

# BMS 알고리즘을 이용한 핵심어 검출기 거절기능 성능 향상 실험

김병돈(전남대), 김진영(전남대), 최승호(전남대)

## <차 례>

- |                     |                           |
|---------------------|---------------------------|
| 1. 서론               | 3.2. 기존의 신뢰도 기반 반응소 모델 선정 |
| 2. 인식 시스템           | 3.3. Background Model Set |
| 2.1. 핵심어 검출기        | 3.4. BMS를 이용한 반응소 모델 선정   |
| 2.2. 데이터베이스 구성      | 4. 실험 결과                  |
| 2.3. 검출 실험 결과 평가 기준 | 5. 결론                     |
| 3. 후처리 시스템          |                           |
| 3.1. 기존의 신뢰도 기반 후처리 |                           |

## <Abstract>

### **Improvement of Confidence Measure Performance in Keyword Spotting using Background Model Set Algorithm**

**Byoung-Don Kim, Jin-Young Kim, Seung-Ho Choi**

In this paper, we proposed Background Model Set algorithm used in the speaker verification to improve calculating confidence measure(CM) in speech recognition. CM is to display relative likelihood between recognized models and antiphone models. In previous method calculating of CM, we calculated probability and standard deviation using all phonemes in composition of antiphone models. At this process, antiphone CM brought bad recognition result. Also, recognition time increases. In order to solve this problem, we studied about method to reconstitute average and standard deviation using BMS algorithm in CM calculation.

\* Keywords: pitch recognition, BMS (Background Model Set)

## 1. 서 론

음성은 인간만이 사용하는 가장 오래되고 가장 효율적인 정보 교환 수단이다. 컴퓨터의 발달로 인하여 사회가 급속히 산업화되고 정보화되어 인간의 생활이 풍족해지면서 인간은 좀더 편리한 것을 찾게 되고 사람과 기계와의 정보 교환에서 보다 효율적인 방법을 찾게 되었다. 이에 따라 인간은 가장 효율적인 정보 교환 수단인 음성을 이용하여 기계와의 정보 교환을 피하게 되었다. 이에 따라 음성인식에 대한 연구가 활발하게 진행되었다.

음성인식은 사람의 발성을 실시간으로 처리하여 언어의 의미를 파악하여 시스템이 그에 합당한 조치를 취하는데 판단 자료를 제공한다. 현재 음성인식의 큰 흐름은 대어휘 연속 음성인식기를 이용한 연속 음성인식과 해당 분야에 적합한 핵심어를 탑재한 핵심어 검출기를 이용한 핵심어 검출이다. 그 중에서 연속으로 발음된 음성 중에서 핵심어 출현을 인식하는 핵심어 검출기는 연속 음성인식기에 비해서 검출부의 구조가 간단하여 연산 비용이 저렴하다. 이 때문에 연속 음성인식기에 비해서 우월한 특성을 갖는다.

이러한 음성인식 과정에서 발생하는 에러의 패턴에는 인식어를 잘못 검출하는 FA (false alarm)와 인식어 출현을 검출하지 못하는 MD (missed detection)가 있다. 이 중에서 후처리 과정에서 검출된 인식어의 실제 발화 여부를 검증하여 FA를 처리한다. 이와 같은 후처리의 최종 목표는 MD의 증가를 최소한으로 줄이면서 FA 감소를 최대로 하는 것이다.

후처리 성능 향상에 사용된 핵심어 검출기는 입력 음성의 핵심어부와 비핵심어부를 독립적으로 구축하여 변별력을 높일 수 있는 핵심어 모델과 필터 모델을 사용하는 일반적인 핵심어 검출 시스템을 이용하였다[1].

신뢰도 측정은 오검출을 처리하는 후처리 방법 중 음소 모델과 반응소 모델의 유사도를 이용하는 신뢰도를 측정한다. 이 과정에서 사용되는 반응소 모델은 자기 음소를 제외한 유사 음소 집합을 말한다. 이러한 유사 음소의 집합은 정의된 유사 음소가 많을수록 반응소 모델로서의 역할을 충실히 수행한다. 그러나 유사 음소수의 증가는 그에 상응하는 훈련 데이터베이스의 증가를 요구하기 때문에 모든 음소의 계산이 필수적이다. 이에 따라서 음소수의 증가는 검출기의 성능을 떨어지게 할 뿐만 아니라 검출 시스템의 계산 속도를 증가시켜 전체적인 시스템 메모리의 과다한 소요와 인식 시간의 증대를 가져오게 되었다.

