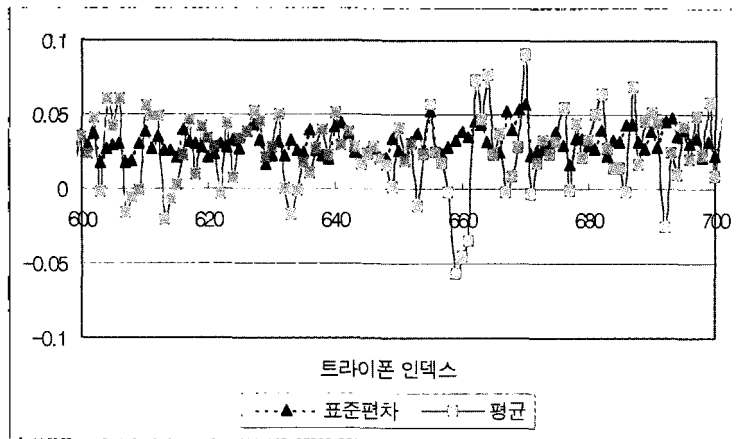


그림 2. 정규화 적용 전의 음소 단위 신뢰도의 평균과 표준편차 분포 (부분)
 Fig 2. The distribution of Mean and standard deviation of phoneme unit confidence measure, before using normalization (part).

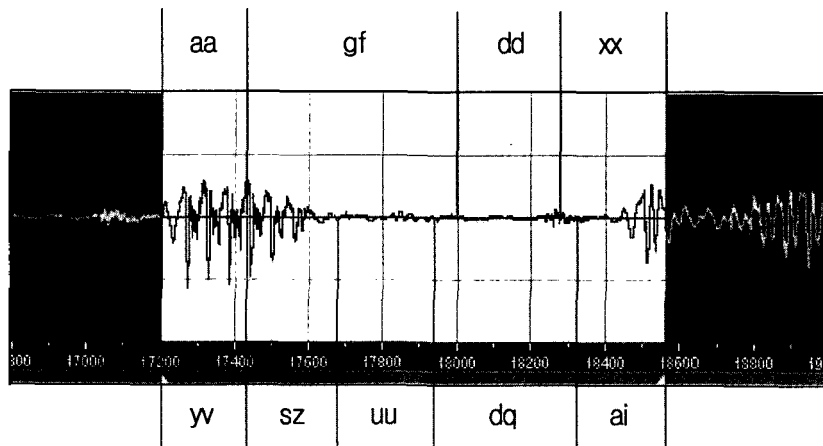


그림 3. NCM 추정 음소 구간과 ACM 추정 음소 구간 비교 (FA의 경우)
 Fig 3. The comparison phoneme section that is assumed by ACM with phoneme section that is assumed by NCM (in the case of FA).

까다로운 음소들이 배정되기 때문에 이에 대해서 고정 문턱치를 적용하는 것은 비현실적이 된다. 때문에 이러한 특성에 강인하여 변별력을 유지할 수 있는 가변 문턱치가 필요하게 되고, 이러한 요인들에 의해서 시스템의 복잡도와 계산비용이 증가하게 된다.

의와 같은 반음소 모델의 단점을 극복하기 위해서 본 논문에서는 새로운 반음소 모델의 개념을 제안하였다. 기존의 반음소 모델은 자신을 제외한 나머지 음소들로 정의되었다. 그러나 본 논문에서는 반음소 모델을 핵심어의 음소 모델과 상관없이 발생되었을 가능성이 가장 높은 음소 수열을 추출하는 방법을 사용하였고 이를 위해서 음소 인식을 사용하여 음소 수열을 추정하였다. 그리고 음소 인식기에 의해 구해진 음소들을 반필터 모델이라 정의하고 반음소 모델로 사용하였다. 그림 3은 핵심어가 발생되지 않은 영역에서 핵심어를 검출하는 FA의 경우에 기존

