

# 바타차라 거리 측정 기법을 사용한 가우시안 모델 기반 음소 인식 향상

오 상 엽<sup>\*\*</sup>

## 요 약

기존의 어휘 인식에서는 일반적인 벡터 값을 데이터베이스를 이용하여 구하므로 탐색 중에 형성되는 음소를 처리하지 못하는 문제점을 제공하며, 음소 데이터에 대한 모델을 구성할 수 없는 단점으로 인하여 가우시안 모델의 정확성을 확보하지 못하게 된다. 따라서 본 논문에서는 음소가 갖는 특징을 기반으로 바타차라 거리 측정법을 이용하여 정확한 음소로 인식할 수 있도록 유도하였으며 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다. 연속 확률 분포의 공유로부터 가우시안 모델 최적화를 실험한 결과 향상된 신뢰도로 인해 높은 인식 성능을 확인하였으며, 본 논문에서 제안한 바타차라 거리 측정법을 이용하여 실험한 결과 기존의 방법들에 비하여 평균 1.9%의 성능 향상을 나타내었으며, 신뢰성을 바탕으로 인식율에서 평균 2.9%의 성능 향상을 나타내었다.

## Improving Phoneme Recognition based on Gaussian Model using Bhattacharyya Distance Measurement Method

Sang-Yeob Oh<sup>\*\*</sup>

### ABSTRACT

Previous existing vocabulary recognition programs calculate general vector values from a database, so they can not process phonemes that form during a search. And because they can not create a model for phoneme data, the accuracy of the Gaussian model can not secure. Therefore, in this paper, we recommend use of the Bhattacharyya distance measurement method based on the features of the phoneme -thus allowing us to improve the recognition rate by picking up accurate phonemes and minimizing recognition of similar and erroneous phonemes. We test the Gaussian model optimization through share continuous probability distribution, and we confirm the heighten recognition rate. The Bhattacharyya distance measurement method suggest in this paper reflect an average 1.9% improvement in performance compare to previous methods, and it has average 2.9% improvement based on reliability in recognition rate.

**Key words:** Gaussian model(가우시안 모델), Bhattacharyya distance measurement(바타차라 거리 측정법), confidence(신뢰도)

### 1. 서 론

어휘 인식에서는 유사한 음소와 부정확한 어휘 제

공으로 인해 오류가 존재하며, 부정확한 어휘의 입력으로 특징을 추출하여 인식할 경우 오인식의 결과가 나타나거나 유사한 음소로 인식되며 특징 추출이 제

※ 교신저자(Corresponding Author): 오상엽, 주소: 경기도 성남시 수정구 복정동 산65(461-702), 전화: (031)750-5798, FAX: (02)426-9159, E-mail: syoh@kyungwon.ac.kr  
접수일: 2010년 9월 1일, 수정일: 2010년 12월 12일

완료일: 2011년 1월 13일

<sup>\*</sup> 종신회원, 경원대학교 IT대학 컴퓨터소프트웨어 교수  
※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

대로 이루어지지 않으면 음소 인식 시 유사한 음소로 인식하게 된다. 이를 위한 신호 처리 단계에서의 어휘 인식 오류 보정에 대한 여러 가지 연구가 진행되고 있다[1].

사용범위가 넓고 다양한 화자 독립적인 최근의 시스템에서 전처리인 신호 처리만으로 인식의 효율을 높이는 것은 매우 제한적이므로 어휘의 단순한 신호 처리 위주의 인식 결과로부터 좀 더 신뢰할 수 있는 결과를 얻기 위한 어휘 후처리에서 오류 보정에 대한 연구가 진행되고 있다[2]. 또한, 미등록어 거절 방식은 구현 방식에 따라 발화 검증 방식과 핵심어 검출 방식으로 구분된다. 핵심어 검출 방식은 문법을 설계할 때 핵심어만 고려하고 나머지 단어는 가비지(garbage) 모델을 사용하여 불필요한 단어를 제거하여 사용하는 방법이다[3]. 발화 검증 방식은 인식 결과를 확인하는 과정이 추가되며 필러(filler) 모델을 이용한다. 필러 모델은 구성방식이 단어 기반이므로 가변 어휘 단어 인식 시스템을 위한 발화 검증 구현을 위해서는 매 음소단위의 검증기능이 있어야하며 반음소 모델을 사용하는 방식이 제안되고 있다[4].

어휘 인식에서는 인식 학습 시 나타나지 않는 미등록어의 트라이 폰이 나타나므로 모델 파라미터들의 초기 추정치를 생성할 수 없으므로 음소 데이터에 대한 모델을 구성할 수 없다. 이러한 단점으로 인하여 가우시안 모델의 정확성이 떨어지게 되어 인식의 신뢰도가 저하되고, 일반적인 벡터 값을 데이터베이스를 이용하여 구하므로 탐색 중에 형성되는 음소를 처리하지 못하는 문제점을 제공하지만, 본 논문에서는 이를 개선하기 위하여 본 논문에서는 바타차라 거리 측정법을 이용한 가우시안 모델을 최적화한 방법을 제안하여 가우시안 모델의 정확성을 향상하고, 음소를 관리 및 제어할 수 있도록 하였다.