위의 단점을 해결하기 위해서 본 논문에서는 인식 향상과 계산 시간 단축을 위해서 화자 인증에 사용되는 BMS (Background Model Set) 알고리즘을 이용한 새로운 신뢰도를 제안한다. BMS 알고리즘을 이용하여 반응소 모델의 수의 감소를 가져왔고, 또한 기존의 신뢰도 검출과 BMS 적용 후의 신뢰도를 비교하였다.

## 2. 인식 시스템

핵심어 인식이란 연속적인 화자의 여러 형태의 발성, 즉 연속어 혹은 대화체에서 미리 정해 놓은 단어 혹은 어절들 중에서 사전에 정의된 핵심어를 검출하고 이를 분석하여 화자의 언의를 파악하는 것이다. 연속된 화자의 발성 중 상대적으로 정보가 적은 비핵심어 구간 및 비음성 구간은 필터 모델을 사용하여 처리하고, 상대적으로 정보를 많이 가진 핵심어 구간에 대하여 인식을 실시한다.

핵심어 검출은 먼저 업무에 맞게 선정된 핵심어 사전을 필요로 한다. 핵심어 사전은 적용 분야에서 업무를 처리하기 위한 핵심적인 정보를 가지는 단어들로 정해진다.

핵심어 모델 네트워크는 위에서 정의된 핵심어 사전에 대하여 문장 종속형 음소 모델을 이용하여 구축하는 것이 일반적이다. 이러한 핵심어 모델 네트워크는 각 핵심어 모델을 이용하여 구축하는 것이 일반적이다. 이러한 핵심어 모델 네트워크는 각 핵심어 모델을 토큰 패싱 알고리즘을 이용한 리니어 결합 네트워크로 이루어진다. 핵심어 모델 네트워크에서의 핵심어 인식은 기존의 잘 훈련된 핵심어 사전상의 단어 모델의 패턴과 입력 음성의 유사도를 이용한다[2].

필터 모델 네트워크는 일반적으로 비핵심어 구간을 처리하는 모델과 비음성 구간을 처리하는 모델로 구성된다. 비핵심어 모델은 일반적으로 문장 독립형 음소 모델을 사용하며, 연속된 화자의 발성 중 상대적으로 정보가 적은 비핵심어 구간을 처리한다. 비음성 모델은 발성 간 배경 소음, 입술소리, 숨소리 등을 처리한다. 본 연구에서는 비음성 모델로서 입술소리와 숨소리를 사용하였다. 필터 모델 네트워크는 이러한 비핵심어 모델과 비음성 모델을 토큰 패싱 알고리즘을 이용한 네트워크들의 리니어 결합 네트워크로 이루어진다.

### 2.1. 핵심어 검출기 구조

핵심어 검출기는 입력 음성에 대하여 인식 네트워크를 이용한 핵심어 검출을 실행한다. 핵심어 검출기는 전체적인 구조를 살펴보면 크게 3가지로 구성되어 있다. 핵심어 검출기는 전처리부, 인식부, 후처리부로 구분된다.

전처리부는 입력 음성을 A/D 변환한 후 인식에 사용되는 특징 파라미터로 변환하는 역할을 한다. 구체적으로 살펴보면 전처리부에서는 칼리브레이션, 시작점/끝점 검출, 특징 파라미터 추출 과정을 거친다.

칼리브레이션은 보다 견인한 음성인식을 위해 초기 환경 잡음을 계산하여 차후의 음성 구간 인식을 위한 기초 자료를 생성한다. 입력 신호의 음성 여부에 관계없이 초기 에너지와 ZCR (zero crossing rate)의 평균과 분산을 계산하여 다음 단계인 시작점/끝점 검출에서의 음성 여부 판단을 위한 비음성 구간의 기준으로 사

용한다. 즉, 환경 잡음을 고려하기 위해서 초기 5프레임을 비음성 구간으로 가정하고 그 구간에서의 에너지와 ZCR의 평균과 분산을 구하여 비음성구간의 평균 에너지와 ZCR로 사용한다.

