韓國컴퓨터情報學會 論文誌
第15卷 第6號, 2010. 6.

2010-15-6-4-1

바타챠랴 거리 측정법을 이용한 음소 유사율 오류 보정 개선 시스템

안 찬 식*, 오 상 엽**

Phoneme Similarity Error Correction System using Bhattacharyya Distance Measurement Method

Chan-Shik Ahn*, Sang-Yeob Oh**

요 약

어휘 인식 시스템은 부정확한 어휘 제공과 유사한 음소 인식으로 인식률이 저하되며 이는 유사한 음소인식 오인식과 효율적 특징 추출 처리를 위한 방법을 필요로 한다. 따라서 본 논문에서는 음소가 갖는 특징을 기반으로 바타챠랴 거리 측정법을 이용한 음소 유사율 오류 보정 개선 시스템을 제안하였다. 음소 유사율은 모노폰으로 훈련시킨 훈련 데이터의 음소에 HMM 특징 추출 방법을 이용하였으며 유사한 음소는 바타챠랴 거리 측정법을 이용하여 정확한 음소로 인식할 수 있도록 유도하여 인식률 향상 효과를 얻을 수 있었다. 이를 유클리디안 거리 측정법과 동적 타임 워핑 시스템에 비교한 시스템 성능 평가 결과 1.2%의 향상된 97.91% 인식률을 보였다.

Abstract

Vocabulary recognition system is providing inaccurate vocabulary and similar phoneme recognition due to reduce recognition rate. It's require method of similar phoneme recognition unrecognized and efficient feature extraction process. Therefore in this paper propose phoneme likelihood error correction improvement system using based on phoneme feature Bhattacharyya distance measurement. Phoneme likelihood is monophone training data phoneme using HMM feature extraction method, similar phoneme is induced recognition able to accurate phoneme using Bhattacharyya distance measurement. They are effective recognition rate improvement. System performance comparison as a result of recognition improve represent 1.2%, 97.91% by Euclidean distance measurement and dynamic time warping(DTW) system.

▸ Keyword : 음소 유사율(phoneme likelihood), 어휘 인식(vocabulary recognition), 오류 보정(error correction), 바타챠랴 거리 측정법(Bhattacharyya distance measurement)

• 제1저자 : 안찬식 교신저자 : 오상엽
• 투고일 : 2010. 03. 08, 심사일 : 2010. 03. 10, 게재확정일 : 2010. 05. 04.
* 광운대학교 컴퓨터공학과 박사과정 **경원대학교 IT대학 컴퓨터소프트웨어
※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

Ⅰ. 서 론

어휘 인식에서는 유사한 음소와 부정확한 어휘 제공에서 오류가 존재한다. 이를 위한 신호 처리 단계에서의 어휘 인식 오류 보정에 대한 여러 가지 연구가 진행되고 있다[1].

사용범위가 넓고 화자 독립적인 최근의 시스템에서 전처리인 신호 처리만으로 인식의 효율을 높이는 것은 상당히 어려운 일이다. 따라서 어휘의 단순한 신호 처리 위주의 인식 결과로부터 좀 더 신뢰할 수 있는 결과를 얻기 위한 어휘 후처리에서 오류 보정에 대한 연구가 진행되고 있다[2].

기존의 오류 보정을 수행하는 잡음 채널 모델 기반의 오류 보정 방법에서는 어휘 인식기의 적용 환경과 실제 인식할 때의 차이를 극복하지 못하는 단점이 있다[3]. 이는 인식 과정에서 일정한 패턴의 오류가 발생한다는 가정 하에 발화 문장과 인식 문장을 비교하여 오류 패턴을 학습하고 적은 비용과 시간으로 오류를 보정할 수 있는 후처리 보정 방법으로 인식하려는 어휘에 대한 오류 패턴 DB가 필요하다[4]. 정보 검색 영역에서 사용되는 문장은 문장이 간결하고 사용자가 검색하고자 하는 핵심어로만 이루어진 경우가 많으므로 정보 검색 영역의 문장은 의미적으로 분석하기 힘들며 문장이 전체적으로 오인식 될 경우 적용이 불가능한 단점이 있다[5].

통계적 어휘 인식에서는 인식 단계에서 유클리디안 알고리즘, DTW(Dynamic Time Warping)알고리즘, 바타챠랴(Bhattacharyya) 거리 등의 패턴 정합 을 이용하여 학습되어진 패턴 중에서 유사도가 가장 높은 것을 인식 결과로 출력한다.

