

음소 유사율 오류 보정을 이용한 어휘 인식 후처리 시스템

안찬식*, 오상엽**

Vocabulary Recognition Post-Processing System using Phoneme Similarity Error Correction

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

요약

어휘 인식 시스템에서 인식을 저하의 요인으로는 유사한 음소 인식과 부정확한 어휘 제공으로 인해 오인식 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 음소 유사율을 이용한 어휘 인식 후처리에서의 오류 보정 후처리 시스템을 제안하였다. 음소 유사율은 모노폰으로 훈련시킨 훈련 데이터를 각각의 음소에 MFCC와 LPC 특징 추출 방법을 이용하여 구하였다. 유사한 음소는 정확한 음소로 인식할 수 있도록 유도하여 부정확한 어휘 제공으로 인하여 오인식되는 오류를 최소화하였다. 음소 유사율과 신뢰도를 이용하여 오류 보정율을 구하였으며, 어휘 인식 과정에서 오류로 판명된 어휘에 대하여 오류 보정을 수행하였다. 예러 패턴 학습을 이용한 시스템과 의미기반을 이용한 시스템에 비해 시스템 성능 평가 결과 MFCC와 LPC는 각각 7.5%와 5.3%의 인식 향상률을 보였다.

Abstract

In vocabulary recognition system has reduce recognition rate unrecognized error cause of similar phoneme recognition and due to provided inaccurate vocabulary. Input of inaccurate vocabulary by feature extraction case of recognition by appear result of unrecognized or similar phoneme recognized. Also can't feature extraction properly when phoneme recognition is similar phoneme recognition. In this paper propose vocabulary recognition post-process error correction system using phoneme likelihood based on phoneme feature. Phoneme likelihood is monophone training phoneme data by find out using MFCC and LPC feature extraction method. Similar phoneme is induced able to recognition of accurate phoneme due to inaccurate vocabulary provided unrecognized reduced error rate. Find out error correction using phoneme likelihood and confidence when vocabulary recognition perform error correction for error proved vocabulary. System performance comparison as a result of recognition improve represent MFCC 7.5%, LPC 5.3% by system using error pattern and system using semantic.

• 제1저자 : 안찬식 교신저자 : 오상엽

• 투고일 : 2010. 02. 08, 심사일 : 2010. 04. 26, 게재확정일 : 2010. 05. 04.

* 광운대학교 컴퓨터공학과 박사과정 ** 경원대학교 IT대학 컴퓨터소프트웨어

※ 본 연구는 2010년도 경원대학교 지원에 의한 결과임.

▶ Keyword : 오류 보정(error correction), LPC(Linear Predictive Coefficient)
MFCC(Mel Frequency Cepstrum Coefficient), 음소 유사율(phoneme likelihood)

I. 서론

모바일 어휘 인식 기술이 발달함에 따라 어휘 기반 검색 시스템, 자동 응답 시스템 등 어휘 인식을 인터페이스로 하는 시스템들이 개발되고 있다. 운전 중인 차량 환경의 경우, 손의 사용이 제한되기 때문에 어휘를 입력으로 하는 텔레매틱스 단말기가 필요하지만 어휘 인식에는 여전히 유사한 음소와 부정확한 어휘 제공에서 오류가 존재한다. 이를 위한 신호 처리 단계에서의 어휘 인식 오류 보정에 대한 여러 가지 연구가 진행되어 왔다[1]. 하지만 사용범위가 넓고, 화자 독립적인 최근의 시스템에서 어휘 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 방법에서는 어휘 인식기의 적용 환경과 실제 인식할 때의 조건상의 차이가 있다는 점을 전제로 하여, 오류 보정을 수행하는 잡음 채널 모델 기반의 오류 보정 방법이 있다. 이 방법은 단순한 언어 모델이 가지는 한계점을 극복하지 못한다는 단점이 있다[3]. 이는 인식 과정에서의 오류는 일정한 패턴을 가지고 발생한다는 점을 전제로 하여 발화 문장과 인식 문장을 비교하여 오류 패턴을 학습하고 후처리 모듈에서 보정하는 방법으로 적은 비용과 시간으로 오류를 보정할 수 있지만, 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 오인식될 경우 적용이 불가능한 단점이 있다[5].