의 NCM이 추정하는 잘못된 음소 구간과 본 논문에서 제안한 ACM이 추정하는 음소 구간을 비교한 것이다. 실제로는 “부탁드립니다”에서 “ㄱ ㄱ ㄷ ㅡ”이 발생된 구간인데 핵심어 검출기에 의해서 “여수대”로 오검출되고 그 결과에 따라 NCM이 잘못된 음소 구간을 추정한 사례이다. 그림에서 보는 것처럼 FA의 경우에 검출된 핵심어의 음소 모델들에 한정된 탐색영역을 가지는 NCM에 비해서 모든 음소를 대상으로 한 탐색영역을 가지는 ACM에서 훨씬 정확한 음소 수열을 얻을 수 있다. 그림에서 보듯이 FA가 발생한 경우에 NCM에 의해서 추정된 음소 구간과 ACM에 의해서 추정된 음소 구간은 큰 차이를 보이게 된다. 즉, 잘못된 정보를 바탕으로 하여 NCM에 의해서 추정된 음소 모델 수열이 낮은 확률값을 갖는 반면에 음소 인식기의 인식결과를 바탕으로 하여 ACM에 의해서 추정된 음소 모델 수열은 높은 확률값을 가지게 되어 신뢰도가 이를 거부

할 수 있는 충분한 변별력을 확보할 수 있다.

그러나 FA 등의 경우에 서로의 음소 수열과 음소 구간이 상이하기 때문에, 새로운 개념의 반응소 모델과 기존의 음소 모델을 연계하기 위해서는 새로운 유사도 비교 방법이 필요하다. 이러한 상황을 해결하기 위해서는 검출된 음소 수열끼리 비교하여 치환 또는 삭제 에러 등의 감점요인에 의해서 신뢰도를 평가하는 방법이 있다. 하지만 이것은 현재 음소 인식기의 결과에 에러가 많으므로 실험에 적용하는데는 무리가 있었다. 때문에 본 논문에서는 음소 인식기를 사용하여 추정된 각 음소들의 경계를 새로운 음소 경계로 인정하고 기존의 NCM에서 새로운 음소 경계에서의 비터비 확률값을 구하고, 이를 음소 모델의 확률값으로 정하였다. 이렇게 하면 음소 인식기의 결과에 에러가 있더라도 음소 모델의 확률에 비해서 반응소 모델의 확률값이 상대적으로 낮은 값을 지니게 되므로 최종 판단에 영향을 덜 미치게 된다.

본 논문에서는 위와 같은 방법에 의해서 계산된 새로운 신뢰도를 ACM이라 정의하였다. 제안된 ACM을 계산하기 위한 수식은 다음과 같다.

$$acm_p = \frac{(\log pr_{interbi} - \log pr_{p-mono})}{|\log pr_{p-mono}|} \quad (8)$$

$$ACM = \frac{1}{f_{acm}} \log \left(\frac{\sum_p \exp(f_{acm} \cdot acm_p)}{nP_{acm}} \right) \quad (9)$$

위의 수식에서 $\log pr_{interbi}$ 는 음소 인식기에 의해서 검출된 음소 경계에서 기존 NCM이 가지는 비터비 검색 로그 확률이다. $\log pr_{p-mono}$ 은 음소 인식기에 의해서 계산된 음소의 로그확률로서 반응소 역할을 한다. f_{acm} 는 ACM에서 사용되는 음의 상수이다. 그림 4는 제안된 ACM의 계산과정을 나타낸 것이다.