본 연구에서는 모노폰으로 훈련시킨 훈련 데이터의 음소에 HMM(Hidden Markov Model) 특징 추출 방법을 이용하며, 유사한 음소는 바타차라 거리 측정법을 이용하여 정확한 음소로 인식할 수 있도록 유도하였으며 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다.

연속 확률 분포의 공유로부터 가우시안 모델 최적화를 실험한 결과 향상된 신뢰도로 인해 높은 인식 성능을 확인하였으며 미등록어 거절 알고리즘이 기존의 방법들에 비하여 평균 거절율이 1.9%의 성능

향상을 나타내었다. 따라서 등록되지 않은 어휘를 거절하여 등록된 어휘에서 인식할 수 있는 신뢰성을 향상시켰으며 신뢰성을 바탕으로 인식율에서 평균 2.9%의 성능 향상을 나타내었다. 본 연구의 결과인 2.9%의 성능 향상은 미등록어 거절과 신뢰성 향상으로 인한 높은 인식율의 향상을 가져왔다.

## 2. 관련연구

### 2.1 미등록어

음성 인식 시스템은 인식 대상 단위에 따라 고립 단어 인식 시스템과 연속 음성 인식 시스템으로 구분된다. 고립 단어 인식 시스템은 사용자가 한 단어만을 말하거나 단어와 단어 사이에 구분을 줌으로써 시스템이 단어 단위로 인식을 하는 시스템이다. 연속 음성 인식 시스템은 여러 단어나 문장을 자연스럽게 말할 때 시스템은 그 결과를 여러 단어나 문장 단위로 보여주는 것이다. 그러나 두 시스템 모두 미리 정해 놓은 특정 인식 대상 단어가 입력될 것이라는 가정 하에 음성 인식 기능을 수행하며 사용자의 실수 또는 고의로 인식 대상 단어를 말하면 대상 단어중의 하나로 잘못 인식하는 결과를 보여주므로 다른 단어로 인식해 버리는 문제점이 있다[5].

고립 단어 인식 시스템의 경우는 그런 입력이 들어왔을 때 그 단어에 대해서만 오인식을 하게 될 것이므로 그리 큰 문제가 되지 않는다고 볼 수 있다. 그러나 연속음성인식 시스템의 경우에는 음향학적 처리기가 동작을 하게 되므로 더 큰 문제를 야기시킬 수 있다. 사용자가 발화한 문장 중 한 단어만 음성 인식 대상 단어가 아니라 하더라도 그 단어가 오인식됨으로 인해서 그 뒤에 발화된 단어에까지 영향을 미쳐 오인식을 높하게 되므로 더 치명적인 결과를 초래하게 된다.

음소 필러 모델을 사용한 핵심어 검출 방식을 이용해서 미등록어를 거절시키는 방법이다. 핵심어 검출 방식은 핵심어 모델과 필러 모델을 사용하는 연결 단어 인식 알고리즘을 기반으로 사용하고 있다. 필러 모델들은 핵심어에 해당하지 않는 음성 구간들인 비핵심어들과 비음성, 묵음 또는 배경 잡음 구간들을 표현하는데 사용된다[6].

기존 미등록어 거절 방법에는 가변 어휘 단어 시스템에서 비터비 탐색시 사용되는 네트워크망을 그

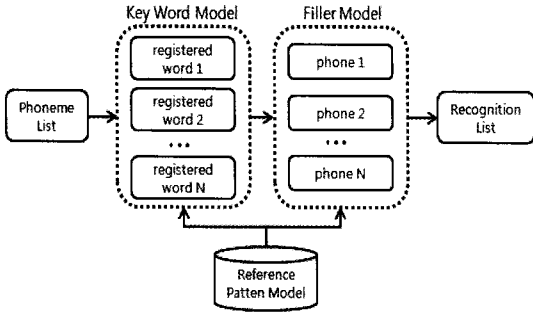


그림 1. 핵심어 모델과 필러 모델 분류 과정

림 1과 같이 구성한다. 구성된 네트워크망에서 인식 된 결과는 등록어들과 음소들의 열로 나타나게 된다.

### 2.2 HMM(Hidden Markov Model)

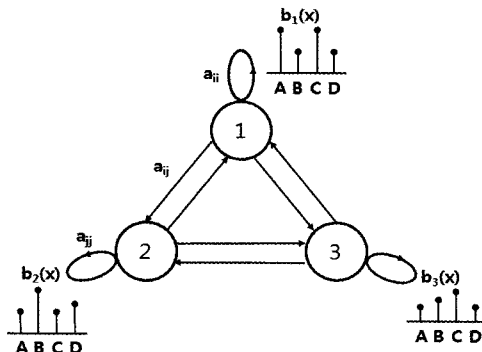
HMM은 관측할 수 없는 “hidden”과정과 음성 신호로부터 “hidden”과정의 상태로 유도되는 음향학적 벡터를 연결하는 관측 과정으로 구성된다.

