

음성인식 후처리에서 음소 유사율을 이용한 오류보정에 관한 연구*

A Study on Error Correction Using Phoneme Similarity in Post-Processing of Speech Recognition

한동조[†]
(Dong-Jo Han)

최기호^{**}
(Ki-Ho Choi)

요 약

최근 텔레매틱스 단말기 등과 같이 음성인식을 인터페이스로 하는 음성기반 검색시스템들이 많이 개발되고 있다. 그러나 음성인식에는 여전히 많은 오류가 존재하며, 이에 오류보정에 대한 여러 가지 연구가 진행되고 있다. 본 논문에서는 한국어의 음소가 갖는 특징을 기반으로 음성인식 후처리에서의 오류보정을 제안하였다. 이를 위해 한국어 음소의 특징을 고려한 음소 유사율을 사용하였다. 음소 유사율은 훈련데이터를 모노폰으로 훈련시켜 한국어 음소 각각에 대하여 MFCC와 LPC 특징추출방법을 사용하여 특징추출을 수행하고, 바타차라 거리 측정법을 사용하여 각 음소 사이의 유사율을 구하였다. 음소 유사율과 신뢰도를 이용하여 오류보정률을 구하였으며, 이를 사용하여 음성인식 과정에서 오류로 판명된 어절에 대하여 오류보정을 수행하고, 음절 복원과 형태소 분석을 재수행하는 과정을 거쳤다. 실험 결과 MFCC와 LPC 각각 7.5%와 5.3%의 인식 향상률을 보였다.

ABSTRACT

Recently, systems based on speech recognition interface such as telematics terminals are being developed. However, many errors still exist in speech recognition and then studies about error correction are actively conducting. This paper proposes an error correction in post-processing of the speech recognition based on features of Korean phoneme. To support this algorithm, we used the phoneme similarity considering features of Korean phoneme. The phoneme similarity, which is utilized in this paper, trains data by mono-phoneme, and uses MFCC and LPC to extract feature in each Korean phoneme. In addition, the phoneme similarity uses a Bhattacharyya distance measure to get the similarity between one phoneme and the other. By using the phoneme similarity, the error of eo-jeol that may not be morphologically analyzed could be corrected. Also, the syllable recovery and morphological analysis are performed again. The results of the experiment show the improvement of 7.5% and 5.3% for each of MFCC and LPC.

Key words : Speech recognition, error correction, phoneme similarity, MFCC, LPC, Bhattacharyya distance

본 연구는 2006학년도 광운대학교 교내학술연구비로 진행되었음.

* 주저자 : 광운대학교 컴퓨터공학과 박사과정

** 공저자 : 광운대학교 컴퓨터공학과 교수

† 논문집수일 : 2007년 11월 6일

I. 서 론

최근 음성인식 기술이 발달함에 따라 음성 기반 검색시스템, 자동 응답 시스템 등 음성인식을 인터페이스로 하는 많은 시스템들이 개발되고 있다. 특히 운전 중인 차량환경의 경우, 손의 사용이 제한되기 때문에 음성을 입력으로 하는 텔레매틱스 단말기가 필요하다. 하지만 음성인식에는 여전히 많은 오류가 존재하며, 이는 텔레매틱스 단말기의 사용을 어렵게 한다. 이를 위한 신호처리 단계에서의 음성인식 오류보정에 대한 여러 가지 연구가 진행되어 왔다. 하지만 사용범위가 넓고, 화자 독립적인 최근의 시스템에서 음성신호처리만으로 인식의 효율을 높이는 것은 상당히 어려운 일이다. 따라서 발화된 음성의 단순한 신호처리 위주의 인식 결과로부터 좀 더 신뢰할 수 있는 결과를 얻기 위한 후처리에서 오류보정에 대한 연구가 진행되고 있다.

기존의 방법으로는 음성인식기의 학습시의 환경과 실제 인식할 때의 조건상의 차이가 있다는 점을 전제로 하여, 오류보정을 수행하는 잠음 채널 모델 기반의 오류보정 방법이 있다 [1]. 이 방법은 단순한 언어 모델이 가지는 한계점을 극복하지 못한다는 단점이 있다. 두번째로는 오류 패턴 매칭 기반의 오류보정 방법이 있다 [2, 3]. 이는 인식과정에서의 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 많이 필요하다. 세 번째로는 어휘 의미 패턴 기반의 오류보정 방법은 인식 문장을 의미적으로 분석하여 인식된 단어별로 의미 정보를 포함한 특정 스트링으로 대치하는 방법이다 [4]. 하지만 정보 검색 영역에서 사용되는 문장은 문장이 간결하고, 사용자가 검색하고자 하는 핵심어만 이루어진 경우가 많다. 따라서 정보 검색 영역의 문장은 의미적으로 분석하기 힘들며, 문장이 전체적으로 오인식 될 경우 적용이 불가능한 단점이 있다.