유클리디안 알고리즘은 두 개체의 속성 값들에 의한 유사도를 구하는 방법이고 DTW 알고리즘은 비선형 시간 정규화를 갖는 패턴정합 방법이며 바타챠랴 거리 측정은 오류율을 측정하여 거리를 계산하는 방법이다[6].

어휘 인식 시스템에서 인식률 저하는 부정확한 어휘 제공과 유사한 음소로 인식되어 오인식되는 오류가 존재한다. 부정확한 어휘의 입력으로 특징을 추출하여 인식할 경우 오인식의 결과가 나타나거나 유사한 음소로 인식되며 특징 추출이 제대로 이루어지지 않으면 음소 인식 시 유사한 음소로 인식하게 된다.

따라서 본 논문에서는 음소가 갖는 특징을 기반으로 바타챠랴 거리 측정법을 이용한 음소 유사율 오류 보정 개선 시스템을 제안하였다.

모노폰으로 훈련시킨 훈련 데이터의 음소에 HMM(Hidden Markov Model) 특징 추출 방법을 이용하여 구하였다. 유사한 음소는 바타챠랴 거리 측정법을 이용하여 정확한 음소로 인식할 수 있도록 유도하였으며 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다. 음소 유사율과 신뢰도를 이용하여 오류 보정률을 구하였으며 문법적인 어절은 인식시키고 비문법적인 어절은 오류 보정을 수행하여 재인식시킨다. 의미적으로 분석하기 힘든 핵심어로만으로 이루어진 문장을 복원하여 전체적으로 인식할 수 있도록 해결하였으며 어휘 인식 과정에서 오류로 판명된 어휘에 대하여 오류 보정을 수행하였다.

시스템 성능 평가는 음소 유사율과 신뢰도를 이용하여 오류 보정률을 구하였으며, 어휘 인식 과정에서 오류로 판명된 어휘에 대하여 오류 보정을 수행하였다. 이를 유클리디안 거리 측정법과 동적 타임 워핑 시스템에 비교한 시스템 성능 평가 결과 1.2%의 인식 향상률을 보였다.

본 논문의 구성은 2장에서 음소 유사율 보류 보정 시스템과 바타챠랴 거리 측정 방법에 대해서 설명하고 3장에서는 바타챠랴 거리 측정법을 이용한 음소 유사율 오류 보정 개선 시스템에 대하여 설명한다. 4장에서는 실험 환경과 실험 결과, 5장에서는 결론을 다룬다.

Ⅱ. 기존 연구

2.1 HMM(Hidden Markov Model)

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

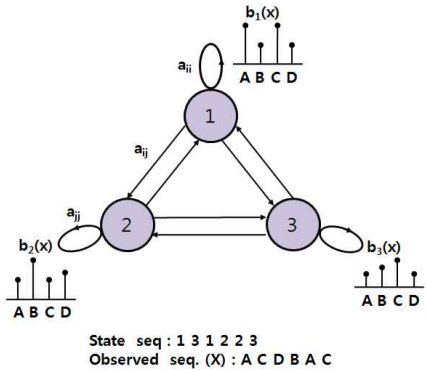

그림 1. 상태의 수가 3인 HMM
Figure 1. HMM with 3 states

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

그림 1은 3-states HMM으로 모델링한 상태 천이가 가능한 음성 모델을 나타냈으며 HMM 모델은 음성 구간의 변이에 의한 상태천이 확률 A_i와 각 상태에서의 관측 심볼의 출력 확률 B_j, 초기 확률 π로 표현할 수 있으며 다음 식(1)과 같이 정리되며 각 상태에서는 A, B, C, D의 관측 심볼을 출력한다.

$$\lambda = (A, B, \pi) \quad \cdots\cdots (1)$$

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

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

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

상태들은 다른 어떤 상태에도 갈 수 있도록 서로 연결되어 있으며 각 상태는 $\{1, 2, \ldots, N\}$로 나타내며 시간 t에서의 상태는 q_t와 같이 나타낸다. $A = \{a_{ij}\}$는 상태 전이 확률 분포를 나타내며 $B = \{b_j(k)\}$는 관측 기호 확률 분포를 나타낸다. $\pi = \{\pi_i\}$는 초기 상태 분포를 나타내며 $\lambda = (A, B, \pi)$는 모델 파라미터의 집합으로 표현된다.