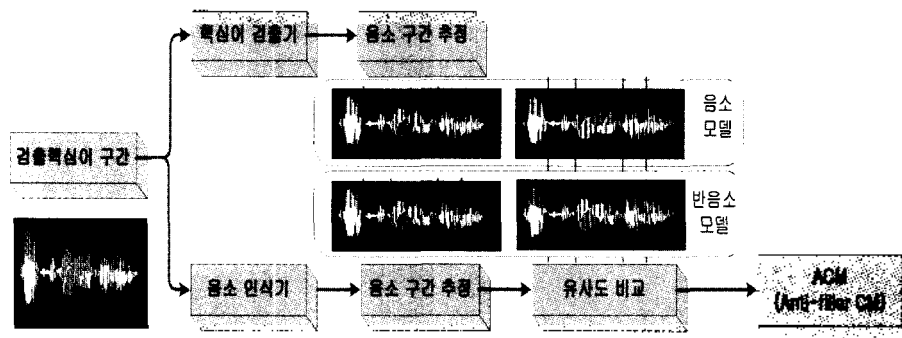


그림 4. 제안된 ACM 시스템 구조
Fig. 4. Block diagram of proposed ACM.

3.4. NCM과 ACM의 통합 (Hybrid CM)

3.2와 3.3에서 언급한 바와 같이 각 신뢰도들은 각기 다른 목적으로 설계되었고, 그로 인해서 특정분야의 에러에 대해서 좋은 성능을 나타낸다.

먼저 NCM의 경우는 각 음소 단위 신뢰도들의 안정성을 향상시켜서 핵심어가 잘못 거부되는 FR을 해결하였다. 그리고 ACM은 반응소 모델의 구조적 특성 때문에 FA에 대한 변별력이 저하되는 현상을 해결하기 위해서 음소 인식기를 이용한 직접적인 음소 수열 측정 방법을 제안하였다.

이러한 두 가지의 신뢰도 측정방법을 고찰해 보면, NCM은 기존의 RLJ-CM의 반응소 모델의 구조를 그대로 사용함으로써 FA에 대한 좋은 해법이 되지 못한다. 또한 ACM은 음소 인식기의 인식 에러와 동일한 음소간의 유사도 비교에 있어서 신뢰도 보상을 고려하지 않는 단점에 의해서 FA 외의 다른 에러들에 대해서 오히려 역효과를 나타내는 단점이 있다. 이렇듯 NCM과 ACM의 장단점은 상보적인 위치에 있음을 알 수 있다. 그러므로 기존의 NCM과 제안한 ACM을 하나로 통합하면 주요 에러들에 대한 신뢰도의 변별력 향상을 기대할 수 있을 것이다.

위와 같은 특성을 고려하여 본 논문에서는 가중치 벡터 α 를 이용하여 NCM과 ACM의 통합을 시도하였다. 두 가중치의 통합에서는 이미 NCM과 ACM에 의해서 시스템의 복잡도와 계산비용이 증가되었기 때문에 추가적인 증가요인을 최소화하는데 주안점을 두었다. 그래서 계산비용 대 성능비가 좋은 가중치 기반 통합이 실험에 사용되었다. HCM 계산에는 식 (10)을 사용하였으며, HCM 시스템의 구조는 그림 5와 같다.