따라서 본 논문에서는 한국어의 음소가 가지는

음성 특징을 고려한 음성인식 후처리에서 오류보정을 제안하였다. 음성인식 후처리에서 오류보정을 위한 본 논문의 접근 방법은 다음과 같다. 인식된 음소열은 음절 복원, 형태소 분석 과정을 통해 음운변동이 적용되기 이전의 문자열로 복원된다. 이때 형태소 분석 단계에서 필터링 된 비문법적 어절에 대하여 오류보정을 수행한다. 오류보정은 한국어의 각 음소와 음소 사이의 유사율을 측정한 음소 유사율과 음성 인식 과정에서 측정된 신뢰도를 사용한다. 오류보정을 수행한 어절에 대하여 다시 음절 복원과 형태소 분석을 통해 최종 인식 결과를 얻게 된다.

본 논문의 구성으로는 2장에서는 본 논문에서 사용하는 음성인식기에 필요한 음절 복원과 형태소 분석에 대해서 살펴보고, 3장에서는 신뢰도와 음성 특징을 사용한 음소 유사율, 그리고 이를 사용하여 오류보정을 수행하는 방법에 대하여 설명한다. 4장에서는 실험 환경과 실험 결과, 5장에서는 결론을 다룬다.

II. 음성 인식기

본 논문에서 사용한 음성인식기는 음소 단위 인식이다. 따라서 한국어의 음운변동이 적용된 인식 음소열에 대하여, 음운 변동이 적용되기 이전의 문자열로 복원하여 주는 음절 복원 과정과 복원된 문자열의 문법적 구성이 올바른지 확인하기 위한 형태소 분석 과정이 필요하다 [5, 6].

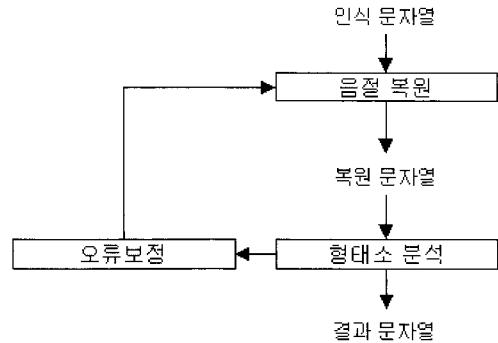
1. 음절 복원

한국어는 하나의 형태소 내부 또는 형태소와 형태소 사이에서 말소리가 바뀌는 음운변동이 일어난다. 따라서 음성인식 과정에서 적용된 음운변동을 역으로 해석하여 원래의 발화 문장으로 바꾸어주는 음절 복원 과정이 필요하다. 다음의 <표 1>은 음절 복원 규칙과 인식 음소열에 대한 복원 문자열에 대한 예를 보여준다.

<표 1> 음절 복원 규칙과 복원 문자열

<Table 1> Syllable recovery rule and recovery string sequence

인식 음소열	복원 규칙	복원 문자열
부사느로	종성(ㅎ) + 초성(ㄴ) → 종성(ㄴ) + 초성(ㅇ)	부산으로
약국	종성(ㄱ) + 초성(ㄱ) → 종성(ㄱ) + 초성(ㄱ)	약국
달꼬	종성(ㄷ) + 초성(ㄷ) → 종성(ㄷ, ㅌ) + 초성(ㄱ, ㅋ)	달고, 닷고
기름	종성(ㄹ) + 초성(ㄹ) → 종성(ㄹ, ㄹᄇ) + 초성(ㅇ, ㅇ)	길음, 깊음
바깥	종성(ㅎ) + 초성(ㅍ) → 종성(ㅍ, ㅍ) + 초성(ㅇ, ㄱ)	밖에, 바깥



<그림 1> 오류보정 흐름도

<Fig. 1> Flow chart of error correction

2. 형태소 분석

형태소 분석이란 실제의 문장에 사용되는 단어의 원래의 구조를 파악하는 것으로 어절을 기본 처리 단위로 한다. 형태소 분석 과정에서는 음절복원 단계에서 출력된 복원 문자열들에 대해서 올바른 형태소들의 집합으로 구성되어 있는지를 분석하고 구조를 파악함으로써 복원 문자열이 문법적으로 올바른 어절인지를 검사한다. 그리고 이를 통해 비문법적 후보 어절을 필터링함으로써 음성인식의 최종 결과 문자열을 출력한다. <표 2>는 복원 문자열에 대한 형태소 분석 결과를 나타낸다.