HMM에서는 관측할 수 없는 음성의 통계적인 특성을 관측 가능한 벡터열을 통해 추정함으로써 음성의 통계적인 변이성을 반영한다[7].

그림 2는 3-states HMM으로 음성 구간의 변이에 의한 상태전이 확률  $A_i$ 와 각 상태에서의 관측 심볼의 출력 확률  $B_j$ , 초기 확률  $\pi$ 로 표현할 수 있으며 다음 식(1)과 같이 정리되고 모델 파라미터의 집합으로 표현된다.

$$\lambda = (A, B, \pi) \tag{1}$$

식 (1)에서 상태전이 확률  $A_i$ 를 나타내면 식 (2)와 같이 표현된다.



State seq : 1 3 1 2 2 3  
Observed seq. (X) : A C D B A C

그림 2 상태의 수가 3인 HMM

$$A = \{a_{ij}\}, \quad a_{ij} = p[q_{t+1} = j | q_t = i] \tag{2}$$

상태전이 확률  $A$ 는  $a_{ij}$ 로 표시하며  $i$ 상태에서  $j$ 상태로 전이할 확률을 나타낸다.

식 (1)에서 관측 심볼의 출력 확률  $B_j$ 를 나타내면 식 (3)과 같이 표현된다.

$$B = \{b_j(k)\}, \quad b_j(k) = p[o_t = v_k | q_t = j] \tag{3}$$

관측 심볼의 출력 확률  $B$ 는  $b_j(k)$ 로 표시하며  $j$ 상태에서  $k$ 가 관측될 확률을 나타낸다.

식 (1)에서 초기 확률  $\pi$ 를 나타내면 식 (4)와 같이 표현된다.

$$\pi = \{\pi_i\}, \quad \pi_i = p[q_1 = i] \tag{4}$$

초기 확률  $\pi$ 는  $\pi_i$ 로 표시하며 초기 상태를 표현한다. 상태들은 다른 어떤 상태에도 갈 수 있도록 서로 연결되어 있으며 각 상태는  $\{1, 2, \dots, N\}$ 로 나타내며 시간  $t$ 에서의 상태는  $q_t$ 와 같이 나타낸다.

### 2.3 Bhattacharyya 거리 측정

어휘 인식, 문자 또는 영상 인식과 같은 패턴인식에서 거리의 개념은 패턴들이 특징 공간상에서 서로 얼마나 떨어져 있는지를 통하여 패턴들 사이의 비슷한 정도를 측정하기 위한 기준으로 사용한다. 특징 공간상에서 매우 근접한 거리에 있는 두 패턴은 거의 동일한 특징을 가지므로 큰 유사도를 갖는다. 이러한 거리 측정 방법에는 유클리디안 알고리즘, DTW 알고리즘, 바타차라 거리 측정 알고리즘이 사용되고 있다[8].

유클리디안 알고리즘은 실제 거리를 구하기 위해 사용하며 두 가지 개체의 속성 값들이 여러 개 일 경우 이들 속성 값들에 의한 두 개체 사이의 유사도를 구하여 확률이 높은 값을 인식한다. DTW 알고리즘은 비선형 시간 정규화를 갖는 패턴정합 알고리즘으로 공통적이고 균일하게 샘플간격을 갖게 음성패턴을 시간적으로 샘플화하여 인식한다.

바타차라 거리 측정은 오류율을 측정하여 거리를 계산하는 방법이다. 단순 거리 계산을 수행하는 방법이므로 실시간을 요구하는 인식과정에서는 일반적으로 동적 프로그램 기술인 비터비 코딩(Viterbi decoding) 방법을 이용하여 상태경로의 변이와 최적의 모델을 추정하여 인식한다[9].

전향, 후향 확률에 의한 연산을 이용하여 상태경

로를 추적하는 경우 인식율이 다소 우수한 반면 주어진 모든 상태에서의 출력 심볼의 확률을 전부 추정하므로 계산량과 복잡도가 증가하게 된다. 비터비 코딩의 경우 전향, 후향 확률추정을 이용한 코딩에 비해 인식율이 다소 저하되나 연산에서의 부하를 월등히 감소시키므로 일반적인 인식과정에서 이용한다.

### 3. 시스템 모델

#### 3.1 가우시안 최적화 모델

가우시안 최적화 모델은 주어진 표본 데이터 집합의 분포밀도를 단 하나의 확률밀도함수로 모델링하는 방법을 개선한 밀도 추정 방법으로 복수 개의 가우시안 확률밀도함수로 데이터의 분포를 모델링하는 방법이다. 단일한 가우시안으로는 모델링 할 수 없는 복수개의 중심점을 가지는 1차원 데이터와 2차원 환형 데이터에 대하여 견고하게 모델링된다[10].