2.2 음소 유사율 오류 보정

음소 유사율 오류 보정은 전처리에서의 잡음 제거 방법과 후처리에의 어휘 보정 방법으로 이루어진다. 전처리의 잡음 제거 방법은 워너 필터 이론에 의해서 제거된다. 입력 신호로부터 잡음 제거가 이루어지고 신호대 잡음비에 의해 잡음제거가 이루어진다. 입력된 어휘 신호의 잡음 제거는 프레임 단위로 실행된다. 입력된 신호가 프레임 단위로 나뉜 후에는 각 프레임 별로 스펙트럼을 추정하고 스펙트럼 평활화를 수행한다. 특징 추출은 인식에 유용한 성분을 어휘 신호로부터 뽑아내는 과정이다. 특징 추출은 일반적으로 정보의 압축, 차원의 감소 과정과 관련되며 특징의 좋고 나쁨은 어휘 인식률로 판단된다. 흔히 사용되는 특징추출 과정에서 청각 특성을 반영한 멜 필터 뱅크 분석과 프리엠퍼시스 필터를 사용한다[8].

음소 유사율 오류 보정은 오류로 발생된 인식 결과를 올바른 인식 결과로 보정한다. 화자 독립 적인 어휘 인식 시스템에서 신호 처리만으로 보정하기 힘든 오류를 어휘 인식 후처리에서 보정하는 것으로 어휘 인식 후처리에서 오류 보정은 인식된 결과에 대하여 오류를 보정한다.

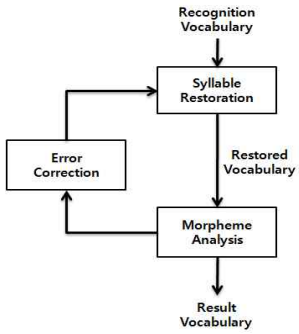

그림 2. 어휘 오류 보정
Fig. 2. Vocabulary error restoration

그림 2는 어휘 오류 보정의 구성도를 나타내고 있으며, 인식된 음소열에 대해 음절 복원하여 음운 변동이 적용되기 이전의 문자열로 복원시킨다[9].

2.3 Bhattacharyya 거리 측정 방법

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 동인한 특징을 가지므로 큰 유사도를 갖는다. 이러한 거리 측정 방법에는 유클리디안 알고리즘, DTW 알고리즘, 바타챠랴 거리 측정 알고리즘이 사용되고 있다[10].

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 패턴 정합 알고리즘으로 공통적이고 균일하게 샘플간격을 갖게 음성패턴을 시간적으로 샘플화하여 인식한다.

바타챠랴 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[11].

전향, 후향 확률에 의한 연산을 이용하여 상태경로를 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.

III. 음소 유사율 오류 보정 개선 시스템

3.1 시스템 모델

어휘 인식 시스템에서 부정확한 어휘 제공으로 인한 인식률의 저하는 유사한 음소인식으로 오인식되므로 특징 추출을 효율적 처리하여 유사한 음소로의 인식을 보정한다. 훈련 데이터의 음소에 HMM을 이용하여 특징을 추출하고 거리 측정을 이용하여 오류율을 측정하여 거리를 계산한 후 후처리에서 오류 보정하여 인식된 결과에 대하여 오류를 보정한다.

음소 특징을 기반으로 바타챠랴 거리 측정법을 이용하여 음소 유사율 오류 보정 개선 시스템을 그림 3와 같이 구성하였다.

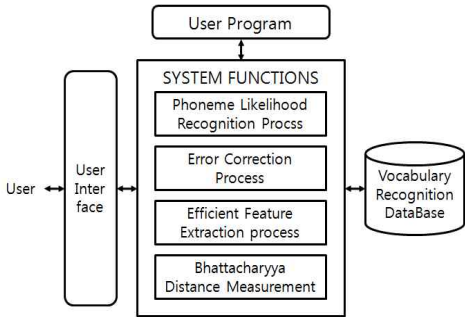

그림 3. 음소 유사율 오류 보정 개선 시스템 구성도
Figure 3. phoneme similarity error correction system architecture