따라서 본 논문에서는 음소가 가지는 어휘 특성을 고려한 어휘 인식 후처리에서 오류 보정 시스템을 제안한다. 어휘 인식 후처리에서 오류 보정을 위한 본 논문의 접근 방법은 인식된 음소열을 형태소 분석 과정을 통해 음운 변동이 적용되기 이전의 문자열로 복원한다. 부정확한 어휘의 입력으로부터 특징을 추출하여 인식할 경우 유사한 음소로 인식하거나 오인식 오류로 나타나게 되므로 입력되어진 어휘를 형태소 분석 과정을 통해 유사 음소 인식과 오인식 오류를 최소화하여 인식률을 향상시켰다. 형태소 분석은 문법적 어절과 비문법적 어절의 필터링하여 문법적 어절은 인식시키고 비문법적 어절은 오

류 보정을 수행하여 재인식시킨다. 의미적으로 분석하기 힘든 핵심어로만 이루어진 문장을 복원하여 전체적으로 인식할 수 있도록 해결하였다.

본 논문의 구성은 2장에서 어휘 인식기에 어휘 인식 탐색 시스템과 어휘 오류 보정에 대해서 살펴보고, 3장에서는 신뢰도와 어휘 특징을 사용한 음소 유사율 처리 시스템을 사용하여 오류 보정을 수행하는 방법에 대하여 설명한다. 4장에서는 실험 환경과 실험 결과, 5장에서는 결론을 다룬다.

II. 기존 연구

2.1 어휘 인식 탐색 시스템

어휘 인식에는 음향학적 지식을 이용하는 방법, 통계적 방법, 인공 지능을 이용한 방법, 신경 회로망을 이용한 방법 등이 널리 연구되었다. 통계적 어휘 인식에서는 다음 그림 1에서와 같이 학습 단계에서 추출된 특징 벡터를 이용하여 기준이 되는 시간적으로 변화하는 음성 신호의 특징을 모델링한 음향 모델 또는 기준 패턴을 구한 다음, 어휘 인식 단계에서 부분 최적인 각 상태를 통하여 최상의 단일한 상태열을 찾는 비터비 알고리즘 또는 상태열의 길이가 일정하지 않은 두 열의 유사도를 측정하는 DTW(Dynamic Time Warping)와 같은 패턴 정합 알고리즘을 이용하여 미리 학습된 패턴 중에서 가장 유사한 것을 찾아서 인식 결과로 출력한다. 일체형 인식 구조는 탐색 과정에서 모든 가능한 지식 정보들을 가져오므로 일체형 탐색 방법은 복잡한 단어간 문법을 고려한 언어 모델과 시간적으로 변화하는 음성 신호의 특징을 모델링한 음향 모델을 사용한다[6].

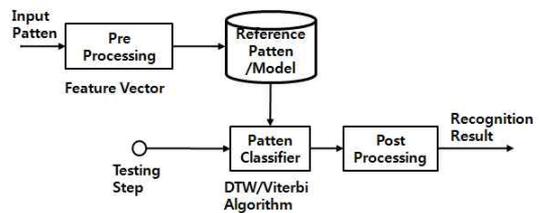


그림 1. 어휘 인식 탐색 시스템
Fig. 1. Vocabulary recognition search system

일체형 어휘 인식의 장점은 검색의 효율성에 있다. 어휘에서 음향적 혼란이 많아 어휘 및 언어 모델에 의해 제공되는 정보를 일찍 포함하는 것은 탐색공간으로부터 가장 가능성이 낮은 부분들을 삭제하기 위해 효과적이다.