시작점/끝점 검출은 입력 음향에서 음성의 시작과 끝을 감시하여 인식기에 입력으로 부여하는 역할을 한다. 입력 음성의 시작점과 끝점의 검출은 실제 인식기의 성능에 큰 영향을 미치므로 매우 중요하다고 할 수 있다. 시작점을 결정하기 위해서는 rVSE(음성 시작 최소 에너지), rUCZCR(음성 시작 최소 ZCR)을 사용한다. 끝점을 결정하기 위해서는 rVEE(음성 종료 최소 에너지), rFZCR(음성 종료 최소 ZCR)을 사용한다.

특징 파라미터는 12차의 MFCC와 1차의 에너지, 그리고 이들 파라미터들의 델타를 사용하였다.

핵심어 검출부는 전처리 부에서 인계된 특징 파라미터를 바탕으로 HMM 기반의 토큰 패싱 알고리즘을 이용한 인식 네트워크에서 예상 핵심어를 검출한다. 핵심어 검출 네트워크는 핵심어 모델 네트워크와 필터 모델 네트워크로 구성되며, 토큰 패싱 알고리즘이 적용된다. 토큰 패싱 알고리즘을 적용한 핵심어 검출 네트워크에서는 입력된 음성을 분석하여 입력 음성을 필터 모델과 핵심어 모델간의 결합으로 표현한다. 핵심어 검출 네트워크는 토큰 패싱 알고리즘에 의해서 입력단의 끝에서 각 모델들의 exit 노드에 도달한 토큰들 중 가장 큰 확률을 가진 토큰을 선택하고 이 토큰의 경로 정보를 분석하여 전체 검출 결과를 도출한다.

후처리부는 인식부에서 인계된 예상 핵심어를 신뢰도의 측정을 통해 검증해서 예상 핵심어의 수용 여부를 결정한다. 신뢰도는 핵심어로 결정된 구간에 대해서 핵심어를 구성하는 모노폰과 각각의 반음소 모델간의 유사도를 이용하여 계산하였다.

## 2.2. 데이터베이스 구성

훈련 데이터베이스는 인식 모델과 필터 모델을 위한 음성부 훈련 데이터베이스와 비음성부 훈련 데이터베이스로 구성된다. 필터 모델과 인식 모델은 HTK (Hidden Markov Model Tool Kit)를 이용하여 훈련을 실시하였다. 훈련용 데이터베이스는 트라이폰을 분석한 결과 선정된 1,031개의 녹음 문장을 20대 남성 화자 33명을 대상으로 녹음하였다. 평가용 데이터베이스는 핵심어 검출기의 성능평가를 위해 20대 남성 화자 6명의 496문장에서 최대 3개까지의 핵심어를 출현하도록 허용하여 이를 평가용 데이터베이스로 사용하였다. 데이터베이스 녹음 환경은 조용한 사무실 환경이고 8kHz, 16bit로 A/D 변환하여 PCM 파일로 저장하였다

### 2.3. 검출 실험 결과 평가 기준

인식 성능 평가를 위하여 다음과 같은 평가 기준을 선정하였다.

첫째, MDR (Missed Detection Rate)은 핵심어 발생 지점에서 출현한 핵심어를 검출기가 제대로 검출하지 못한 경우이고, 둘째 FAR로는 각 핵심어 당 FA의 출현 횟수를 평가 시간으로 정규화한 FA/KW/HR을 사용하였다[3, 9].

다음의 식은 실험에 사용된 미검출률과 FA/KW/HR을 식으로 정의한 것이다.

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

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

## 3. 후처리 시스템

### 3.1. 기존의 신뢰도 기반 후처리(RLJ-CM, NCM)

본 연구에서는 참고논문 [4]에서 제안된 RLJ-CM과 참고논문 [5]에서 제안된 NCM (Normalized CM)을 이용하여 신뢰도를 측정하였다. RLJ-CM의 경우 각 음소 단위의 신뢰도를 음소 모델과 반음소 모델의 유사도를 이용하여 계산하고 이를 기준으로 단어 단위로 확장하여 검출어의 신뢰도를 계산하였다[4].

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

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

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

위의 수식에서  $\log pr_a$ 는 반음소 모델의 평균 로그 확률이고,  $\log pr_p$ 는 핵심어 구성 음소 모델의 로그 확률이다.  $n_p$ 는 핵심어의 구성 음소의 수이고,  $M$

은 반응소 모델의 수이다.  $cm_p$ 는 음소 단위의 신뢰도이고  $f_{cm}$ 은 음의 가중치이다.