확률밀도함수는 가우시안 분포뿐 아니라 다른 분포가 될 수도 있다. 가우시안 최적화 밀도는 단지 확률밀도함수를 가우시안 분포로 가정하는 경우이다. 결국 최종적인 전체 확률밀도함수는  $M$ 개의 가우시안 확률밀도함수의 선형 결합으로 식(5)와 같이 표현된다.

$$p(x|\theta) = \sum_{i=1}^M p(x|w_i, \theta_i) F(w_i) \quad (5)$$

여기서  $p(x|w_i, \theta_i)$ 는 데이터  $x$ 에 대하여  $w_i$  번째 성분 파라미터  $\theta_i$ 로 이루어진 확률밀도함수를 의미하며,  $F(w_i)$ 는 혼합 가중치로 각 확률밀도함수의 상대적인 중요도를 의미한다. 혼합 가중치를 사전확률과 같은 형태로  $\alpha_i$ 라고 하면 식(6)과 같은 제약조건이 따른다.

$$0 \leq \alpha_i \leq 1, \text{ and } \sum_{i=1}^M \alpha_i = 1 \quad (6)$$

확률밀도함수가 가우시안 분포를 따를 경우  $\theta_i$ 는 식(7)와 같은 파라미터 집합이 된다.

$$\theta_i = (\mu_1, \mu_2, \dots, \mu_M, \sigma_1, \sigma_2, \dots, \sigma_M, \alpha_1, \alpha_2, \dots, \alpha_M) \quad (7)$$

전체 모델을 이루는 각 가우시안 성분은 완전대각 또는 정방형 공분산 행렬의 형태를 가질 수 있다. 또한 혼합 성분의 개수는 학습 데이터 집합의 크기에 따라 조절 가능하다. 가우시안 최적화 모델로 데이터

의 분포를 모델링할 경우에 혼합 성분의 개수가 충분히 주어지고 적절한 파라미터 값들만 주어진다면 이론적으로는 어떠한 연속적인 분포도 완벽하게 추정하여 모델링한다.

단일 가우시안 출력 확률 밀도 함수를 갖는 3상태 단 음소 모델 초기 집합을 생성하고 훈련한다. 단 음소의 상태 출력 분포는 재 추정하여 훈련된 트라이폰 모델 집합을 초기화하기 위해 복사한다. 동일한 각 음소로부터 유도된 트라이폰들의 각 집합에 대해 대응되는 상태들을 군집화한다. 각 결과 군집에서 대표적인 상태가 선택되고 모든 군집 내의 상태들은 대표 상태로 묶이게 된다. 각 상태의 혼합 요소의 수를 증가시켜 재추정하여 모델의 정밀도를 향상시킨다.

그림 3은 입력되는 음선 신호들로부터 신호에 대한 특징 벡터를 추출한 "일", "이", "삼" 특징 벡터 값을 구성한다. 구성되어진 특징 벡터들로부터 가우시안 모델을 생성하기 위해 가우시안 값을 추정하고 모델화하는 과정을 나타내었다. 최적화된 모델들은 미리 생성되어진 가우시안 모델값들과 비교하여 인식하기 위해 사용된다.

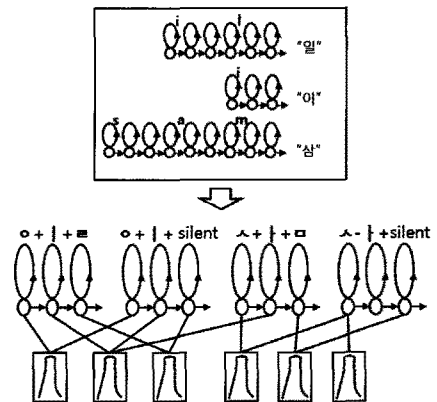


그림 3. 가우시안 최적화 모델

#### 3.2 유사 음소 인식 처리

유사 음소 인식 처리는 모노폰으로 훈련시킨 훈련 데이터의 음소에 HMM을 이용하여 특징을 추출하여 모델링하고 추출된 특징들의 모델들은 유사한 음소로 군집화하여 각 대표값을 구성한다.

음소 유사율은 두 음소 사이의 거리를 나타내며, 두 클래스간의 분리도를 측정하는 통계적 수단으로 두 개의 가우시안 분포 사이의 거리를 측정한다. 제

산의 단순함과 오류의 대한 경계값을 제공함으로써 유연성을 갖는다[11].

식 (8)은 두 클래스에 대해서 정의를 한다.

$$b = \int_{\Omega} [P(X|w_1)P(X|w_2)]^{\frac{1}{2}} dX \quad (8)$$

$P(X|w_i)$ 는 클래스  $w_i (i=1,2)$ 의 확률 밀도 함수이고,  $\Omega$ 는 확률 분포상에서 정의되는 랜덤 함수  $X$ 의 영역이다. 클래스의 확률 분포를 정규 분포로 표현하면 식 (9)와 같이 나타낸다.

$$b = \frac{1}{2}(u_2 - u_1)^T \left[ \frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (u_2 - u_1) \quad (9)$$

$u_i$ 와  $\Sigma_i$ 는 각각 클래스  $w_i$ 의 평균 벡터와 공분산 행렬을 표시하였다.