- NN(명사), JO(조사), EM(어말 어미), AJ(형용사), V(동사), NR(미동복어)

<표 2> 형태소 분석을 사용한 결과 문자열

<Table 2> Result string sequence using morphological analysis

복원 문자열	형태소 분석	결과 문자열
부산으로	부산/NN + 으로/JO	부산으로
달고, 닷고	달/V + 고/EM, 닷고/NR	달고
길음, 깊음	길/NN + 음/JO, 깊/NR + 음/JO	길음
없고, 옆고	없/Aj + 고/EM, 옆/V + 고/EM	없고, 옆고

III. 음소 유사율을 이용한 오류보정

오류보정은 음성인식에서 발생한 오류를 올바른 인식 결과로 보정하는 것으로, 사용범위가 넓고, 화자 독립적인 최근의 음성인식 시스템에서 음성신호 처리만으로 보정하기 힘든 오류를 음성인식 후처리에서 보정하는 연구가 활발히 진행 중이다. 음성인식 후처리에서 오류보정은 인식된 결과에 대하여 오류를 보정한다.

<그림 1>은 음성인식기와 오류보정의 전체적인 구성도이다. 인식된 음소열을 음절 복원을 통해 음운변동이 적용되기 이전의 문자열로 복원시켜준다. 그리고 복원된 어절의 형태소 분석을 통해 문법적 어절과 비문법적 어절을 필터링하고, 비문법적 어절, 즉 형태소 분석이 실패한 어절에 대해 오류보정을 수행한다. 그 후 음절 복원기와 형태소 분석 단계를 다시 거치게 된다.

1. 신뢰도

인식된 결과인 음소나 단어에 대해서, 그 외의 음소나 단어로부터 그 말이 발화되었을 확률에 대한 상대값을 의미하는 신뢰도는 음성인식 결과에 대해서 그 결과가 얼마나 믿을만한 것인가를 나타내는 척도이다 [7-9].

어떤 O 를 실제 음성의 관측 세그먼트(segment)라고 하면 음성인식 과정에서 O 가 입력되었을 때는 크게 두 가지의 가정이 가능하다. 즉, 그 O 가 실제 음성 세그먼트 k 일 것이라는 것이라는 가정이 가능한데 이를 영 가설이라 하고, H_0 로 표시한다. 반면 그 O 가 실제 음성 세그먼트 k 가 아닌 다른 유사 음성에서 발화되었을 것이라 가정할 수 있는데 이를 대립 가설이라 하고 H_1 으로 표현한다. 그러면 주어진 테스트 세그먼트 O 에 대해 발화 검증 과정은 영 가설에 대한 확률과 대립 가설의 확률을 비교하여 영 가설에 대한 확률이 크면 인식하고 아니면 잘못 인식된 핵심어일 가능성이 커진다.

$$P(O|H_0) > P(O|H_1) \quad (1)$$

위 식 (1)을 Bayes rule에 의해 다시 쓰면

$$P(H_0|O)P(H_0) > P(H_1|O)P(H_1) \quad (2)$$

$$\frac{P(H_0|O)}{P(H_1|O)} > \frac{P(H_1)}{P(H_0)} \quad (3)$$

이 된다.

여기서 $P(H_0|O)$ 는 HMM 모델 λ_k 에서 O 가 관측될 확률이고, $P(H_1|O)$ 는 그 외는 다른 모델링에서 O 가 관측될 확률이다. H_1 을 모델링 하기 위해서 각 음소마다 유사한 음소들, 즉 유사음소 집합(cohort set)를 구하여 이를 HMM 파타미터로 훈련하고, 이렇게 훈련된 HMM 파타미터를 안티모델(anti-model)이라고 하고 $\lambda_{\bar{k}}$ 로 표현한다.(각 음소나 단어에 대한 가장 혼돈하기 쉬운 유사한 것을 찾아서 그에 대한 HMM모델을 만든다.) 안티모델 식에서 log를 취하면 우도(log-likelihood)가 되는데, 이를 $LLR_k(O, \lambda_k)$ 또는 줄여서 LLR 로 표현한다.

$$LLR_k(O, \lambda_k) = \log P(O|\lambda_k) - \log P(O|\lambda_{\bar{k}}) \quad (4)$$

그리고 우도 값이 너무 큰 범위에서 나타나지 않

도록 정규화 한다. 이 정규화 함수로 시그모이드(log 값의 범위를 한정)함수를 사용하고, 최종적인 음소 신뢰도는 다음의 식 (5)에 의해서 계산된다.