유사 음소 인식 처리는 모노폰으로 훈련시킨 훈련 데이터의 음소에 HMM을 이용하여 특징을 추출하여 모델링하고 추출된 특징들의 모델들은 유사한 음소로 군집화하여 각 대푯값을 구성한다. 오류 보정 처리에서 군집에 포함된 모델들은 인식에 사용하게 되고 군집에 포함되지 않는 모델들은 분류하여 효율적 특징 추출 처리 과정으로 보내진다. 효율적 특징 추출 처리 과정에서는 군집 모델에 포함되지 않는 특징들에 대해 HMM을 이용하여 이미 구성되어진 모델들 함께 특징을 재추출하여 군집들의 모델에 포함되는 특징을 추출하게 되고 추출되어진 모델들은 바타챠랴 거리 측정법을 이용하여 군집들에 포함되는 모델로 재구성되어 인식에 사용된다.

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 대푯값을 구성한다.

음소 유사율은 두 음소 사이의 거리를 나타내며, 두 클래스간의 분리도를 측정하는 통계적 수단으로 두 개의 가우시안 분포 사이의 거리를 측정한다. 계산의 단순함과 오류의 대한 경계값을 제공함으로써 유연성을 갖는다[12].

식 (2)는 두 클래스에 대해서 정의를 한다.

$$b = -\ln \int_{\Omega} \left[P(X|w_1) P(X|w_2) \right]^{\frac{1}{2}} dX \quad \cdots\cdots (2)$$

$P(X|w_i)$는 클래스 $w_i\,(i=1,2)$의 확률 밀도 함수이고, Ω는 확률 분포상에서 정의되는 랜덤 함수 X의 영역이다. 클래스의 확률 분포를 정규 분포로 가정할 경우 식 (3)과 같이 나타낸다.

$$b = \frac{1}{8}(u_2 - u_1)^T \left[\frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (u_2 - u_1) + \frac{1}{2} ln \frac{|(\Sigma_1 + \Sigma_2)/2|}{|\Sigma_1|^{1/2} |\Sigma_2|^{1/2}} \quad \cdots\cdots (3)$$

u_1와 Σ_i는 각각 클래스 w_i의 평균 벡터와 공분산 행렬이다.

식 (4)에 의해 음소 유사율을 측정하였고 식 (5)에 의해 [0, 1]사이에 분포하도록 정규화한다.

$$d(i,j) = \left[\frac{1}{n} \sum_{k=1}^{n} \frac{(\mu_{ik} - \mu_{jk})^2}{\sigma_{ik}\ \sigma_{jk}} \right]^{\frac{1}{2}} \quad \cdots\cdots (4)$$

$$S_p = \frac{b_{\max} - b_{xy}}{b_{\max} - b_{\min}} \quad \cdots\cdots (5)$$

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

3.3 오인식 오류 처리

오류 보정 처리에서는 군집에 포함된 모델들은 인식에 사용하게 되고 군집에 포함되지 않는 모델들은 분류하여 효과적인 특징 추출 처리 과정으로 보내진다.

오류 보정을 수행하기 위해 우선순위 어절을 선정하여 어절을 구성하고 있는 음소의 신뢰도와 음소에 대한 타 음소와의 유사율을 사용하여 각각의 음소에 대한 후보 음절을 생성한다.

인식 오류의 경우 상위에 정확한 인식 결과를 포함하고 있으므로 후보 음절을 생성할 때에는 초성 자음과 모음에 대해서 상위의 음소 유사율을 사용한다.

오류 보정 처리는 신뢰도에만 의존하지 않고 신뢰도와 음소 유사율을 모두 고려하여야 한다. 음소의 신뢰도가 높아도 음소 유사율이 높은 비슷한 음소로 대치되어 오류가 발생할 수 있으며, 신뢰도가 낮아도 제대로 인식될 수 있기 때문이다[13]. 신뢰도가 낮으며 음소 유사율이 높은 음소를 가지고 있는 음소부터 오류 보정이 수행되어야 정확한 오류 보정을 수행할 수 있다. 식 (6)은 하나의 음절에 대한 오류 보정률을 나타낸다.

$$R_{syllable} = \frac{1}{n}\sum_{k=1}^{n}(1-\alpha_k)b_k \quad \cdots\cdots (6)$$

n은 한 음절의 음소 수, α_k는 k번째 음소의 신뢰도, b_k는 k번째 음소에 대한 유사율을 나타낸다. $\alpha=1$일 때, 오류 보정률은 0의 값을 가지며, $\alpha \neq 1$일 때는 $0 < R_{syllable} < 1$의 값을 가진다. 어떤 음소의 신뢰도가 낮으며 음소와 음소 유사율이 높은 음소일수록 오류 보정율은 커지게 되며 음소의 신뢰도가 높거나 음소와 음소 유사율이 낮은 음소일수록 오류 보정률은 작아진다.