2.2 어휘 오류 보정

어휘 오류 보정은 전처리에서의 잡음 제거 방법과 후처리에서의 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. 특성 추출은 인식에 유용한 성분을 어휘 신호로부터 뽑아내는 과정이다.

멜 필터뱅크 분석은 인간의 청각은 주파수 영역에서 음성을 비선형적으로 분석하는 특징을 갖는다. 그러나 1kHz 이상에서는 로그 주파수(logarithmic frequency) 영역에서 거의 선형적으로 분석한다. 또한, 임계 대역(critical band)은 1kHz 이하에서는 로그 주파수 영역에서 거의 일정하고, 1kHz 이상에서는 선형적으로 증가한다. 이러한 근거를 통하여 1kHz의 선형 주파수가 1,000멜(mel)에 대응되도록 정의하고 분석한다[7].

프리엠퍼시스란 저주파는 상대적으로 줄이고 고주파를 상대적으로 키우는 것으로 어휘 스펙트럼이 기본적으로 낮은 주파수에서 에너지가 많고 높은 주파수에서 에너지가 적은 특성을 가지고 있기 때문에 프리엠퍼시스를 통해서 스펙트럼의 경사를 평탄화한다. 프리엠퍼시스 필터를 사용하는 이유는 인간의 외이와 중이의 주파수 특성을 모델링하기 위하여 고대역 필터링을 한다. 이는 입술에서의 방사에 의하여 20dB로 감쇄되는 것을 보상하게 되어 음성으로부터 성도 특성만을 얻게된다. 청각시스템이 1kHz이상의 스펙트럼 영역에 대하여 민감한 것을 어느 정도 보상하기 위해서 사용한다.

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

그림 2는 어휘 오류 보정에 구성도를 나타내고 있으며, 인식된 음소열을 음절 복원을 통해 음운 변동이 적용되기 이전의 문자열로 복원시킨다.

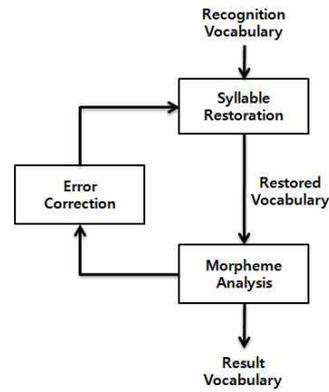


그림 2 어휘 오류 보정
Fig. 2. Vocabulary error correction

복원된 어절의 형태소 분석을 통해 문법적 어절과 비문법적 어절의 필터링을 하고 비문법적 어절에 대해 오류보정을 수행한다. 음절 복원기와 형태소 분석 단계를 다시 거치게 된다[8].

2.3 음소 특성 추출

특성 추출은 인식에 유용한 성분을 신호로부터 뽑아내는 과정이다. 특성 추출은 일반적으로 정보의 압축, 차원의 감소 과정과 관련되며 특성의 좋고 나쁨은 인식률로 판단된다. 흔히 사용되는 특성 추출 과정에서 청각 특성을 반영한 것으로 달팽이관의 주파수 응답을 필터 뱅크 분석, 주파수에 따른 대역폭의 증가, 프리엠퍼시스 필터 등이 사용된다. 어휘 신호의 동적 특성을 반영하기 위하여 캡스트럼 1차, 2차 미분값을 사용한다.

미분은 시간축 방향의 필터링으로 생각할 수 있으며 시간축 방향으로의 특징 벡터를 얻는 과정이다. 인식을 위하여 주로 사용되는 특징은 MFCC(Mel Frequency Cepstrum Co-efficient)와 LPC(Linear Predictive Coefficient)가 주로 사용된다[9].