$$f(LLR) = \log \frac{1}{1 + \exp(-a \cdot LLR)} \quad (5)$$

2. MFCC (Mel Frequency Cepstrum Coefficient)

음향심리학 연구에서 물리적 주파수와 인간이 인지하는 주관적인 주파수의 피치 사이의 관계는 선형이 아님이 밝혀져 있다. 따라서 순음의 주관적인 피치를 정의하는데, 인간이 주관적으로 느끼는 주파수 스케일을 멜 (mel)이라고 하고 실제 Hz로 표현되는 물리적인 주파수 f 와 구별지어준다. 기준점으로 1 kHz 순음의 피치를 1000 mel 이라 정의한다. 다른 주관적인 피치 값은 기준점에 해당하는 순음의 반 혹은 두 배로 인지되는 순음에 대한 실제 주파수를 측정하여 얻는다. 음성의 특징벡터 추출 방식에서는 사람이 주관적으로 인지하는 주파수 특성을 반영하여 주파수를 멜스케일 (mel-scale)로 warping해서 필터뱅크를 비선형적으로 분포시키는 방식을 사용하는데, 이러한 필터뱅크를 사용해서 구한 음성 벡터를 MFCC라 한다.

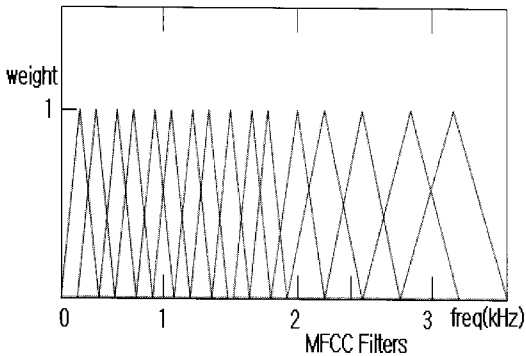
음성신호는 성도를 통과하면서 고주파 성분이 감쇠되는데 이를 보상하기 위해서 식(6)과 같은 전달함수를 사용하였다.

$$X(z) = 1 - aZ^{-1} \quad 0.9 > a > 1.0 \quad (6)$$

위 전달함수를 시간 축으로 해석하면 다음과 같이 나타낼 수 있다.

$$S'_n = S_n - aS_{n-1} \quad (7)$$

n 은 시간인덱스(time index)이고, a 는 프리엠퍼시스(pre-emphasis)계수로써 보통 0.95와 1사이의 값을 주로 사용한다. 프리엠퍼시스를 거친 음성신



<그림 2> Mel-scale 필터뱅크
<Fig. 2> Mel-scale filter bank

호에 대해 해밍윈도우(Hamming window)를 씌운 후, FFT(Fast Fourier Transform) 분석을 통하여 파워스펙트럼(power spectrum)을 구한다. 이 결과를 멜스케일 필터뱅크에 통과 시킨 후 log를 취하고 DCT(Discrete Cosine Transform)를 취하여 MFCC를 얻게 된다.

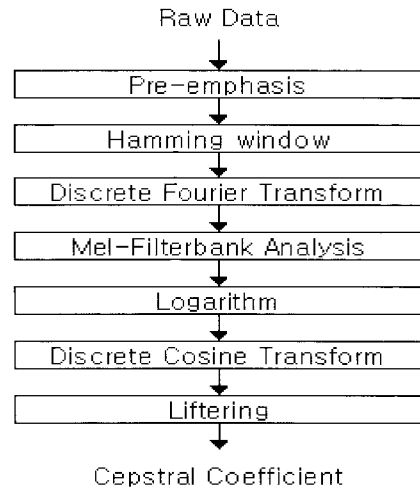
$$MFCC_k = \sqrt{\frac{2}{N} \sum_{i=1}^N \log(x[i]) \cos\left(\frac{2\pi k}{N}(i-9.5)\right)} \quad (8)$$

식 (8)에서 $x[i]$ 는 멜스케일 필터뱅크의 출력이고, N 은 필터뱅크의 대역통과 필터 개수이며, m 은 캡스트럼의 차수이다. <그림 2>는 MFCC 추출과정에서 사용되는 비선형(Non-linear) 삼각필터뱅크(triangular filter bank)이다.

캡스트럼은 전송채널의 변화, 화자의 특성 등으로 인한 영향에 민감하고, 고차의 캡스트럼은 잡음에 민감한 특성을 가지기 때문에 이러한 음성학적 내용과 관계없는 요소에 의한 캡스트럼의 변화를 줄이기 위해서 캡스트럼에 weighting을 가한다. 이를 liftering이라고 한다.

$$c'_m = \left[1 + \frac{Q}{2} \sin\left(\frac{\pi m}{Q}\right)\right] c_m \quad 1 \leq m \leq Q \quad (9)$$

식 (9)에서 Q 는 lifter weight를 나타내고, m 은 캡스트럼의 차수이다. 전체적인 MFCC 특징 추출



<그림 3> MFCC 특징 추출
<Fig. 3> MFCC feature extraction

과정은 <그림 3>과 같다.