다음은 오인식 오류 처리에 대한 알고리즘이다.

```
int frameX;
string Bhatta, BhattaPhone;
float like;
BPtr* bestBptr=0;
for(frameX=0; ; frameX++){
FVector obs=m_featureP->GetObservation(frameX);
if(obs.Size()==0)
break
if(obs.Size() != m_amodelP->VecSize())
Error("SearchBase","Recognize: Feature vector size mismatch.\n");
if(m_trace>0){
obs.Print(cout,"obs");}
}
```

3.4 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 i의 확률 $\pi_i = \Pr(s_0 = i)$, 상태 i에서 j로의 천이 확률 $a_{ij} = \Pr(s_t = i, s_{t+1} = j)$, 상태 j에서 심볼 k를 관측할 확률 $b_j(k) = \Pr(x_t = k | s_t = j)$로 표현한다. 임의의 음성 특징벡터의 관측열 $O = (o_1, o_2, \ldots, o_T)$이 사실임을 가정할 때 주어진 N-states HMM 모델에서의 상태열이 $q = (q_1, q_2, \ldots, q_T)$라면 결국 관측열의 확률은 다음 식 (7)과 같이 주어진다[14].

$$\begin{aligned} P(O|q,\lambda) &= \prod_{t=1}^{T} P(o_t|q_t,\lambda) \quad \cdots\cdots (7) \\ &= \sum_{q\in Q} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \cdots a_{q_{T-1}q_T} b_{q_T}(o_T) \\ &= \sum_{q\in Q} \prod_{t=1}^{T} a_{q_{t-1}q_t} b_{q_t}(o_t) \end{aligned}$$

초기 상태 $t=1$에서 확률 $a_{s_0 s_1}(\pi_{s_1})$로 천이가 시작되며, 관측 O_1는 출력 확률 $b_{s_1}(O_1)$로서 생성이 된다. 초기 상태 s_1에서 상태 s_2로의 천이는 천이 확률 $a_{s_1 s_2}$로 이루어지며, 대응되는 상태 s_2에서의 관측 O_2를 생성될 확률은 $b_{b_2}(O_2)$가 된다. 이러한 과정은 상태 s_{T-1}에서 마지막 상태 s_T로 $a_{s_{T-1}s_T}$의 확률로 천이되어 기호 O_T를 출력 확률 $b_{s_T}(O_T)$로 생성할 때까지 계속된다. 이러한 과정을 정의에 의하여 직접 계산하면 모든 시간 $t = 1, 2, \cdots, T$에서는 진행 가능한 상태 수는 N개가 되어 계산의 복잡도는 $O(N^T)$이 된다.

이러한 확률계산은 음성 구간에 따라 모델이 지수 함수적으로 증가하는 상태 열을 갖기 때문에 쉽게 계산할 수 없고 계산량이 지나치게 방대해지므로 전향, 후향 알고리즘을 이용하여 HMM 모델의 관측열의 확률을 추정한다[15].

3.5 바타챠랴 거리 측정

추출되어진 모델들은 바타챠랴 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 오류율 측정에 기반을 두고 있는 바타챠랴 거리 측정법을 사용하였으며 다음 수식 (8)과 같다.

$$d(i,j) = \left[\frac{1}{n}\sum_{k=1}^{n}\frac{(\mu_{ik}-\mu_{jk})^2}{\sigma_{ik}-\sigma_{jk}}\right]^{\frac{1}{2}} \quad \cdots\cdots (8)$$

n은 데이터의 차수를 나타내고 μ_{ik}와 σ_{ik} 상태 s (i 혹은 j)의 가우시안 분포의 k번째 평균과 분산이다. 혼합 거리는 두 가우시안 사이의 겹치는 부분을 측정한다[16]. 거리 범위는 0에

서 ∞까지의 값을 가지며 각각 가우시안과 완전히 겹치거나 전혀 겹치지 않음을 나타낸다. 실제로 가우시안 간의 겹치는 부분이 조금이라도 있게 되면 거리 측정은 절대 ∞가 되지 않는다. 혼합 거리 측정의 자세한 식은 다음 식 (9)과 같다.