MFCC는 신호를 안티 앨리어싱 필터(anti-aliasing filter)로 거친 다음, A/D 변환을 거쳐서 디지털 신호 $x(n)$ 로 변환한다. 디지털 신호는 고대역 통과 특성을 갖는 디지털 프리엠퍼시스 필터를 거친다. 프리엠퍼시스된 신호는 해밍 윈도우(hamming window)를 씌워서 블록 단위의 프레임으로 나누어지고 프레임 단위로 만들어진다. 프레임의 크기는 보통 20-30ms이며 프레임 이동은 10ms가 사용한다. 한 프레임의 신호는 FFT(Fast Fourier Transform)를 이용하여 주파수 영역으로 변환하여 주파수 대역을 여러 개의 필터뱅크로 나누고 각 बैं크에서의 에너지를 구한다. 밴드 에너지에 로그를 취한 후 DCT(Discrete Cosine Transform)를 취하면 최종적인

MFCC가 얻어진다. MFCC 계수는 12개를 사용하며 이와는 별도로 구한 프레임 로그 에너지가 추가적으로 사용되어 인식의 입력으로 사용되는 특성 벡터는 13차 벡터로 구성되어 사용된다[10].

LPC는 신호의 스펙트럼 및 FFT 켈스트럼으로 얻은 포락을 나타낸다. 선형 분석법 자체가 스펙트럼의 폴(Pole)만을 고려하는 올 폴(all-pole) 모델링에 기반하므로 LPC 켈스트럼으로 얻은 스펙트럼 포락을 사용한다[11].

III. 음소 유사율 오류 보정 시스템

오류 보정 방법은 단순한 언어 모델이 가지는 한계점을 극복하지 못한다는 단점과 오류 패턴 DB가 필요하다. 문장이 간결하고 사용자가 검색하고자 하는 핵심어만 이루어진 경우가 많으므로 의미적으로 분석하기 힘들다. 이러한 단점을 보완하기 위해 음소가 가지는 어휘 특성을 고려한 어휘 인식 후처리에서 오류보정 시스템 모델을 그림 3과 같이 구성하였다.

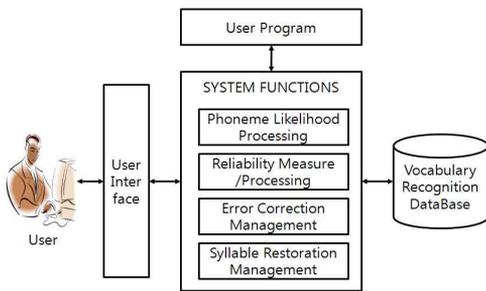


그림 3. 시스템 모델
Fig. 3. System Model

입력되어진 음소를 음소 유사율 처리를 통해 음소 사이의 거리를 측정하여 수치로 나타내고 신뢰도 측정을 통하여 인식되어진 결과를 확인한다. 인식 결과로부터 오류로 판명되어진 음소를 확인하여 오인식되어진 음소열은 형태소 분석을 통해 문법적 어절과 비문법적 어절의 필터링하여 문법적 어절은 인식시키고 비문법적 어절은 오류 보정을 수행하여 오류 복원율을 확인하여 관리하여 재인식을 수행한다.

음소 유사율 처리는 음소 사이의 거리를 측정하여 비슷한 음소를 찾아 관리하며 신뢰도 측정 및 처리에서는 음소의 신뢰도를 측정 및 처리한다. 오류 보정을 관리하는 유사율과 신뢰도를 비교하여 오류 보정을 위한 음소와 음절을 관리하며 음절 복원 관리에서는 오류 보정된 음소와 음절을 복원하여 후보 음절을 생성한다.

3.1 음소 유사율 처리

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

음소 유사율 처리 알고리즘은 다음과 같다.

```

int SearchTree::ExpandRCTriphone(int frameX)
{
    PhoneNodeParm phoneNodeParm;
    SetPhoneNodeParm(&phoneNodeParm);

    float thresh = m_maxLike -
                  m_searchParm.m_phoneBeamWidth;
    int n=0;

    for(int i=0;i<m_currActiveAP->size();i++){
        PhoneNode* node=(*m_currActiveAP)[i];
        if (!node->IsRootRightContext()) continue
        if (node->FinalLike()<=thresh) continue
        n += node->ExpandRCTriphone(frameX,
                                   m_rightContextBaseA, &phoneNodeParm);
    }
    return n;
}
    
```

식 (1)은 두 음소 집합에 대해서 정의를 한다.