3. 음소 유사율의 구성

음소 유사율은 두 음소 사이의 거리를 나타내는 것으로, 바타차랴 거리 측정법(Bhattacharrya Distance Measure)을 사용하여 구한다 [10].

바타차랴 거리 측정법은 두 클래스간의 분리도를 측정하는 통계적 수단으로, 두 개의 가우시안 분포 사이의 거리를 측정한다. 또한 계산이 매우 단순하고, 정확히 거리를 계산하는 것보다 오류에 대한 경계를 제공함으로써 유연성을 갖는다. 바타차랴 거리는 두 개의 클래스에 대하여 다음 식 (10)과 같이 정의된다.

$$b = -\ln \int_{\Omega} [P(X|w_1)P(X|w_2)]^{\frac{1}{2}} dX \quad (10)$$

여기서, $P(X|w_i)$ 는 클래스 w_i ($i=1,2$)의 확률밀도함수이고, Ω 는 확률 분포상에서 정의되는 랜덤 변수 X 의 영역이다. 특히, 클래스의 확률 분포를 정규분포로 가정할 경우 위 식은 다음 식 (11)과 같이 나타낼 수 있다.

$$b = \frac{1}{8}(u_2 - u_1)^T \left[\frac{\Sigma_1 + \Sigma_2}{2} \right]^{-1} (u_2 - u_1) + \frac{1}{2} \ln \frac{|\Sigma_1 + \Sigma_2|/2}{|\Sigma_1|^{1/2} |\Sigma_2|^{1/2}} \quad (11)$$

여기서 u_i 와 Σ_i 는 각각 클래스 w_i 의 평균 벡터와 공분산 행렬이다. 이 경우 위의 식에서와 같이 적분 계산 과정은 필요로 하지 않는다. 본 논문에서 사용한 거리 측정법은 식 (12)와 같다.

$$d(i, j) = \left[\frac{1}{n} \sum_{k=1}^n \frac{(\mu_{ik} - \mu_{jk})^2}{\sigma_{ik} \sigma_{jk}} \right]^{\frac{1}{2}} \quad (12)$$

또한, 다음 식 (13)에 의해 [0, 1] 사이에 분포하도록 정규화 시킨다.

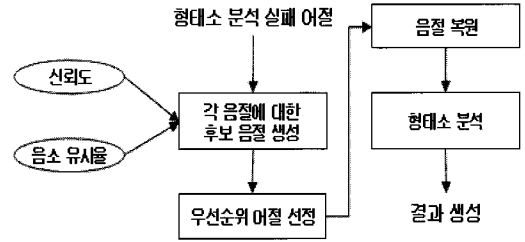
$$S_p = \frac{b_{\max} - b_{xy}}{b_{\max} - b_{\min}} \quad (13)$$

여기서 S_p 는 두 음소의 유사율, b_{\max} 와 b_{\min} , b_{xy} 는 각각 측정된 바타차라 거리의 최대값과 최소값, 두 음소 x 와 y 의 바타차라 거리를 나타낸다. 즉, 유사한 음소일수록 1에 가깝고, 유사하지 않은 음소일수록 0에 가깝게 된다.

음소 유사율은 초성 자음 18개, 모음 22개, 종성 자음 7개로 구성되며, 초성 자음 ‘ㅇ’과 무종성은 음가가 없기 때문에 음성인식 분야에서는 다루지 않는다. 또한, 본 논문의 실험에 사용한 유사음소단위(PLUs)는 모음 “ㅣ”와 “ㄱ”을 같은 음소로 정의하였다 [11]. 따라서 두 모음은 같은 음소 유사율을 가지며, 둘 사이의 유사율은 1이다.

4. 오류보정 알고리즘

<그림 4>와 같이 오류보정을 수행하기 위하여 우선순위 어절을 선정한다. 먼저 어절을 구성하고 있는 음소의 신뢰도와 그 음소에 대한 타 음소와의 음소 유사율을 사용하여 각각의 음소에 대한 후보 음절을 생성하고, 생성한 후보 음절을 통하여 우선순위 어절을 선정한다. 후보 음절을 생성할 때에는 초성 자음과 모음에 대해서는 상위 3개의 음소 유사율을 사용하며, 종성 자음에 대해서는 상위 2개의



<그림 4> 오류보정 알고리즘
<Fig. 4> Error correction algorithm

음소 유사율을 사용한다. 이는 인식 오류의 경우 상위 2개나 3개 안에 대부분의 제대로 된 인식결과들이 포함되기 때문이다[12].