정규화 신뢰도(Normalized CM)은 RLJ-CM의 통계적 분포가 불안정한 현상을 해결하기 위해서 정규화라는 방법을 선택하였다. 먼저 각 음소 단위 신뢰도들이 정규분포를 가진다고 가정하고 평균과 표준편차를 조사하여 사전에 계산된 음소 단위 신뢰도의 평균과 표준편차를 이용하여 각 음소 단위 신뢰도들을 표준 정규분포로 정규화하였다[5, 6].

$$(6) \quad ncm_p = \frac{cm_p - TriP_q.mean}{TriP_q.sd} + a$$

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

$f_{ncm}$ 은 음소 구성 신뢰들의 가중치이다.  $cm_p$ 는 기존의 음소 단위 신뢰도이고,  $ncm_p$ 는 정규화된 음소 단위 로그 확률이다.  $a$ 는 음소 단위 신뢰도의 정규화에 사용되는 가중치이다.

### 3.2. 기존의 신뢰도 기반 반응소 모델 선정

신뢰도란 인식된 모델과 인식되지 않은 모델간의 상대적인 유사도이므로 각 모델에 대한 가장 혼동되기 쉬운 유사한 것을 찾아 그에 따른 모델을 구성하는 것을 반응소 모델이라고 한다.

핵심어 검출기에서 반응소 모델을 구성하기 위해서 사용된 방법은 Bhattacharyya 거리 측정을 통하여 음소간의 거리를 계산하였다. 이 방법은 GMM 기반 음소 모델에 대해 평균과 공분산 매트릭스만을 가지고 음소들 간의 거리를 계산하는 방법으로 별도의 사전 계산과 준비 과정을 필요로 하지 않는다. 이 측정 방법은 계산이 매우 단순하고, 정확히 거리를 계산하는 것보다 에러에 대한 경계를 제공하여 유연성을 갖는다.

식(8)은 Bhattacharyya 거리를 구하는 공식이며  $M$ 은 음소의 평균,  $\Sigma$ 는 음소의 공분산 매트릭스를 나타낸다[7].

$$(8) \quad D_{bhat} = \frac{1}{8} (M_1 - M_2)^T \left[ \frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (M_2 - M_1) + \frac{1}{2} \ln \frac{|\Sigma_1 - \Sigma_2|}{\sqrt{|\Sigma_1| |\Sigma_2|}}$$

### 3.3. Background Model Set

BMS에서의 평균 로그 유사도는 Background Model에 의해 계산되어진다. 식(9)는 평균 로그 유사도 계산식을 나타낸 것이다.

$$(9) \quad \iota(X|\lambda_c) = \log \left[ \frac{1}{N_B} \sum_{b=1}^{N_B} \exp \iota(X|\lambda_b) \right]$$

위 식에서  $N_B$ 는 mixture 개수이다.

Background Model Set의 각각의 클라이언트들은 그 클라이언트 모델 세트로부터 선택된다. 트레이닝 데이터를 사용함으로써 클라이언트 모델들 간의 거리를 계산할 수 있게 된다. 모델  $\lambda_D$ 와  $\lambda_E$ 는 모델 구성 시에 사용되었던 특징 벡터 세트인  $X_D$ 와  $X_E$ 에 대응된다. Set의 거리를 구하는 식은 식(10)과 같다.

$$(10) \quad d(\lambda_D, \lambda_E) = [\iota(X_D|\lambda_D) - \iota(X_D|\lambda_E)] + [\iota(X_E|\lambda_E) - \iota(X_E|\lambda_D)]$$

식(10)에서 거리 계산식은 모델  $\lambda_D$ 와  $\lambda_E$  둘간의 유사함 또는 둘간의 접근 상태를 계산할 수 있다. BMS는 각 모델들이 클라이언트 모델과 얼마나 근접해 있는지, 혹은 얼마나 떨어져 있는가의 정보를 가지고 있다. 이 정보를 통해서 우리는 가장 클라이언트로부터 근접해 있는 모델들을 사용하고 클라이언트로부터 가장 멀리 떨어져 있는, 즉 취약함을 보이는 모델을 제거할 수 있다.