$$HCM = \alpha \times NCM + (1 - \alpha) \times ACM \quad (10)$$

이 때 각 신뢰도에 대한 가중치는 각 신뢰도의 상보적

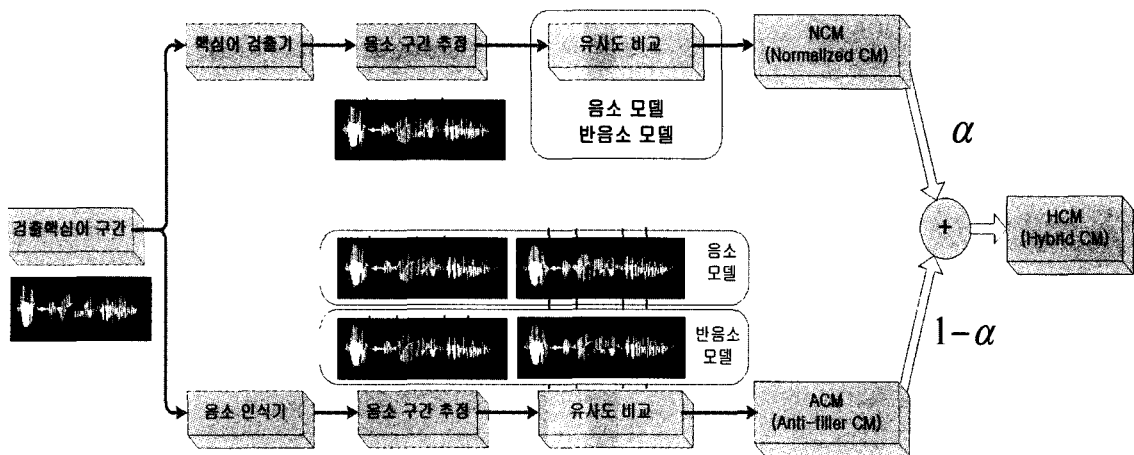


그림 5. 제안된 HCM 시스템 구조
Fig. 5. Block diagram of proposed HCM.

특성을 반영하기 위해서 1에 대한 상대적인 값으로 결정된다. 즉, NCM에 대한 가중치에 따라서 ACM의 가중치가 종속적으로 변화한다.

IV. 실험 결과

실험은 기존 RLJ-CM의 시스템 성능과 FR을 처리하기 위해서 도입된 NCM은 참고논문 [3]의 결과를 인용하였고, 이러한 결과들과 동일한 조건하에서 본 논문에서 제안한 ACM의 결과를 비교하였다.

4.1. 기존 신뢰도의 실험결과 (RLJ-CM & NCM)

참고문헌 [5]에서는 가변어휘 단어 인식 시스템에서 인

식된 단어의 신뢰도를 측정하였다. 이는 핵심어 모델과 비핵심어 모델들의 병렬 결합 네트워크를 이용하여 핵심어를 검출하는 핵심어 검출기의 신뢰도를 평가하는 본 논문의 실험결과와 절대적인 비교를 하는데는 무리가 있다. 참고문헌 [5]에서의 실험결과는 정인증 (CA: correct accept) 89%, 정거부 (CR: correct reject) 90%이다.

본 논문에서의 실험은 기존의 RLJ-CM과 이것의 FR 단점을 보완한 NCM의 성능을 비교하는 것이다. 그림 6에서 보는바와 같이 기존의 RLJ-CM보다 참고문헌 [3]에서 제안된 NCM이 좋은 결과를 나타내었다. 이는 유효한 핵심어가 잘못 거부되는 FA를 막기 위해 적용된 NCM이 역할을 충실히 수행했기 때문이다. MDR이 10% 정도인 경우에 기존 RLJ-CM의 경우 FA/KW/HR이 0.349, NCM 경우 FA/KW/HR이 0.280로 NCM을 적용한 경우가 기존의 RLJ-

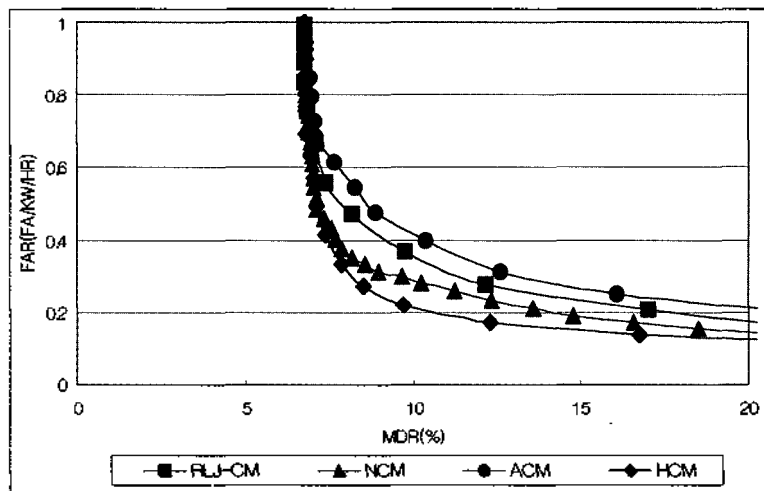


그림 6. HCM 실험결과
Fig. 6. Experiment result of HCM.