후보 음절을 생성하기 위한 오류보정률은 신뢰도에만 의존하여 결정하지 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. 따라서 신뢰도가 낮으며, 음소 유사율이 높은 음소를 가지고 있는 음소부터 오류보정이 수행되어야 정확한 오류보정을 수행할 수 있다.

하나의 음절에 대한 오류보정률(R_{syllable})은 식 (14)와 같이 나타낼 수 있다.

$$R_{\text{syllable}} = \frac{1}{n} \sum_{k=1}^n (1 - a_k) b_k \quad (14)$$

여기서 n 은 한 음절의 음소 수, a_k 는 k 번째 음소의 신뢰도, b_k 는 k 번째 음소에 대한 유사율을 나타낸다. $a=1$ 일 때, 오류보정률은 0의 값을 가지며, $a \neq 1$ 일 때, $0 < R_{\text{syllable}} < 1$ 의 값을 가진다. 즉, 어떤 음소의 신뢰도가 낮으며 그 음소와 음소 유사율이 높은 음소일수록 오류보정률은 커지게 되며, 음소의 신뢰도가 높거나 그 음소와 음소 유사율이 낮은 음소일수록 오류보정률은 작아진다. 본 논문에서는 각 음절에 대한 상위 10개의 오류보정률을 사용하여 후보 음절을 선정하였다.

후보 음절을 선정한 후 우선순위 어절을 선정한

다. 우선순위 어절은 오류보정률의 평균으로 구하게 되며, 음절 복원과 형태소 분석을 재수행한다. 이때 신뢰도가 높은 음절의 경우 다음 후보 음절과의 오류보정률 차이는 작으며, 신뢰도가 낮은 음절의 경우 다음 후보 음절과의 오류보정률 차이가 크게 되며, 우선순위 어절을 선정할 때 올바른 어절이 낮은 순위로 나타나게 된다.

따라서 식 (15)와 같이 같이 음절의 신뢰도에 따른 임계값을 주어 적절한 오류보정률을 갖게 한다.

$$R_{\text{syllable}} = \lambda \frac{1}{n} \sum_{k=1}^n (1 - \alpha_k) \beta_k \quad \lambda = \sum_{k=1}^n \alpha_k \quad (15)$$

IV. 구현 및 실험 결과

1. 구현

본 논문에서 제안하는 음소 유사율을 구하기 위한 훈련데이터는 ETRI의 PBW445DB를 수동 레이블링된 데이터로 모노폰을 훈련시켜 사용하였다. 또한 특징추출방법에 따른 음소 유사율의 정확성을 알아보기 위해 MFCC와 LPC의 두 가지 특징을 사용하였다.

<표 3> 음소 유사율에 사용된 훈련 데이터
<Table 3> Training data used in the phoneme similarity

번호	어 절	번호	어 절
1	청와대	221	대통령
2	컴퓨터	222	넘도록
3	그에게	223	암컷이
4	위대한	224	주의해야
5	당뇨병	225	예측할
6	그야말로	226	소유의
...
215	이웃고	447	자율화
216	세포의	448	느껴졌다
217	빼놓을	449	위로부터의
218	표적수사	450	양심과
219	터져나온	451	때때로
220	짜증이	452	할까요

<표 4> 실험에 사용된 테스트 데이터
<Table 4> Test data used in the experiment

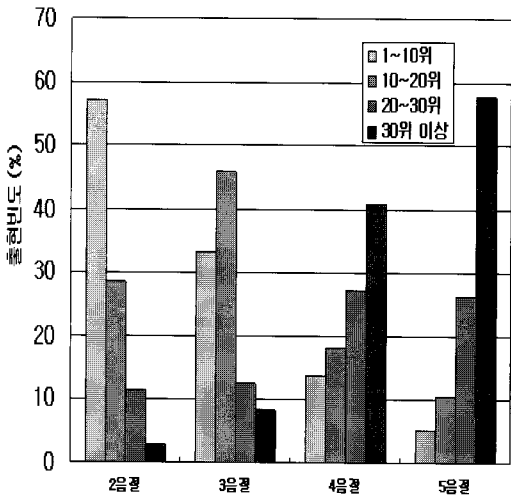
번호	어 절	번호	어 절
1	기본	501	방법도
2	상호	502	복잡한
3	정보	503	남한과
4	전달	504	가지는
5	지식	505	명사를
6	사회	506	학자도
...
495	집단이	995	상대적으로
496	속하며	996	존재하지만
497	따라서	997	분석하려는
498	명사를	998	절대적이다
499	공학적	999	시스템들이
500	붙는다	1000	재배치하고

음성인식기는 음소 인식 방법으로 HTK를 사용하였다. 음성인식기의 훈련 데이터는 ETRI의 PBW445DB를 사용하였으며, 테스트 데이터로는 전공서적에서 선별한 1000개의 어절(2음절 357개, 3음절 243개, 4음절 217개, 5음절 183개)을 사용하였다. <표 3>과 <표 4>는 각각 음소 유사율에 사용된 훈련 데이터와 실험에 사용된 테스트 데이터의 일부이다.