$$B_{distance} = \frac{1}{8}\left(\overrightarrow{\mu_1} - \overrightarrow{\mu_2}\right)^T \times \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} \times \left(\overrightarrow{\mu_1} - \overrightarrow{\mu_2}\right) + \frac{1}{2} ln \frac{\left|\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)\right|}{\sqrt{|\Sigma_1||\Sigma_2|}} \quad \cdots\cdots (9)$$

μ_1과 μ_2은 각 분포의 평균이며 Σ_1과 Σ_2은 공분산이다. 혼합 거리는 비슷한 가중치를 갖는 가우시안들이 조합되기가 더 쉽기 때문에 축적된 형태로 구하게 된다[17].

다음은 바타차랴 거리 측정에 대한 알고리즘이다.

```
int frameX;
string Bhatta, BhattaPhone;
DoOneFrame(frameX,obs);
if(frameX>0 && m_backtracePeriod>0 && frameX%
   m_backtracePeriod==0){
bestBptr = BacktraceBhatta(frameX,Bhatta);
like=(bestBptr ? bestBptr->m_like : MIN_LIKELIHOOD);
cout << "\rframeX="<<frameX<<" like="<<like<<
          " Bhatta=" << Bhatta << endl;
}
InitOncePerFrame();
m_timer.m_recognize.Stop();
m_featureP->Stop();
int frameN = frameX;
if(frameN>0)
{
bestBptr= BacktraceHypo
            (frameN-1,hypo,m_searchResultP);
like=(bestBptr ? bestBptr->m_like : MIN_LIKELIHOOD);
frameN=(bestBptr ? bestBptr->m_frameX+1 : frameN);
}
if(frameN>0){
BacktracePhone(frameN-1,hypoPhone,m_searchResultP);
}
```

연속적으로 흡수하는 이웃 가우시안들로부터 공분산이 큰 단일 가우시안이 생겨나고, 외부의 것이 흡수되지 않는 것을 막아준다. 가중치 크기(weighting scalar) B_{scale}는 가우시안의 가중치 w_1과 w_2의 함수이다.

$$B_{scale} = \sqrt{\frac{w_1^2 + w_2^2}{2w_1w_2}} \quad \cdots\cdots (10)$$

B_{scale}는 $w_1 \to w_2$일때, $B_{scale} \to 1$이 된다. 반대로 $w_1 \gg w_2$이거나 $w_1 \ll w_2$일 때 $B_{scale} \to \infty$가 된다. 이것은 가우시안 간의 가중치 차이를 최소화한다[18].

IV. 실험결과 및 분석

본 논문에서 제안한 음소가 갖는 특징을 기반으로 바타차랴 거리 측정법을 이용한 음소 유사율 오류 보정 개선 시스템의 성능 검증을 위하여 인식 실험을 수행하였다. 훈련과정과 실험환경과의 불일치 문제를 해결하기 위해 잡음처리는 워너 필터를 사용하였다. 어휘 목록은 서울 시내의 지역명 200개, 지하철역명 100개로 구성하였다. 인식 실험에서는 실험에 참가한 화자가 어휘 목록을 5회 발음하여 총 1500단어를 대상으로 실험을 수행하였다. 어휘는 실내 환경과 잡음 환경에서 이동기기 등에 내장되어 있는 내장형 마이크로폰을 사용하여 16kHz Mono로 녹음 하였고, 16bit PCM 양자화를 사용하였다. 실험 어휘는 실내 10명, 실외 5명 등 총 15명의 성인 남성이 참가하였다. 실내 환경은 50~55dB이고, 실외 환경은 70~75dB의 소음환경 하에서 실험하였다.

제안한 시스템의 성능 평가를 위하여 기존 방식과 비교 실험을 하였으며 표 1에 기존의 방식인 Euclidean, DTW, 제안 방법을 대상으로 인식률과 오류 보정률을 실험하여 나타내었다.

표 1. 오류 보정률 비교
Table 1. Comparison of error correction

오류 보정	인식률(%)	보정률(%)
Euclidean	97.31	3.8
DTW	97.51	4.5
제안 방법	98.71	4.7

Euclidean거리 측정 방법과 DTW 그리고 제안 방법에 대해 인식률을 측정한 결과 인식률과 보정률이 비슷하게 나타났지만 제안 방법이 조금 높게 나타난 것을 확인할 수 있었으며 오류 보정률에서도 근소하게 높게 나타난 것을 확인할 수 있었다.