클라이언트 모델이  $\lambda_c$ 로 주어지고, 모델 세트  $B$ 에서 가장 근접해 있는 모델을  $\lambda_b$ 라고 하면 BMS 수식은 아래와 같다.

$$(11) \quad \lambda_i = \arg \max_{\lambda_i} \left[ \frac{1}{N_{B_{close}}} \sum_{\lambda_b \in B_{close}} \frac{d(\lambda_b, \lambda_j)}{d(\lambda_c, \lambda_j)} \right]$$

$\lambda_i$ 는 모델 세트  $B$ 에서  $\lambda_c$ 에는 가장 가깝고  $\lambda_b$ 와는 가장 멀리 떨어진 client model argument이고  $d(\lambda_c, \lambda_j)$ 는  $\lambda_c$ 와  $\lambda_j$ 의 거리를 나타낸다.

BMS argument를 계산하기 위해서는 3단계의 과정이 필요하다.

첫째, 모델 세트  $B$ 에서 가장 가까운 곳에 있는 음소 모델의 거리를 측정한다.

둘째, 식(11)을 이용하여 argument  $\lambda_i$ 를 계산한다.

셋째, 위의 두 과정을  $N_{B_{close}} = N_B/2$ 이 나올 때까지 반복한다[8].

이와 같은 방법을 통해서 반응소 세트를 결정한다. 이 과정에서 BMS 알고리즘에서 필요한 거리 함수를 계산하기 위하여 Bhattacharyya 거리를 사용한다.

<그림 1>은 기존의 거리 세트 개념과 BMS를 이용한 거리와의 관계를 나타낸 것이다.

### 3.4. BMS를 이용한 반응소 모델 선정

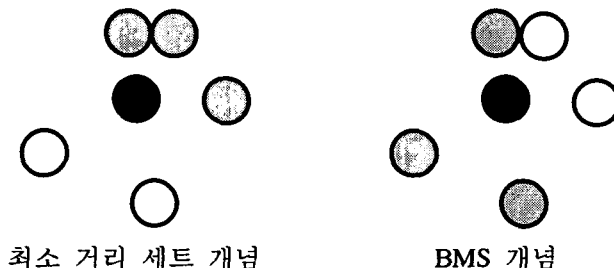
BMS 알고리즘을 이용한 반응소 모델 선정 과정에서 BMS 알고리즘 적용을 두 가지 방법을 통하여 실험을 실시하였다.

첫 번째 방법으로 임계값을 사용하여 반응소 모델을 선정하였다. 이 경우 각각의 음소마다 반응소 세트의 개수가 달라짐을 알 수 있다. 즉 BMS 모델 안의 거리의 합을 구하고 이 값이 임계값에 이를 때까지 계속해서 반응소 모델을 추가하고 임계값을 넘어가면 반응소 모델 생성을 멈추고 인식 단계를 계속 수행한다.

두 번째 방법은 반응소 개수에 따른 반응소 모델을 선정하였다. 반응소의 개수에 따라서 반응소 세트를 결정할 수 있다. 즉 음소마다 반응소의 개수를 임의의  $n$ 개로 고정하고 고정된 반응소들을 이용하여 반응소 세트를 결정하였다.

## 4. 실험 결과

실험 결과는 6명의 결과의 평균한 값을 데이터로 나타낸 것이다. <그림 2>는 핵심어 검출기에서 기존의 CM과 BMS-CM과의 비교를 나타낸 것이다. MDR 22% 부근에서 BMS 알고리즘을 적용했을 때 RLJ-CM은 0.322FA/KW/HR (false alarm/keyword/hour)로서 BMS 적용 전의 반응소 세트에 비해서 성능이 29% 향상됨을 알 수 있다.