2. 실험 결과

음소 유사율의 구성에 따른 음소 유사율의 정확성을 알아보기 위하여 각각 MFCC와 LPC의 특징 추출방법을 사용하여 음소 유사율을 구성하고, 이를 오류보정에 적용하였다.

음절의 개수에 따라서 올바른 어절이 우선순위 후보 어절중 몇번째에 나타나는지를 확인하기 위해 음절의 개수에 따른 올바른 어절의 순위에 대해서 알아보았다. 실험은 MFCC의 특징추출방법을 사용한 음소 유사율을 사용하였으며, 실험 결과 <그림 5>와 같이 2음절과 3음절의 경우 1~20위 사이에 올바른 어절이 자주 나타났으며, 4음절과 5음절의 경우 20위 이상에 자주 나타났다.

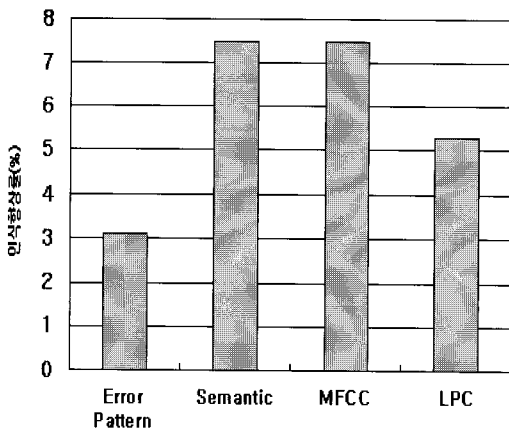


<그림 5> 음절의 개수에 따른 올바른 어절의 출현 빈도

<Fig. 5> Frequency of correct eo-jeol by number of syllable

이는 음절의 개수가 증가함으로써 우선순위 후보 어절에 대한 형태소 분석과 음절 복원의 계산량이 증가함을 나타내며, 이와 같은 불필요한 계산을 줄여주기 위하여 우선순위 어절 선정시 에러패턴 학습을 이용한 오류보정 방법 또는 의미기반의 오류보정 방법을 같이 사용한다면, 계산량 감소와 더 높은 인식 향상률을 얻을 수 있다.

또한 기존의 에러패턴 학습을 이용한 방법[2, 3]



<그림 6> 오류보정률 비교

<Fig. 6> Comparison of error correction

<표 5> 어절 ‘부분적인(부분저기)’의 음소 유사율을 사용한 오류보정의 예
<Table 5> Example of error correction using phoneme similarity

순위	R_{syllable}	음절	R_{syllable}	음절	R_{syllable}	음절	R_{syllable}	음절
1	0.2887	뷰	0.2779	분	0.1958	저	0.2858	긴
2	0.2826	베	0.2750	빈	0.1935	머	0.2836	건
3	0.2819	부	0.2747	분	0.1925	서	0.2831	간
4	0.2812	뵤	0.2729	벤	0.1918	써	0.2826	겐
5	0.2808	꾸	0.2599	꾼	0.1911	지	0.2798	친
6	0.2802	계	0.2576	둔	0.1895	재	0.2796	찐
7	0.2790	꾸	0.2570	뀐	0.1888	미	0.2776	천
8	0.2751	뵤	0.2567	꾼	0.1879	시	0.2774	찐
9	0.2744	뵤	0.2558	뵤	0.1876	자	0.2771	찬
10	0.2741	데	0.2549	젠	0.1872	매	0.2769	짚

과 의미기반의 방법[4], 그리고 본 논문에서 제안한 오류보정 방법의 결과를 <그림 6>에 나타내었다. 에러패턴 학습을 이용한 오류보정의 경우 3.1%, 의미기반의 오류보정의 경우 7.5%의 인식 향상률을 보였으며, 본 논문에서 제안한 MFCC, LPC를 이용한 음소 유사율을 사용할 경우 각각 7.5%, 5.3%의 인식 향상률 결과를 보였다. 또한 MFCC가 LPC보다 우수한 성능을 보였으며, 전체적으로 71.8%, 69.6%의 인식률을 보였다. 단어 단위의 음성인식기에 비하여 낮은 인식률이지만 음성인식기의 사전에 구매받지 않는 음소 단위의 음성인식기인 점을 고려한다면 우수한 성능이라고 할 수 있다. <표 5>는 음소 유사율을 사용하여 오류보정을 수행한 후보 음절의 예를 보여준다.