ACM에 적용한 경우에 비해서 20% 성능이 향상되었다.

4.2. 제안한 ACM의 실험결과

이번 실험은 기존의 RLJ-CM과 이것의 FA 단점을 보완한 ACM의 성능을 비교하는 것이다. MDR이 10% 부근인 경우 ACM은 0.397로 기존의 RLJ-CM에 비해서 성능이 12% 저하되었다.

이러한 결과가 나타난 원인으로는 2가지를 생각할 수 있다. 첫째는, 실제 음소 수열을 추정하기 위해서 사용한 음소 인식기의 인식 어려이다. 현재 음소 인식기의 인식 능력은 단어 인식기 등에 비해서 상대적으로 열악하기 때문에 이에 의해서 ACM의 성능이 저하되었다. 두 번째는 ACM을 구하는 공식에서 음소 모델과 반음소 모델이 동일한 음소를 갖게 되었을 경우에 대한 보상이 없다는 것이다. 이러한 경우 음소 모델과 반음소 모델이 서로 비슷한 수준의 확률값을 가지게 되어서 음소 단위 신뢰도가 저하되는 현상이 발생하게 된다. 그리고 이것은 단어 단위 신뢰도 저하와 직접적으로 연결된다. 이러한 단점에도 불구하고 ACM은 실제 음소 수열에 대한 정보를 가지고 반음소 모델을 구축한다는 점에서 새로운 신뢰도로서의 의미가 충분하다 할 수 있다.

4.3. 제안한 HCM의 실험결과

3.2와 3.3에서 언급한 바와 같이 NCM과 ACM은 각기 다른 영역에 최적화하여 설계되었다. 따라서 본 논문에서는 가중치 벡터 α 를 사용하여 이들을 통합하였다.

실험결과 가중치 벡터 α 가 0.05일 때 가장 좋은 결과를 보였다. 이 때 가중치 벡터는 NCM에 0.05를 ACM에는 0.95를 할당하였지만 표 2에서 보는바와 같이 NCM과 ACM의 단어 단위 신뢰도들의 평균의 차이를 고려하면

표 2. 각 신뢰도의 평균과 표준편차
Table 2. Mean and standard deviation of each confidence measure.

구분	RLJ-CM	NCM	ACM
평균	0.016	1.986	0.015
표준편차	0.024	0.947	0.029

표 3. 신뢰도 실험결과 (FA/KW/HR)
Table 3. The result of the experiment of confidence measure (FA/KW/HR).

구분	RLJ-CM	NCM	ACM	HCM
MDR 7.5%	0.560	0.431	0.614	0.417
MDR 10%	0.349	0.280	0.397	0.219

NCM이 상대적으로 큰 비중을 차지한다.

표 3에서 나타낸 바와 같이, MDR 10% 정도에서 제안한 HCM은 0.219 FA/KW/HR로서 기존의 RLJ-CM에 비해서는 성능이 37% 향상되었고, NCM에 비해서는 22%의 성능 향상을 보였다. 또한 이를 참고하면, 제안한 ACM은 비록 성능 면에서는 기존의 RLJ-CM에 비해서 나쁜 결과를 보였지만, 실제로 발생된 음소열에 대한 정보를 가지고 있어서 이것이 가중치 벡터를 이용한 신뢰도의 통합과정에서 장점으로 작용해서 기존 NCM의 약점을 보완했다고 볼 수 있다.

V. 결론

본 논문에서는 핵심어 검출기의 후처리부로 기존에 사용되고 있던 RLJ-CM과 NCM의 장단점을 분석하고 이를 보완하였다. 기존의 RLJ-CM과 NCM의 취약점이었던 FA에 대한 해결책으로 ACM을 제안하였고 기존의 NCM과 제안한 ACM을 통합한 HCM을 제안하였다.

제안한 ACM은 기존의 신뢰도들의 반음소 모델 구축 방법을 음소 인식기에서 인식한 음소 수열로 추정하는 방법으로 대체하여 FA 발생시에 보다 견인한 성능을 보인다.