표 2와 3은 기존의 방식인 Euclidean, DTW 그리고 제안한 방법을 실내 환경에서의 실험과 실외 환경에서의 실험을 나타낸다.

표 2. 실내 환경 인식률
Table 2. Indoor Environment Recognition Rate

어휘	인식률(%)		
	Euclidean	DTW	제안 방법
어휘 종속	97.31	97.51	98.71
어휘 독립	96.97	96.91	97.91

표 2는 실내 환경인 50~55dB에서 실험 하였으며 결과에서 보는 것과 같이 시스템 성능 평가 결과 어휘 종속 인식률은 Euclidean과 DTW는 각각 97.31%와 97.51%로 나타났고 제안한 방법은 98.71%로 나타났다. 어휘 독립 인식률에서도 Euclidean과 DTW는 각각 96.97%와 96.91%로 나타났고 제안한 방법은 97.91%로 나타났다.

표 3. 실외 환경 인식률
Table 3. Outdoor Environment Recognition Rate

어휘	인식률(%)		
	Euclidean	DTW	제안 방법
어휘 종속	95.11	95.18	95.31
어휘 독립	93.67	93.58	93.88

표 3은 실외 환경은 70~75dB의 소음환경 하에서 실험하였으며 결과에서 보는 것과 같이 어휘 종속 인식률은 Euclidean과 DTW는 각각 95.11%와 95.18%로 나타났고 제안한 방법은 95.31%로 나타났다. 어휘 독립 인식률에서도 Euclidean과 DTW는 각각 93.67%와 93.58%로 나타났고 제안한 방법은 93.88%로 나타났다. 실험 결과 실내•외 환경 인식률에서 비교적 비슷한 성능을 보이고 있지만 제안 방법이 1.2%의 향상된 인식률을 보이고 있다.

V. 결론

본 논문은 음소가 갖는 특징을 기반으로 바타챠랴 거리 측정법을 이용한 음소 유사율 오류 보정 개선 시스템을 제안하여 실험하였다.

어휘 인식 시스템에서 인식률 저하는 부정확한 어휘 제공과 유사한 음소로 인식되어 오인식되는 오류가 존재하여 부정확한 어휘의 입력으로 특징을 추출하여 인식할 경우 오인식의 결과가 나타나거나 유사한 음소로 인식되며 특징 추출이 제대로 이루어지지 않으면 음소 인식 시 유사한 음소로 인식하게 된다.

바타챠랴 거리 측정의 오류율 측정에 기반을 두고 비슷한 가중치를 갖는 가우시안의 조합이 쉽고 외부의 가우시안이 흡수되지 않는 장점을 활용함으로써 모델의 군집화를 원활화 할 수 있었으며 벡터 값과 오류율을 조절함으로 원하는 결과를 얻을 수 있었다.

모노폰으로 훈련시킨 훈련 데이터의 음소에 HMM 특징 추출 방법을 이용하여 구하였다. 유사한 음소는 바타챠랴 거리 측정법을 이용하여 정확한 음소로 인식할 수 있도록 유도하였으며 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다. 음소 유사율과 신뢰도를 이용하여 오류 보정률을 구하였으며 문법적인 어절은 인식시키고 비문법적인 어절은 오류 보정을 수행하여 재인식시킨다. 의미적으로 분석하기 힘든 핵심어로만으로 이루어진 문장을 복원하여 전체적으로 인식할 수 있도록 해결하였으며 어휘 인식 과정에서 오류로 판명된 어휘에 대하여 오류 보정을 수행하였다.

시스템 성능 평가 유사한 음소는 바타챠랴 거리 측정법을 이용하여 정확한 음소로 유도하여 인식시켰고 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다. 어휘 인식 과정에서 오류로 판명된 어휘에 대하여 오류 보정을 수행하여 재인식을 수행하였다. 이를 유클리디안 거리 측정법과 동적 타임 워핑 시스템에 비교한 시스템 성능 평가 결과 1.2%의 인식 향상률을 보였다.

참고문헌

[1] Eiichi Tanaka and Tamotsu Kasai, “Synchronization and Substitution Error-correcting codes for the Levenshtein Metric,” IEEE Trans. Information Theory, Vol. IT-22, No. 2, pp.156-176, 1976.