V. 결론

본 논문에서는 신호처리 위주의 인식결과로부터 좀 더 신뢰할 수 있는 결과를 얻기 위하여 음소 유

사율을 사용한 음성인식 후처리에서의 오류보정을 제안하였다. 음소 유사율은 두 음소 사이의 거리를 나타내는 것으로, MFCC와 LPC의 특징추출방법을 사용하여 특징을 추출한 후, 바타차라 거리 측정법을 사용하여 구성하였다.

본 논문에서 제안한 음소 유사율을 사용하여 오류보정을 수행할 경우 음질의 개수가 증가 할수록 형태소 분석과 음절 복원의 계산량이 증가함을 보였으나, 우선순위 어절 선정시 에러패턴 학습을 이용한 오류보정 방법 또는 의미기반의 오류보정 방법을 같이 사용하여, 계산량을 감소 시킬 수 있다.

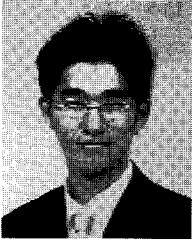
또한 MFCC의 음소 유사율을 오류보정에 적용하였을 때 7.5%의 인식 향상률을 보임으로써 LPC의 음소 유사율이나 에러패턴 학습을 이용한 방법에 비해 우수함을 보였으며, 의미기반의 오류보정 방법과는 비슷한 인식 향상률을 보였다.

본 논문에서 제안한 음소 유사율을 사용한 오류보정은 음성인식을 인터페이스로 하는 텔레매틱스 단말기 등에 적용될 수 있을 것으로 기대된다.

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] 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.
- [3] 김용현, 정민화, "에러패턴 학습과 후처리 모듈을 이용한 연속 음성 인식의 성능향상," *Proc. KISS Spring Semiannual Conf.* vol. 27, no. 1, pp. 441-443, 2000. 4.
- [4] 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.
- [5] 박미성, 김미진, 김계성, 최재혁, 이상조, "연속 음성인식 후처리를 위한 음절 복원 rule-based 시스템과 형태소분석기법의 적용," *대한전자공학회논문지*, 제36권, 제3호, pp. 47-57, 1999. 3.
- [6] 박미성, 김미진, 김계성, 김성규, 이문희, 최재혁, 이상조, "형태소 분석 기법을 이용한 음성 인식 후처리," *대한전자공학회논문지*, 제36권, 제4호, pp. 65-77, 1999. 4.
- [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] 문광식, 김희린, 정재호, 이영직, "가변어휘 단어 인식에서의 미등록어 거절 알고리즘의 성능비교," *신호처리합동학술대회 논문집*, vol. 12, no. 1, pp. 305-308, 1999. 10.
- [9] 문광식, *가변어휘 단어 인식에서의 미등록어 거절 알고리즘*, 석사학위논문, 인하대학교 공과대학, 2000. 2.
- [10] 안태옥, "혼합 가우시안 군집화를 이용한 상태 공유 음향모델 최적화," *대한전자공학회논문지*, 제42권, 제6호, pp. 167-176, 2005. 11.
- [11] 음성정보기술산업지원센터, *한국어 음성인식 플랫폼 사용자 매뉴얼 (ECHOS Manual)*, 2006.
- [12] 정두경, *한국어 연결숫자인식 발화검증 시스템에서 강인한 신뢰도 측정에 관한 연구*, 석사학위논문, 부산대학교전자공학과, 2003. 2.

저자소개



한 동 조 (Han, Dong-Jo)

2007년 9월~현재 : 광운대학교 박사과정 (컴퓨터공학전공)
2007년 2월 : 광운대학교 공학석사 (컴퓨터공학전공)
2005년 3월~2007년 2월 : 광운대학교 공학석사 (컴퓨터공학전공)



최 기 호 (Choi, Ki-Ho)

1990년 4월~현재 : 광운대학교 컴퓨터공학과 교수
2006년 1월~2006년 12월 : 한국 멀티미디어학회 회장
2005년 1월~2005년 12월 : 한국 ITS 학회 회장
1987년 2월 : 한양대학교 공학박사 (전자공학전공)
1977년 3월~1979년 2월 : 한국과학기술연구소 연구원