식 (10)에 의해  $[0, 1]$ 사이에서 분포하도록 정규화 한다.

$$S_p = \frac{b_{\max} - b_{xy}}{b_{\max} - b_{\min}} \quad (10)$$

$S_p$ 는 두 음소의 유사율을 측정하였으며  $b_{\max}$ 와  $b_{\min}$ ,  $b_{xy}$ 는 최대값, 최소값, 두 음소  $x$ 와  $y$ 의 거리를 나타낸다. 유사한 음소일수록 1에 가깝고, 유사하지 않을수록 0에 가깝게 나타낸다.

### 3.3 효율적 특징 추출 처리

효율적 특징 추출 처리 과정에서는 군집 모델에 포함되지 않는 특징들에 대해 HMM을 이용하여 이미 구성되어진 모델들 함께 특징을 재추출하여 군집들의 모델에 포함되는 특징을 추출한다.

HMM은 초기 ( $t=0$ )에 상태  $i$ 의 확률  $\pi_i = \Pr(s_0 = i)$ , 상태  $i$ 에서  $j$ 로의 천이 확률  $a_{ij} = \Pr(s_t = i, s_{t+1} = j)$ , 상태  $j$ 에서 심볼  $k$ 를 관측할 확률  $b_j(k) = \Pr(x_t = k | s_t = j)$ 로 표현한다. 임의의 음성 특징 벡터의 관측열  $O = (o_1, o_2, \dots, o_T)$ 이 사실임을 가정할 때 주어진 N-states HMM 모델에서의 상태열이  $q = (q_1, q_2, \dots, q_T)$  라면 결국 관측열의 확률은 다음 식(11)과 같이 주어진다[12].

$$P(O|q, \lambda) = \prod_{t=1}^T P(o_t | q_t, \lambda) \quad (11)$$

초기 상태  $t=1$ 에서 확률  $a_{s_0, s_1}(\pi_{s_1})$ 로 천이가 시작되며, 관측  $O_1$ 는 출력 확률  $b_{s_1}(O_1)$ 로써 생성이 된다. 초기 상태  $s_1$ 에서 상태  $s_2$ 로의 천이는 천이 확률  $a_{s_1, s_2}$

로 이루어지며, 대응되는 상태  $s_2$ 에서의 관측  $O_2$ 를 생성될 확률은  $b_{s_2}(O_2)$ 가 된다. 이러한 과정은 상태  $s_{T-1}$ 에서 마지막 상태  $s_T$ 로  $a_{s_{T-1}, s_T}$ 의 확률로 천이되어 기호  $O_T$ 를 출력 확률  $b_{s_T}(O_T)$ 로 생성할 때까지 계속된다. 식 (11)를 다시 표시하면 식 (12)과 같이 표현된다.

$$P(O|q, \lambda) = \sum_{q \in Q} \prod_{t=1}^T a_{q_{t-1}, q_t} b_{q_t}(o_t) \quad (12)$$

이러한 확률 계산은 음성 구간에 따라 모델이 저수 함수적으로 증가하는 상태 열을 갖기 때문에 쉽게 계산할 수 없고 계산량이 지나치게 방대해지므로 전향, 후향 알고리즘을 이용하여 HMM 모델의 관측열의 확률을 추정한다[13].

### 3.4 바타차라 거리 측정

추출되어진 모델들은 바타차라 거리 측정법을 이용하여 군집들에 포함되는 모델로 재구성되어진다. 거리 측정의 목적은 가장 비슷한 것을 찾아내는 것을 의미한다. 출력 확률 분포를 연속 확률 밀도로 갖는 가우시안의 경우 오류를 측정에 기반을 두고 있는 바타차라 거리 측정법을 사용하였으며 다음 수식 (13)과 같다.

$$d(i, j) = \left[ \frac{1}{n} \sum_{k=1}^n \frac{(\mu_{ik} - \mu_{jk})^2}{\sigma_{ik} - \sigma_{jk}} \right]^{\frac{1}{2}} \quad (13)$$

$n$ 은 데이터의 차수를 나타내고  $\mu_{ik}$ 와  $\sigma_{ik}$  상태  $s$  ( $i$  혹은  $j$ )의 가우시안 분포의  $k$ 번째 평균과 분산이다. 혼합 거리는 두 가우시안 사이의 겹치는 부분을 측정한다. 거리 범위는 0에서  $\infty$ 까지의 값을 가지며 각각 가우시안과 완전히 겹치거나 전혀 겹치지 않음을 나타낸다. 실제로 가우시안 간의 겹치는 부분이 조금이라도 있게 되면 거리 측정은 절대  $\infty$ 가 되지 않는다. 혼합 거리 측정의 자세한 식은 다음 식 (14)과 같다.

$$B_d = \frac{1}{2}(\overline{\mu_1} - \overline{\mu_2})^T \times \left( \frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} \times (\overline{\mu_1} - \overline{\mu_2}) \quad (14)$$

$\mu_1$ 과  $\mu_2$ 는 각 분포의 평균이며  $\Sigma_1$ 과  $\Sigma_2$ 은 공분산이다. 혼합 거리는 비슷한 가중치를 갖는 가우시안들이 조합되기가 더 쉽기 때문에 축적된 형태로 구하게 된다[14].