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

$P(X|w_i)$ 는 음소 집합 $w_i (i = 1, 2)$ 의 확률 밀도 함수이고, Ω 는 확률 분포상에서 정의되는 랜덤 함수 X 의 영역이다. 음소 집합의 확률 분포를 정규 분포로 가정할 경우 식 (2)와 같이 나타낼 수 있다.

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

u_1 와 Σ_i 는 각각 음소 집합 w_i 의 평균 벡터와 공분산 행렬이다. 식 (3)에 의해 음소 유사율을 측정하였고 식 (4)에 의해 [0, 1]사이에 분포하도록 정규화하였다.

$$d(i, j) = \left[\frac{1}{n} \sum_{k=1}^n \frac{(\mu_{ik} - \mu_{jk})^2}{\sigma_{ik} \sigma_{jk}} \right]^{\frac{1}{2}} \dots\dots\dots (3)$$

$$S_p = \frac{b_{\max} - b_{xy}}{b_{\max} - b_{\min}} \dots\dots\dots (4)$$

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 음소 일수록 1에 가깝고, 유사하지 않을수록 0에 가깝게 나타낸다.

3.2 신뢰도 측정 및 처리

인식된 결과는 음소나 단어로부터 발화되었을 확률에 대한 상대값을 의미하며 신뢰도는 인식 결과에 대해 그 결과가 얼마나 믿을 만한 것인가를 나타내는 척도이다[13].

어떤 O 를 관측 세그먼트라고 하면 인식 과정에서 O 가 입력되었을 때는 두 가지의 가정이 가능하다. O 가 실제 세그먼트 k 일 것이라는 가정이 가능하며 영가설이라 하고 H_0 로 표시한다. 또한 O 가 실제 세그먼트가 아닌 다른 유사 발화라 가정이 가능하며 대립가설이라 하고 H_1 으로 표현한다. 주어진 테스트 세그먼트 O 에 대해 발화 검증 과정은 영가설에 대한 확률과 대립가설에 대한 확률을 비교하여 영가설에 대한 확률이 크면 인식하고 아니면 잘 못된 인식으로 판단한다. 따라서 식 (5)가 만들어 진다.

$$P(O|H_0) > P(O|H_1) \dots\dots\dots (5)$$

위의 식 (5)를 베이시안 룰(bayes rule)에 의해 정리하면 식 (6)과 식 (7)로 나타나고 이를 신뢰도 측정을 위해 사용한다.

$$P(H_0|O)P(H_0) > P(H_1|O)P(H_1) \dots\dots\dots (6)$$

$$\frac{P(H_0|O)}{P(H_1|O)} > \frac{P(H_1)}{P(H_0)} \dots\dots\dots (7)$$

$P(H_0|O)$ 는 모델 λ_k 에서 O 가 관측될 확률이고, $P(H_1|O)$ 는 다른 모델에서 O 가 관측될 확률이다. H_1 을 모델링하기 위해서 각 음소마다 유사한 음소들을 구하여 파라미터로 훈련하고 훈련된 파라미터를 λ_k 로 표현하고 이를 안티 모델이라 한다. 안티모델을 log를 취해서 우도(log-likelihood)로 사용하고 식(8)로 표현한다.

$$LLR_k(O, \lambda_k) = \log P(O|\lambda_k) - \log P(O|\lambda_{\bar{k}}) \dots\dots\dots (8)$$

우도 값이 너무 큰 범위에서 나타나지 않도록 정규화하며 시그모이드 함수를 사용하고 최종적인 음소 신뢰도는 식 (9)에 의해서 계산된다.