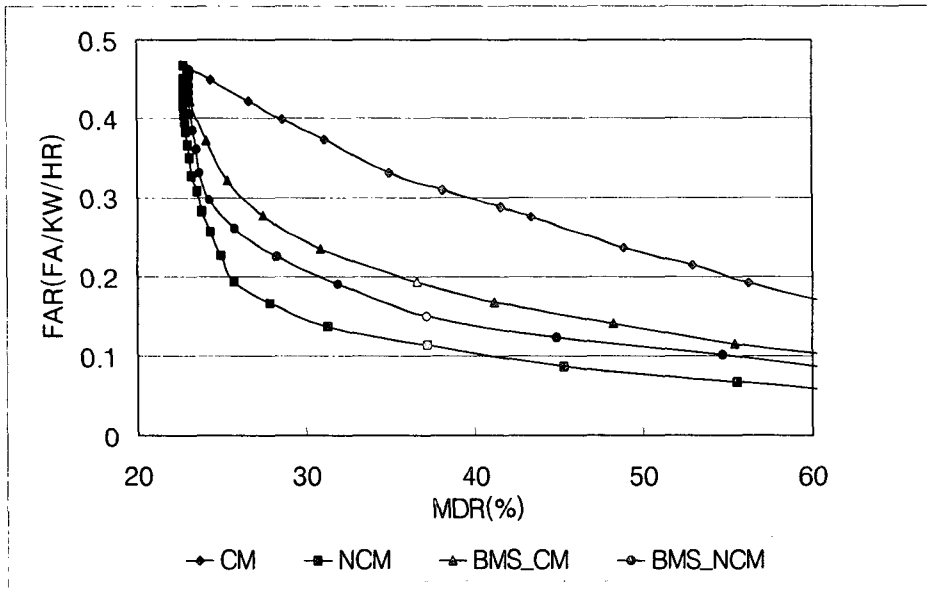
<그림 1> 최소거리 세트와 BMS 비교



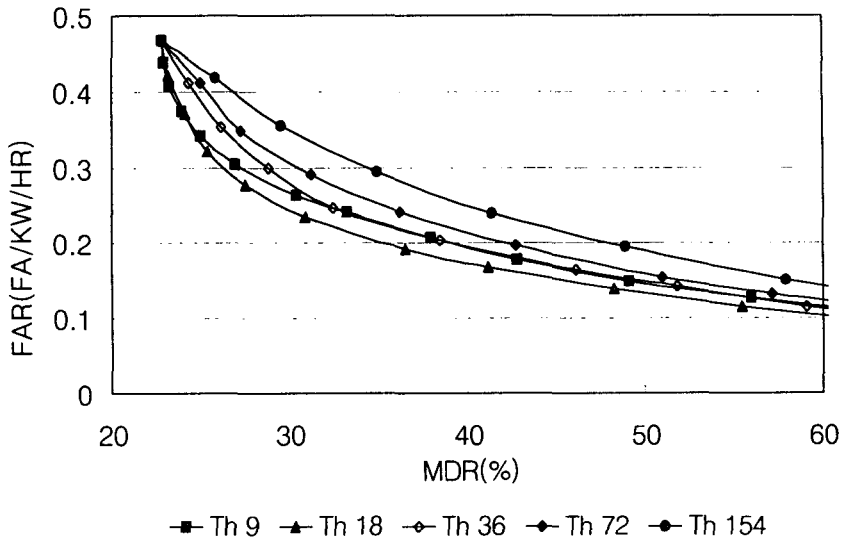
하지만 NCM에서는 MDR 25%에서 0.385 FA/KW/HR (false alarm/keyword /hour)로서 BMS 적용 시에 기존의 NCM보다 뒤떨어지는 결과를 보인다.

<그림 3>과 <그림 4>에서 보는 바와 같이 값이 커질수록 인식 성능이 향상되지만 특정 범위를 넘게 되면 인식 성능이 저하됨을 알 수 있었다. 실험 결과 임계값 18을 주었을 때와 반응소 개수가 7개일 때 인식 성능이 가장 좋음을 알 수 있다. <표 1>은 임계값에 따라 사용된 음소의 평균 개수를 나타낸 것이다.

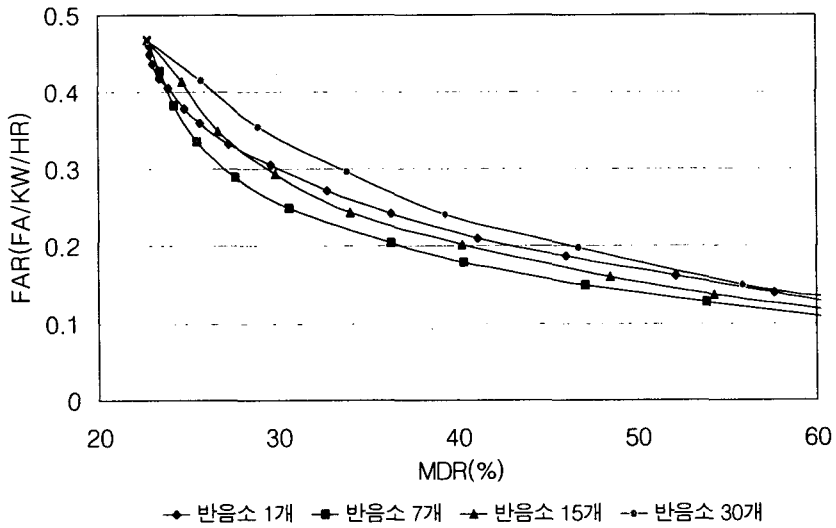
<표 1>을 통해서 임계값 18을 사용했을 때 모든 음소가 사용하는 반응소의 평균 개수가 7.04임을 알 수 있다. 모든 반응소 세트를 사용했을 때보다 전체 반응소 세트의 14%만을 사용했을 때 가장 좋은 성능을 보임을 알 수 있었다. 이 결과를 통해서 우리는 인식 음소를 제외한 모든 음소를 반응소 세트로 사용했을 때의 인식 성능에 비해 BMS 알고리즘을 이용하여 반응소 세트의 생성에 제한을 두었을 때의 신뢰도 성능이 향상됨을 알 수 있었다.