다음은 바타차라 거리 측정에 대한 알고리즘이며

입력되어진 두 모델로부터 벡터를 구하여 각각의 벡터 값들로부터 평균과 분산으로 계산하여 바타차라 거리 측정법을 사용하여 근접한 벡터들을 테이블로 구성하여 모델로 사용한다.

```

Begin
  Initialize the BhattaPhone
Job1 : Get Input Vector X1, X2
  If Vector X1, X2 exist then next
  else goto end
  Clear MeanList
  Add the MeanList to Evaluation(n)
  Process the phone transition and
  Add the Expand(n) to MeanList
  Insert BhattaPhone to BhattaPhoneframe
  Get the frame of word list BhattaPhone for likelihood
  Process word transition
  Process the likelihood(ALL n∈CovList)
  If BhattaPhone=NL Job2 else Job 3
Job2 : Recognition
  Process word transition
  Create the max likelihood of BhattaPhone table
Job3 : Non-Recognition
  Replace CovList, MeanList
  goto Job1
end
    
```

연속적으로 흡수하는 이웃 가우시안들로부터 공분산이 큰 단일 가우시안이 생겨나고, 외부의 것이 흡수되지 않는 것을 막아준다. 가중치 크기 (weighting scalar)  $B_s$ 는 가우시안의 가중치  $w_1$ 과  $w_2$ 의 함수이다.

$$B_s = \sqrt{\frac{w_1^2 + w_2^2}{2w_1w_2}} \quad (15)$$

$B_s$ 는  $w_1 \rightarrow w_2$  일때,  $B_s \rightarrow 1$ 이 된다. 반대로  $w_1 \gg w_2$  이거나  $w_1 \ll w_2$  일 때  $B_s \rightarrow \infty$ 가 된다. 이것은 가우시안 간의 가중치 차이를 최소화한다.

패턴인식에서 거리의 개념은 패턴들이 특징 공간 상에서 서로 얼마나 떨어져 있는지를 통하여 패턴들 사이의 비슷한 정도를 측정하기 위한 기준으로 사용하며 특징 공간상에서 매우 근접한 거리에 있는 두 패턴은 거의 동일한 특징을 가지므로 유사도가 크게 나타나게 되며 이러한 유사도를 측정하기 위해 바타차라 거리 측정 방법을 사용하였다.

바타차라 거리 측정 방법은 알고리즘에서 보이듯이 입력되어진 벡터들로부터 평균과 분산을 이용하여 근접한 거리를 측정하여 바타차라 테이블에 저장되어진 유사율을 통해 오류율을 측정하여 거리를 계

산하는 방법이며 단순 거리 계산을 수행하는 방법이므로 실시시간을 요구하는 인식 과정에서 사용된다.

#### 4. 실험 결과 및 분석

본 논문에서 제안한 바타차라 거리 측정법을 이용한 가우시안 모델을 최적화한 방법을 제안하여 가우시안 모델을 구성하여 인식 실험을 수행하였다. 실험은 다양한 가우시안 모델을 포함하고 있는 한국전 자통신연구원(ETRI)에서 제작한 음성 인식용 PBW (Phonetically Balanced Words) 445DB를 이용하였다. 어휘수가 총 445개로 구성되어 있으며 1명이 2회 발성한 것을 1개의 set로 구성하였다. 이러한 set이 남성음이 5set, 여성음이 5set으로 모두 10set으로 구성하였다[15]. 녹음된 데이터는 인식기 학습을 위해 MFCC 특성 추출 방법을 사용하였고 인식기는 SITEC에서 개발한 ECHOS[16]를 이용하였다.

미등록어 거절의 성능은 다음과 같은 항목을 기준으로 평가하였다[17].

##### 1) 등록어

- ① CA(Correctly Accepted for Keyword) 인식 대상 등록어를 제대로 accept한 경우의 확률
- ② FAI(False Accepted In-Grammar Word, Keyword) 인식 대상 등록어로 accept는 했지만 잘못 인식한 경우의 확률
- ③ FR(False Rejected for Keyword) 인식 대상 등록어를 말했는데 reject한 경우의 확률
- ④ CA+FAI+FR = 100%

##### 2) 미등록어

- ① CR(Correctly Rejected for OOV) 미등록어에 대해 reject한 경우의 확률
- ② FAO(False Accepted Out-of-Grammar Word, OOV) 미등록어인데 accept한 경우의 확률
- ③ CR+FAO = 100%

표 1과 표 2는 단어 패널티와 삽입된 음소들의 개수에 따른 성능을 보여준다. 단어 패널티와 인식된 결과의 삽입된 음소 개수를 이용하여 미등록어를 거절시키므로 임계값의 요소는 단어 패널티와 삽입된 음소의 개수가 된다. 표 1에서는 단어 패널티 1과 임계치 -30.0일 때의 발화 검증의 성능을 나타냈으며