$$f(LLR) = \log \frac{1}{1 + \exp(-a \cdot LLR)} \dots\dots\dots (9)$$

3.3 오류 보정율 관리

오류 보정은 신뢰도에만 의존하지 않고 신뢰도와 음소 유사율을 모두 고려하여야 한다[14]. 음소의 신뢰도 높아도 음소 유사율이 높은 비슷한 음소로 대체되어 오류가 발생할 수 있으며, 신뢰도가 낮아도 제대로 인식될 수 있기 때문이다 [15]. 신뢰도가 낮으며 음소 유사율이 높은 음소를 가지고 있는 음소부터 오류 보정이 수행되어야 정확한 오류 보정을 수행할 수 있다. 식 (10)은 하나의 음절에 대한 오류 보정율을 나타낸다.

$$R_{\text{syllable}} = \frac{1}{n} \sum_{k=1}^n (1 - \alpha_k) b_k \dots\dots\dots (10)$$

n 은 한 음절의 음소 수, α_k 는 k 번째 음소의 신뢰도, b_k 는 k 번째 음소에 대한 유사율을 나타낸다. $\alpha = 1$ 일 때, 오류 보정율은 0의 값을 가지며, $\alpha \neq 1$ 일 때는 $0 < R_{\text{syllable}} < 1$ 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 보정율은 작아진다.

후보 음절을 선정한 후 우선 순위 어절을 선정한다. 우선 순위 어절은 오류 보정율의 평균으로 구하며 음절 복원과 형태소 분석을 재 수행한다. 신뢰도가 높은 음절의 경우 다음 후보 음절과의 오류 보정율 차이는 작으며, 신뢰도가 낮은 음절의 경우 다음 후보 음절과의 오류 보정율 차이가 크게 나타나므로 우선순위 어절을 선정할 때 올바른 어절이 낮은 순위로 나타나게 된다. 따라서 음절의 신뢰도에 따른 임계값을 식 (11)과 같이 주어 적절한 오류 보정율을 갖게 한다.

$$R_{\text{syllable}} = \lambda \frac{1}{n} \sum_{k=1}^n (1 - \alpha_k) b_k, \quad \lambda = \sum_{k=1}^n \alpha_k \dots\dots\dots (11)$$

3.4 음절 복원 관리

음절 복원을 수행하기 위하여 어절을 구성하고 있는 음소의 신뢰도와 음소에 대한 타 음소와의 유사율을 사용하여 각각의 음소에 대한 후보 음절을 생성한다. 생성된 후보 음절을 통하여 우선순위 어절을 선정한다.

인식 오류의 경우 상위 2-3위에 정확한 인식 결과를 포함하고 있으므로 후보 음절을 생성할 때에는 초성 자음과 모음에 대해서 상위 3위의 음소 유사율을 사용하며 종성자음에 대해서는 상위 2위의 음소 유사율을 사용한다.

음절 복원 알고리즘은 다음과 같다.

```

void Activate
{
if(HasModel()){
m_rootP->Activate(frameX,like,bpw,bpwPhone);
if(!IsActive(frameX))
nextActiveAP->push_back(this);
}
else if(ignoreNullNode && IsNull()){
for(list<WordLink>::iterator i=m_link.begin();
i!=m_link.end(); ++i){
(*i).m_wordNodeP->Activate(frameX,like,
bpw,bpwPhone,nextActiveAP);
}
}
if(like>m_maxLike){
m_maxLike=like;
m_lastFrameX=frameX;
if(!HasModel() || HasSubnet()){
m_finalLike=m_maxLike;
m_finalPath=bpw;
m_lastBpw=bpw;
m_lastBpwPhone=bpwPhone;
}
}
}
}
    
```

IV. 실험결과 및 분석

본 논문에서 제안한 음소가 가지는 어휘 특성을 고려한 어휘 인식 후처리에서 오류보정 시스템 모델을 구성하여 인식 실험을 수행하였다.