기존의 NCM과 제안한 ACM의 특성을 살펴보면, 신뢰도의 통계적 특성을 이용하여 정규화한 NCM과 음소 인식기를 이용하여 반음소 수열을 직접 추정한 ACM은 서로 상보적인 특성을 가진다. 본 논문에서는 이러한 상보적인 특성을 가지는 두 신뢰도를 가중치 벡터 α 를 사용하여 HCM이라는 새로운 신뢰도로 통합하였다.

실험결과 NCM 10% 정도에서 HCM은 0.219 FA/KW/HR로, 기존의 NCM 단독 사용시에 비해서 22%의 성능향상을 보였고, 제안한 ACM 단독 사용시에 비해서 44%의 성능향상을 보였다.

참고 문헌

1. P. Jeanrenaud, K. Ng, J. R. Siu, M. Rohlicek and H. Gish, "Phonetic-based word spotter : Various configurations and application to event spotting," *Proc. of Eurospeech 93*, 2, 1057-1060, 1993.
2. M. G. Rahim, C.-H. Lee, B.-H. Juang and W. Chou, "Discriminative utterance verification using minimum string verification error (MSVE) training," *Proc. of ICASSP 96*, 3585-3588, 1996.
3. 김 철, 이경록, 김진영, 최승호, 최승호, "정규화 신뢰도를 이용

- 한 핵심어 검출 성능 향상,” 한국음향학회지, 21 (4), 380-386, 2002.
4. S.-J. Young, N. H. Russell and J.-H. S. Thornton, “Token Passing : a simple conceptual model for connected speech recognition systems,” Technical report of Cambridge university engineering department, TR 38, 1989.
5. 김기태, 문광식, 김회린, 이영직, 정재호, “가변어휘 단어 인식에서의 미등록어 거절 알고리즘 성능 비교,” 한국음향학회지, 20 (2), 27-34, 2001.

저자 약력

● 이 경 록 (Kyong-Rok Lee)



1997년 2월: 호남대학교 전자공학과 (공학사)
 2001년 8월: 전남대학교 전자공학과 (공학석사)
 2001년 8월~현재: 전남대학교 정보통신공학부 박사과정
 * 주관심분야: 음성인식, 멀티미디어 인콰싱

● 서 현 철 (Hyun-Chul Seo)



2000년 8월: 전남대학교 전자공학과 (공학사)
 2000년 3월~현재: 전남대학교 전자공학과 석사과정
 * 주관심분야: 음성인식, 화자적응

● 최 승 호 (Seung-Ho Choi)

1981년 2월: 전북대학교 물리학과 (이학사)
 1984년 8월: 명지대학교 전자공학과 (공학석사)
 1992년 2월: 명지대학교 전자공학과 (공학박사)
 1992년 3월~현재: 동신대학교 정보통신공학과 부교수
 * 주관심분야: 음성인식, 멀티미디어통신, 멀티모달 MMI

● 최 승 호 (Seung-Ho Choi)



1991년: 한양대학교 공과대학 전자공학과 (공학사)
 1993년: 한국과학기술원 전기 및 전자공학과 (공학석사)
 1999년: 한국과학기술원 전기 및 전자공학과 (공학박사)
 1996년~2002년 9월: 삼성종합기술원
 2002년 9월~현재: 서울산업대학교 전자공학과 교수
 * 주관심분야: 음성인식, 음성코딩, 음성합성, 디지털통신호처리, 디지털통신

● 김 진 영 (Jin-Young Kim)

1986년 2월: 서울대학교 전자공학과 (공학사)
 1988년 2월: 서울대학교 전자공학과 (공학석사)
 1994년 2월: 서울대학교 전자공학과 (공학박사)
 1994년~1995년: 한국통신 소프트웨어 연구소
 1995년~현재: 전남대학교 정보통신공학부 부교수
 * 주관심분야: 음성인식, 음성합성, 멀티모달 MMI