표 1. 발화 검증의 성능(단어 패널티 1=-30.0일 때)

데이터 종류 삽입된 음소개수	시험용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	84.15	80.12	3.45	19.88	12.40
2개이상 거절	83.24	88.96	2.97	11.04	13.79
1개이상 거절	71.92	94.31	1.88	5.69	26.20
데이터 종류 삽입된 음소개수	평가용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	88.46	75.24	2.74	24.76	8.80
2개이상 거절	85.64	84.52	2.71	15.48	11.65
1개이상 거절	80.17	91.56	1.69	8.44	18.14

표 2. 발화 검증의 성능(단어 패널티 1=-40.0일 때)

데이터 종류 삽입된 음소개수	시험용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	93.21	59.12	4.14	40.88	2.65
2개이상 거절	90.61	70.61	3.92	29.39	5.47
1개이상 거절	82.84	82.11	2.34	17.89	14.82
데이터 종류 삽입된 음소개수	평가용 데이터				
	CA	CR	FAI	FAO	FR
3개이상 거절	96.14	58.14	3.84	41.86	0.02
2개이상 거절	92.53	70.35	2.62	29.65	4.85
1개이상 거절	85.18	74.91	2.11	25.09	12.71

표 2에서는 단어 패널티 1과 임계치 -40.0일 때의 발화 검증의 성능을 나타내었다.

음소 유사율의 신뢰도 정규화에 따른 신뢰도를 확인하기 위하여 바타차라 거리 측정법을 이용한 가우시안 모델을 최적화한 방법에 적용하였다. 가우시안

표 3. 오류 보정율 비교

오류 보정	인식율(%)	거절율(%)
error pattern	91.3	7.9
semantic	90.2	8.3
Gaussian optimization	93.6	9.8

모델을 최적화한 방법에 대한 인식율과 거절율은 식 (13)과 식 (15)를 이용하여 표 3에 나타내었다.

표 3은 기존의 에러 패턴 학습을 이용한 방법[18, 19]인 error pattern과 의미기반의 방법[20]인 semantic 그리고 본 논문의 제안 방법인 가우시안 모델 최적화의 결과를 나타내었다.

에러 패턴 학습을 이용한 미등록어 거절의 경우 7.9%, 의미 기반의 미등록어 거절의 경우 8.3%의 미등록어 거절율을 보였으며, 본 논문에서 제안한 가우시안 최적화 모델을 이용한 미등록어 거절은 9.8%로 미등록어 거절 알고리즘이 기존의 방법들에 비하여 평균 거절율이 1.9%의 성능 향상을 나타내었다. 따라서 등록되지 않은 어휘를 거절하여 등록된 어휘에서 인식할 수 있는 신뢰성을 향상시켰으며 신뢰성을 바탕으로 인식율에서 평균 2.9%의 성능 향상을 나타내었다. 이는 기존 등록되어 학습을 수행한 모델에 대한 인식율을 평가하는 기존에는 부족하지만 등록되지 않아 학습을 수행하지 않은 모델을 대상으로 인식율을 측정할 경우 데이터를 학습하지 않아 기존 모델과의 인식율이 저조하게 나타난다. 따라서 본 연구의 결과인 2.9%의 성능 향상은 미등록어 거절과 신뢰성 향상으로 인한 높은 인식율의 향상을 가져왔다.

### 5. 결론

어휘 인식에서는 인식 학습 시 나타나지 않는 미등록어의 트라이 폰이 나타나므로 모델 파라미터들의 초기 추정치를 생성할 수 없으므로 음소 데이터에 대한 모델을 구성할 수 없다. 이러한 단점으로 인하여 가우시안 모델의 정확성이 떨어지게 되어 인식의 신뢰도가 저하되고, 일반적인 벡터 값을 데이터베이스를 이용하여 구하므로 탐색 중에 형성되는 음소를 처리하지 못하는 문제점을 제공하지만, 본 논문에서는 이를 개선하기 위하여 본 논문에서는 바타차라 거리 측정법을 이용한 가우시안 모델을 최적화한 방법을 제안하여 가우시안 모델의 정확성을 향상하고,

음소를 관리 및 제어할 수 있도록 하였다.

본 연구에서는 모노폰으로 훈련시킨 훈련 데이터의 음소에 HMM(Hidden Markov Model) 특징 추출 방법을 이용하며, 유사한 음소는 바타차라 거리 측정법을 이용하여 정확한 음소로 인식할 수 있도록 유도하였으며 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다.

연속 확률 분포의 공유로부터 가우시안 모델 최적화를 실험한 결과 향상된 신뢰도로 인해 높은 인식 성능을 확인하였으며 미등록어 거절 알고리즘이 기존의 방법들에 비하여 평균 1.9%의 성능 향상을 나타내었으며, 신뢰성을 바탕으로 인식율에서 평균 2.9%의 성능 향상을 나타내었다. 미등록어 거절이 높은 것은 등록되어 있지 않은 어휘에 대해 인식하지 않는 것을 나타내므로 인식 성능 향상과 어휘 인식의 신뢰성에 영향을 미치게 된다. 따라서 전체적인 어휘 인식 시스템의 미등록어의 거절로 인한 인식율의 향상을 가져오게 된다.