본 실험에서는 단말기에서 사용되는 단어를 선정하여 총 20개를 표 1과 같이 선별하였고 후처리 실험을 위해 5명의 사용자가 100회의 데이터를 구축하였다.

어휘는 실내 환경과 잡음 환경에서 이동기기 등에 내장되어 있는 내장형 마이크폰을 사용하여 16kHz Mono로 녹음

을 하였고, 16bit PCM 양자화를 사용하였다. 실험 어휘는 실내 10명, 실외 5명 등 총 15명의 성인 남성이 참가하였다.

녹음된 데이터는 인식기 학습을 위해 MFCC 특성 추출 방법을 사용하였고 인식기는 SITEC에서 개발한 ECHOS[16]를 이용하였다.

HMM 기반 ECHOS는 각 단어별 데이터로 학습된 인식 모델을 이용하여 발화된 단어의 인식 가능한 단어들의 인식 가능도를 우도로 표현하고 최대값을 가지는 단어를 최종 결과로 선정한다.

표 1. 실험 단어
Table 1. Experimentation word

No.	단어	No.	단어	No.	단어	No.	단어
1	위	2	아래	3	시작	4	다음
5	종료	6	상위	7	하위	8	정지
9	크게	10	작게	11	확대	12	축소
13	취소	14	예	15	찾기	16	도착지
17	출발지	18	경로	19	확인	20	아니오

음소 유사율의 구성에 따른 음소 유사율의 정확성을 확인하기 위하여 MFCC와 LPC의 특성 추출 방법을 사용하여 음소 유사율을 구성하고 어휘 인식 후처리에서 오류보정 시스템 모델에 적용하였다.

표 2는 기존의 에러 패턴 학습을 이용한 방법[17][18]인 error pattern과 의미기반의 방법[19]인 semantic 그리고 본 논문의 제안 방법인 MFCC, LPC의 결과를 나타내었다.

표 2. 오류 보정을 비교
Table 2. Comparison of error correction

오류 보정	인식률(%)	보정율(%)
error pattern	76.2	3.1
semantic	79.4	7.5
MFCC	86.5	7.5
LPC	82.9	5.3

에러 패턴 학습을 이용한 오류 보정의 경우 3.1%, 의미기반의 오류 보정의 경우 7.5%의 인식 향상률을 보였으며, 본 논문에서 제안한 MFCC, LPC를 이용한 음소 유사율을 사용할 경우 각각 7.5%, 5.3%의 인식률 향상을 보였다. 또한 MFCC가 LPC보다 우수한 성능을 보였으며, 전체적으로 86.5%, 82.9%의 인식률을 보였다.

V. 결론

본 논문에서는 음소가 가지는 어휘 특성을 고려한 어휘 인식 후처리에서 오류 보정 시스템을 제안하였다.

어휘 인식 후처리에서 오류 보정을 위한 방법은 인식된 음소열을 오인식된 음소열을 관리하여 형태소 분석 과정을 통해 음운변동이 적용되기 이전의 문자열로 복원한다. 부정확한 어휘의 입력으로부터 특징을 추출하여 인식된 유사한 음소와 오인식 오류로 나타난 어휘를 형태소 분석인 문법적 어절과 비문법적 어절의 필터링하여 문법적 어절은 인식시키고 비문법적 어절은 오류 보정을 수행하여 재인식하여 오류를 최소화하고 인식률을 향상시켰다. 의미적으로 분석하기 힘든 핵심어들은 문장을 복원하여 전체적으로 인식할 수 있도록 해결하였다.

제안한 어휘 인식 후처리에서 오류 보정 시스템으로 인하여 인식률을 향상 시킬 수 있는 장점을 확인하였으며 검색 시속도와 인식률에서 기존 시스템인 에러 패턴 학습을 이용한 시스템과 의미기반을 이용한 시스템보다 나은 결과를 얻을 수 있었다. 시스템 성능 평가 결과 MFCC와 LPC는 각각 7.5%와 5.3%의 인식 향상률을 보였다.

