The Journal of Digital Policy & Management 2013 Jun; 11(6): 137-142

MFCC와 LPC 특징 추출 방법을 이용한 음성 인식 오류 보정

오상엽*
가천대학교 글로벌캠퍼스 IT대학 컴퓨터미디어융합학과*

Speech Recognition Error Compensation using MFCC and LPC Feature Extraction Method

Sang-Yeob Oh*
Dept. of Computer Media Convergence, College of IT, Gachon University*

요 약 음성 인식 시스템은 부정확한 음성 신호의 입력으로 특징을 추출하여 인식할 경우 오인식의 결과가 나타나거나 유사한 음소로 인식된다. 따라서 본 논문에서는 음소가 갖는 특징을 기반으로 음소 유사율과 신뢰도 측정을 이용한 음성 인식 오류 보정 방법을 제안하였다. 음소 유사율은 학습 모델의 음소에 MFCC와 LPC 특징 추출 방법을 이용하여 구하였으며 신뢰도로 측정하였다. 음소 유사율과 신뢰도를 측정하여 오인식되는 오류를 최소화하였으며 음성 인식 과정에서 오류로 판명된 음성에 대하여 오류 보정을 수행하였다. 본 논문에서 제안한 시스템을 적용한 결과 98.3%의 인식률과 95.5%의 오류 보정율을 나타내었다.

주제어 : 음성 인식, 특징 추출, 음소 유사율, 신뢰도 측정, 오류 보정

Abstract Speech recognition system is input of inaccurate vocabulary by feature extraction case of recognition by appear result of unrecognized or similar phoneme recognized. Therefore, in this paper, we propose a speech recognition error correction method using phoneme similarity rate and reliability measures based on the characteristics of the phonemes. Phonemes similarity rate was phoneme of learning model obtained used MFCC and LPC feature extraction method, measured with reliability rate. Minimize the error to be unrecognized by measuring the rate of similar phonemes and reliability. Turned out to error speech in the process of speech recognition was error compensation performed. In this paper, the result of applying the proposed system showed a recognition rate of 98.3%, error compensation rate 95.5% in the speech recognition.

Key Words : Speech Recognition, Feature Extraction, Phonemes Similarity Rate, Reliability Measures, Error Compensation

* 이 논문은 2013년도 가천대학교 교내연구비 지원에 의한 결과임.(GCU-2013-R139)
Received 14 May 2013, Revised 12 June 2013
Accepted 20 June 2013
Corresponding Author: SangYeob Oh(The University of Gachon)
Email: syoh1234@gmail.com

ISSN: 1738-1916

1. 서론

모바일 통신 기술이 발달함에 따라 음성 인식 기반 검색 시스템, 자동 응답 시스템 등 음성 인식을 인터페이스로 하는 시스템들이 개발되고 있으며 음성 인식 시스템에서는 여전히 유사한 음소와 부정확한 입력 음성 신호로 인하여 오류가 존재한다. 이를 위한 신호 처리 단계에서의 음성 인식 오류 보정에 관한 여러 가지 연구가 진행되어 왔다[1]. 하지만 사용 범위가 넓고, 화자 독립적인 최근의 시스템에서 음성 신호 처리만으로 인식의 효율을 높이는 것은 상당히 어렵다. 따라서 음성 인식의 단순한 신호 처리 위주의 인식 결과로부터 좀 더 신뢰할 수 있는 결과를 얻기 위한 오류 보정에 대한 연구가 진행되고 있다[2].

기존의 연구 방법에는 잡음 채널 모델 기반의 오류 보정 방법이 있으며 음성 인식기의 적용 환경과 실제 인식할 때의 조건상의 차이가 있다는 점을 전제로 오류 보정을 수행한다. 하지만 단순한 언어 모델이 가지는 한계점을 극복하지 못하는 단점을 가지고 있다[3]. 인식 과정에서의 오류는 일정한 패턴을 가지고 발생한다는 전제로 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 오인식 될 경우 적용이 불가능한 단점이 있다[5].

따라서 본 논문에서는 음소 유사율과 신뢰도 측정을 이용한 음성 인식 오류 보정 방법을 제안하였다. 부정확한 어휘의 입력으로부터 특징을 추출하여 인식할 경우 유사한 음소로 인식하거나 오인식 오류로 나타나게 되므로 음소 유사율과 신뢰도를 측정하여 오류 보정을 수행하므로 인식률을 향상시켰다. 음소 유사율은 가우시안 분포를 이용하여 구하였으며 신뢰도 측정은 후보군을 확보하여 확률적 계산을 이용하여 구하였으며 후보군에서 오류 보정을 실시하였다.