## 참 고 문 헌

- [1] 안찬식, 오상엽, "공유모델 인식 성능 향상을 위한 효율적인 연속 어휘 군집화 모델링," 한국컴퓨터정보학회지, 제15권, 제1호, pp. 177-183, 2010. 1.
- [2] 김우성, 구명완 "반음소 모델링을 이용한 거절 기능에 관한 연구," 한국음향학회지, 제18권, 제3호, pp. 3-9, 1999. 3.
- [3] 문광식, 김희린, 정재호, 이영직, "가변어휘 단어 인식에서의 미등록어 거절 알고리즘의 성능비교," 신호처리합동학술대회논문집, 제12권, 제1호, pp. 305-308, 1999. 10.
- [4] 방기덕, 강철호, "가변 신뢰도 문턱치를 사용한 미등록어 거절 알고리즘에 대한 연구," 한국멀티미디어학회논문지, 제11권, 제11호, pp. 1471-1479, 2008. 11.
- [5] 안찬식, 오상엽, "MLHF 모델을 적용한 어휘 인식 탐색 최적화 시스템," 한국컴퓨터정보학회지, 제14권, 제10호, pp. 217-223, 2009. 10.
- [6] 김용현, 정민화, "에러패턴 학습과 후처리 모듈을 이용한 연속 음성 인식의 성능향상," Proc. KISS Spring Semiannual Conf. 제27권, 제1호, pp. 441-443, 2000. 4.
- [7] A. S. Manos and V. W. Zue, "A study on out-of- vocabulary word modeling for a segment-based keyword spotting system," Master Thesis, MIT, 1996.
- [8] 김동주, 김한우, "문맥가중치가 반영된 문장 유사도 척도," 전자공학회 논문지, 제43권, 제6호, pp. 496-504, 2006.
- [9] L. R. Bahl, P. V. deSouza, P. S. Gopalakrishnan, D. Nahamoo, and M. Picheny, "A Fast Match for Continuous Speech Recognition Using Allophonic Models," InProc. IEEE ICASSP-92, Vol.1, pp. 17-21, 1992.
- [10] L. R. Rabiner, B. H. Juang, "Fundamentals of speech recognition," Prentice Hall, 1993.
- [11] T. Jitsuhiro, S. Takatoshi, and K. Aikawa, "Rejection of out-of-vocabulary words using phoneme confidence likelihood," ICASSP, pp. 217-220, 1998.
- [12] 이경록, 김철, 김진영, 최승호, 최승호, "정규화 신뢰도를 이용한 핵심어 검출 성능 향상," 한국음향학회지, 제21권, 제4호, pp. 380-386, 2002. 5.
- [13] 김동주, 김한우, "문맥가중치가 반영된 문장 유사도 척도," 대한전자공학회논문지, 제43권, 제6호, pp. 496-504, 2006.
- [14] 김상운, 신성효, "ML/MMSE를 이용한 HMM-Net 분류기의 학습에 대한 실험적 고찰," 대한전자공학회논문지C, 제36C권, 제6호, pp. 44-51, 1999. 6.
- [15] S. Young, D. Kershaw, J. Odell, D. Ollason, Valtcher, P. Woodland, "The HTK Book," Cambridge University Engineering Department, 2002.
- [16] 권석봉, 윤성락, 장규철, 김용래, 김봉완, 김희린, 유창동, 이용주, 권오욱, "한국어 음성인식 플랫폼(ECHOS)의 개선 및 평가," 대한음향학회지:말소리, 제59호, pp. 53-68, 2006. 9.
- [17] 최승호, "정규화 신뢰도 기반 가변 어휘 고립 단어 인식기의 거절기능 성능 분석," 한국음향학회지, 제25권, 제2호, pp. 96-100, 2006. 2.
- [18] K. Demuynck, J. Duchateau, and D. Van Compernelle, "A static lexicon network repre-



sentation for cross-word context dependent phones,” In Proc. EUROSPEECH, Vol.1, pp. 143-146, 1997.

- [19] 김기태, 문광식, 김희린, 이영직, 정재호, “가변 어휘 단어 인식에서의 미등록어 거절 알고리즘 성능 비교,” 한국음향학회지, 제20권, 제2호, pp. 27-34, 2001. 2.
- [20] M. W. Jeong, B. C. Kim, and G. G. Lee, “Semantic-oriented error correction for spoken query processing,” Proc. IEEE Workshop on ASRU, pp.156-161, Nov, 2003.



오 상 엽

1999년 광운대학교 전자계산학과 박사

1993년 ~ 현재 경원대학교 IT대학 컴퓨터소프트웨어 교수

관심분야: 소프트웨어공학, 버전 관리, 소프트웨어 재사용, 형상관리, 객체지향, 음성

인식, 분산 처리, 음성/음향 신호처리