<그림 2> 핵심어 검출기에서 실험 결과



<그림 3> 임계값 변화에 따른 실험 결과



<그림 4> 반응소 개수에 따른 실험 결과

&lt;표 1&gt; 임계값에 따라 사용된 반음소 평균 개수 비교

임계값	3	5	9	18	36	72	154
반음소 평균 개수	1.43	2.13	3.88	7.04	12.12	20.37	35.24

위의 실험을 통해서 기존의 신뢰도와 BMS 알고리즘을 이용한 신뢰도를 비교하였다. 위의 실험 과정에서 각각의 인식 과정에서 걸리는 시간을 측정하였다.

<표 2>는 BMS 알고리즘 사용 전의 인식 시간과 BMS 알고리즘 사용 후의 인식 시간을 비교한 것이다. 각 실험에 측정된 시간은 6명의 인식 시간을 평균하여 1인의 총 496문장을 인식했을 때 걸린 평균 시간이며, BMS 알고리즘 사용 후의 인식 시간 측정은 <그림 3>과 <그림 4>에서 보는 바와 같이 실험 과정에서 가장 좋은 성능을 나타내었던 임계값 18을 사용했을 때와, 반음소의 개수를 7개 사용했을 때의 측정된 시간이다. 임계값 18을 사용했을 때 사용된 모든 반음소의 개수는 359개이고, 반음소 개수 7개를 사용했을 때 사용된 모든 반음소의 개수는 357개이다. 시스템 구동 환경은 Pentium IV 1.8GHz, Ram 512MByte를 사용하여 같은 환경에서 실험을 실시하였다.

<표 2>에서 보는 바와 같이 BMS 알고리즘을 적용하기 전에 화자 1명에 인식된 시간은 32분이 걸린 것에 비해 같은 환경에서 BMS 적용 후에 인식 시간은 약 18분으로 약 44% 인식 시간이 단축됨을 알 수 있었다. 또한 후처리에 소요된 시간을 비교하면 BMS 적용 전은 17분이 걸린 것에 비해서 BMS 적용 후에는 4분이 소요되어 76%의 인식 속도 향상을 가져오게 되었다. BMS 적용 후에서 반음소 개수 선정시 임계값을 주었을 때 인식 시간과 반음소 개수를 고정했을 때의 인식 시간을 비교하면 반음소 개수를 7로 고정했을 때의 인식 시간이 더 좋음을 알 수 있다. 이 결과는 인식에 사용된 반음소 개수의 차이가 인식 시간에 그대로 반영됨을 알 수 있다.

위의 결과를 통해서 BMS 알고리즘을 이용했을 때 인식 시간은 사용 전에 비해서 인식 시간의 단축을 가져올 수 있었고, 또한 반음소의 개수에 따라서 인식 시간의 향상을 가져 올 수 있음을 알 수 있다.

&lt;표 2&gt; 인식 시간 비교

	BMS 적용 전	BMS 적용 후	
		임계값 18	반음소 개수 7
전체 인식 시간(분)	32	18	17
1문장 당 처리 시간(초)	3.87	2.18	2.06
전체 인식 중 후처리에 소요된 시간(분)	17	4	2.5

## 5. 결 론

본 논문에서는 기존에 핵심어 검출기 후처리부로 사용되고 있던 RLJ-CM에 BMS 모델을 이용한 신뢰도 개선 방안을 제안하였다. RLJ-CM에서 취약점인 반음소 모델 선정 과정에서 모든 음소 세트를 반음소 모델로 사용하게 되어 신뢰도의 성능이 떨어질 뿐 아니라 계산량의 증가를 가져오게 된다. 이로 인하여 반음소 모델 세트에 따라서 신뢰도가 변하였고, 기존의 RLJ-CM에서의 취약점인 반음소 모델 선정 과정에서 BMS 알고리즘을 이용한 새로운 반음소 세트를 정의하였다.