[2] E. K. Ringer and J. F. Allen, “A fertility channel model for post-correction of continuous speech recognition,” Proc. ICSLP, pp.897-900, Oct, 1996.

[3] 박미성, 김미진, 김계성, 최재혁, 이상조, “연속 음성인식 후처리를 위한 음절 복원 rule-based 시스템과 형태소분석기법의 적용,” 대한전자공학회논문지, 제 36권, 제 3호, 47-57쪽, 1999년 3월.

[4] 문광식, 김회린, 정재호, 이영직, “가변어휘 단어 인식에서의 미등록어 거절 알고리즘의 성능비교,” 신호처리합동학술대회논문집, 제 12권, 제 1호, 305-308쪽, 1999년 10월.

[5] 조시원, 이동욱, “음성 인식 후처리를 위한 연속 음절 문장의 키워드 추출 알고리즘,” 대한전기학회, 학술대회논문집, 심포지엄 논문집 정보 및 제어부문, 170-171쪽, 2008년 4월.

[6] 안찬식, 오상엽, “MLHF 모델을 적용한 어휘 인식 탐색 최적화 시스템,” 한국컴퓨터정보학회지, 제 14권, 제 10호, 217-223쪽, 2009년 10월.

[7] L. Rabiner and B. H. Juang, "Fundamentals of Speech Recognition", Prentice-Hall, 1993.

[8] T. Jitsuhiro, S. Takatoshi, and K. Aikawa, "Rejection of out-of-vocabulary works using phoneme confidence likelihood," Proc. ICSSP, pp.217-220, May, 1998.
[9] 한동조, 최기호, "음성인식 후처리에서 음소 유사율을 이용한 오류보정에 관한 연구," 한국ITS학회논문지, 제 6권, 제 3호, 77-86쪽, 2007년 12월.
[10] 김용현, 정민화, "에러패턴 학습과 후처리 모듈을 이용한 연속 음성 인식의 성능향상," Proc. KISS Spring Semiannual Conf. 제 27권, 제 1호, 441-443쪽, 2000년 4월.
[11] 방기덕, 강철호, "가변 신뢰도 문턱치를 사용한 미등록어 거절 알고리즘에 대한 연구," 한국멀티미디어학회논문지, 제 11권, 제 11호, 1471-1479쪽, 2008년 11월.
[12] 송원문, 김명원, "문맥 및 사용 패턴 정보를 이용한 음성인식 후처리," 정보처리학회논문지, 제 13-B권, 제 5호, 553-560쪽, 2006년.
[13] 김동주, 김한우, "문맥가중치가 반영된 문장 유사도 척도," 대한전자공학회논문지, 제 43권, 제 6호, 496-504쪽, 2006년.
[14] M. Ostendorf, "From HMM's to segment models: a unified view of stochastic modeling for speech recognition," Speech and Audio Processing, IEEE, Vol. 4, pp.360-378, 1996.
[15] S. Young, D. Kershaw, J. Odell, D. Ollason, Valtcher, P. Woodland, "The HTK Book," Cambridge University Engineering Department, 2002.
[16] M. F. Gales, "Model-based techniques for nosie robust speech recognition," Ph. D. dissertation, University of Cambridge, Sept, 1995.
[17] 안찬식, 오상엽, "공유모델 인식 성능 향상을 위한 효율적인 연속 어휘 군집화 모델링," 한국컴퓨터정보학회지, 제 15권, 제 1호, 177-183쪽, 2010년 1월.
[18] Kris Demuynck, Tom Laureys, Dirk van Compernolle, and Hugo van Hamme, "FLaVor:a flexible architecture for LVCSR," In EUROSPEECH-2003, pp.1973-1976, 2003.

저 자 소 개

안 찬 식

2002 : 광운대학교
컴퓨터공학과 공학석사.
2004 : 광운대학교
컴퓨터공학과 박사수료.
관심분야 : 음성인식, 분산처리, 음성/음향 신호처리

오 상 엽

1999 : 광운대학교
전자계산학과 이학박사.
현 재 : 경원대학교 IT대학
컴퓨터소프트웨어 교수
관심분야 : 소프트웨어공학, 버전관리, 소프트웨어재사용, 형상관리, 객체지향, 음성인식, 분산처리, 음성/음향 신호처리