참고문헌

- [1] 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.
- [2] 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.
- [3] 조시원, 이동욱, "음성 인식 후처리를 위한 연속 음절 문장의 키워드 추출 알고리즘," 대한전기학회, 학술대회논문집, 심포지엄 논문집 정보 및 제어부문, 170-171쪽, 2008년 4월.
- [4] 문광식, 김회린, 정재호, 이영직, "가변어휘 단어 인식에서의 미등록어 거절 알고리즘의 성능비교," 신호처리합동 학술대회논문집, 제 12권, 제 1호, 305-308쪽, 1999년 10월.
- [5] 박미성, 김미진, 김계성, 최재혁, 이상조, "연속 음성인식 후처리를 위한 음절 복원 rule-based 시스템과 형태소분석기법의 적용," 대한전자공학학회논문지, 제 36권, 제 3호, 47-57쪽, 1999년 3월.
- [6] 안찬식, 오상엽, "공유모델 인식 성능 향상을 위한 효율적인 연속 어휘 군집화 모델링," 한국컴퓨터정보학회지, 제 15권, 제 1호, 177-183쪽, 2010년 1월.
- [7] T. Jitsuhiro, S. Takatoshi, and K. Aikawa, "Rejection of out-of-vocabulary words using phoneme confidence likelihood," Proc. ICSSP, pp.217-220, May 1998.
- [8] 한동조, 최기호, "음성인식 후처리에서 음소 유사율을 이용한 오류보정에 관한 연구," 한국ITS학회논문지, 제 6권, 제 3호, 77-86쪽, 2007년 12월.
- [9] 안찬식, 오상엽, "MLHF 모델을 적용한 어휘 인식 탐색 최적화 시스템," 한국컴퓨터정보학회지, 제 14권, 제 10호, 217-223쪽, 2009년 10월.
- [10] 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.
- [11] S. Young, D. Kershaw, J. Odell, D. Ollason, Valtcher, P. Woodland, "The HTK Book," Cambridge University Engineering Department, 2002.
- [12] 김동주, 김한우, "문맥가중치가 반영된 문장 유사도 척도," 대한전자공학학회논문지, 제 43권, 제 6호, 496-504쪽, 2006년.
- [13] 송원문, 김명원, "문맥 및 사용 패턴 정보를 이용한 음성 인식 후처리," 정보처리학회논문지, 제 13-B권, 제 5호, 553-560쪽, 2006년.
- [14] 김영진, 김은주, 김명원, "모바일 기기를 위한 음성인식의 사용자 적용형 후처리," 한국정보과학학회논문지, 컴퓨팅의 실제 및 레터, 제 13권, 제 5호, 338-342쪽, 2007년 10월.
- [15] 김원규, "음성 인식을 위한 후처리에 관한 연구," 한국퍼지및지능시스템학회, 학술대회논문집, 421-424쪽, 2008년 4월.
- [16] 음성정보기술산업지원센터, "한국어 음성인식 플랫폼 사용자 매뉴얼(ECHOS Manual)," 135-308쪽, 2006년.
- [17] S. Kaki, E. Sumita, and H. Iida, "A method for correction speech recognition using the statistical features of character co-occurrence," Proc. COLING-ACL, pp.653-657, Aug, 1998.

- [18] 김용현, 정민화, “에러패턴 학습과 후처리 모듈을 이용한 연속 음성 인식의 성능향상,” Proc. KISS Spring Semiannual Conf. 제 27권, 제 1호, 441-443쪽, 2000년 4월.
- [19] 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.

저자 소개



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 컴퓨터소
프트웨어 교수
관심분야 : 소프트웨어공학, 버전관리,
소프트웨어재사용, 형상관리,
객체지향, 음성인식, 분산
처리, 음성/음향 신호처리