제안한 BMS-CM은 반음소 모델 선정 과정에서 신뢰도 계산 대상 음소와 가까운 음소만을 선택하여 반음소 세트로 사용하여 평균과 표준편차를 재구성하는 방법으로 FA 발생 시에 기존의 CM에 비해 보다 견인한 성능을 보임을 알 수 있었다. 또한 BMS 알고리즘을 이용하여 기존의 신뢰도에 비해 반음소 사용을 줄이게 되었다. 이를 통하여 인식 과정에서 후처리 과정에서 필요한 반음소 확률 계산에서 불필요한 계산량을 줄일 수 있게 되어 인식 시간을 단축시키게 되었다. 후처리 과정에서의 계산 시간 감소로 인하여 전체 인식 시간이 감소됨을 알 수 있었다. 그러나 반음소 선정 과정에서 너무 작은 반음소 세트를 사용하게 되면 인식 성능이 저하됨을 알 수 있었다. 이를 통하여 반음소 세트를 적절히 선정한다면 기존의 신뢰도에 비해 보다 나은 성능뿐만 아니라 인식 속도 또한 향상시킬 수 있을 것이다.

## 참 고 문 헌

- [1] P. Jeanrenaud, K. Ng et al., "Phonetic-based word spotter: Various configurations and application to event spotting", *Proc. of Eurospeech 93*, Vol. 2, pp.1057-1060, 1993.
- [2] S. J. Young, N. H. Russell, J. H. S. Thornton, "Token Passing: a simple conceptual model for connected speech recognition systems", Cambridge University Engineering Department, 1989.
- [3] 김철, 이경록 et al., "정규화 신뢰도를 이용한 핵심어 검출 성능 향상", *한국음향학회지*, 21권, 4호, pp.380-386, 2002.
- [4] M. G. Rahim, C. H. Juang, W. Chou, "Discriminative utterance verification using minimum string verification error (MSVE) training", *Proc. of ICASSP 96*, pp.3585-3588, 1996.
- [5] S. J. Young, N. H. Russell, J. H. S. Thornton, "Token Passing: a simple conceptual model for connected speech recognition systems", Cambridge University Engineering Department, 1989.
- [6] 이경록, 서현철 et al., "하이브리드 신뢰도를 이용한 제한 영역 핵심어 검출 성능 향상", *한국음향학회지*, 21권, 7호, pp.632-640, 2002.

- [7] B. Mak, E. Barnard, "Phone clustering using the bhattacharyya distance", *Center for Spoken Language Understanding Oregon Graduate Institute of Science and Technology* 2000, 2000.
- [8] C. Sanderson, K. K. Paliwal, "Likelihood normalization for face authentication in variable recording conditions", *Proc. ICIP*, 2002.
- [9] 김기태, 문광식 et al., "가변어휘 단어 인식에서의 미등록어 거절 알고리즘 성능 비교", *한국음향학회지*, 20권, 2호, pp.27-34, 2001.

접수일자: 2003년 05월 23일

게재결정: 2003년 06월 12일

▶ 김병돈(Byoung-Don Kim)

주소: 500-757 광주광역시 북구 용봉동 300 전남대학교 공과대학 전자정보통신공학과

소속: 전남대학교 공과대학 전자정보통신공학과 IA&R 실험실

전화: 062) 530-0472

FAX: 062) 530-0472

E-mail: doni96@moiza.chonnam.ac.kr

▶ 김진영(Jin-Young Kim)

주소: 500-757 광주광역시 북구 용봉동 300 전남대학교 공과대학 전자정보통신공학과

소속: 전남대학교 공과대학 전자정보통신공학과

전화: 062) 530-1757

FAX: 062) 530-0472

E-mail: beyondi@chonnam.ac.kr

▶ 최승호(Seung-Ho Choi)

주소: 520-714 전남 나주시 대호동 252번지 동신대학교

소속: 동신대학교 정보과학대학 멀티미디어통신공학과

전화: 061) 330-3194

FAX: 061) 330-2209

E-mail: shchoi@dsu.ac.kr