음소 유사율과 신뢰도를 이용하여 오류 보정률을 구하였으며, 어휘 인식 과정에서 오류로 판명된 어휘에 대하여 오류 보정을 수행하였다. 시스템 성능 평가 결과 98.3%의 인식률과 95.5%의 오류 보정율을 나타내었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 언급하고 3장에서는 음소 유사율과 신뢰도 측정을 이용한 음성 인식 오류 보정 방법에 대해 설명하며 4장에서는 시스템 평가를 수행하고 마지막으로 5장에서 결론을 맺는다.

2. 관련 연구

2.1 음소 특징 추출

특징 추출은 인식에 유용한 성분을 신호로부터 얻어내는 과정이며 일반적으로 정보의 압축, 차원의 감소 과정과 관련되어 추출된 특성에 의해 인식률이 좋고 나쁨으로 판단한다. 흔히 사용되는 방법으로는 특성 추출 과정에서 청각 특성을 반영하는 달팽이관의 주파수 응답을 필터 뱅크 분석으로 사용하며 주파수에 따른 대역폭의 증가, 프리엠퍼시스 필터 등이 사용된다[6].

음성 인식 신호의 동적 특성을 반영하기 위하여 켑스트럼 1차, 2차 미분 값을 사용하며 미분 값은 시간축 방향의 필터링으로 표현된다. 시간 축 방향으로의 특징 벡터를 얻는 과정이며 인식을 위하여 주로 사용되는 특징은 MFCC(Mel Frequency Cepstrum Co-efficient)와 LPC(Linear Predictive Coefficient)가 주로 사용된다[7].

MFCC는 신호를 안티 엘리어싱 필터로 거쳐 A/D 변환 후 디지털 신호 $x(n)$로 변환한다. MFCC 계수는 12개를 사용하며 이와는 별도로 구한 프레임 로그 에너지가 추가적으로 사용되어 인식의 입력으로 사용되는 특성 벡터는 13차 벡터로 구성되어 사용된다[8].

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

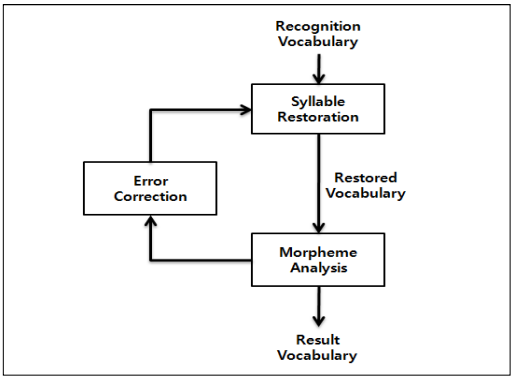
2.2 후처리 오류 보정

후처리 오류 보정은 잡음 제거 방법과 어휘 보정 방법으로 이루어진다. 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[9].

후처리 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 2는 후처리 오류 보정의 구성도를 나타낸다[10].



[Fig. 2] Post-processing Error Correction

3. 음성 인식 음소 유사율 오류 보정

3.1 음소 유사율 측정 및 신뢰도 측정

음소와 음소 사이의 거리를 측정하기 위해 음소 유사율을 사용하며 각 음소의 분리도를 측정하는 통계적 수단으로 가우시안 분포를 사용한다. 계산이 간단하고 오류의 대한 경계값의 조절이 가능하여 많이 사용된다.

음소 유사율의 값이 큰 범위에서 나타나지 않도록 정규화하며 시그모이드 함수를 사용하고 음소 신뢰도는 다음의 식에 의해서 계산된다[11].

$$NCLR(M_i, M_j) = \frac{1}{N_i} log\left(\frac{N(o_i, M_i)}{N(o_i, M_j)}\right) + \frac{1}{N_j} log\left(\frac{N(o_j, M_j)}{N(o_j, M_i)}\right) \quad (4)$$

N_i와 N_j는 o_i와 o_j 데이터의 수를 나타내고 M_i와 M_j는 $N(o_i, M_j)$과 $N(o_j, M_i)$의 각각의 음소 유사율을 나타낸다.

다양한 패턴인식 분야에서 다양한 신뢰도 평가 척도를 사용하고 있다. 그러나 대부분의 연구는 음성인식, 화자인식에 적합한 특징을 사용하고 있으며 아직까지 음성인식에 적합한 신뢰도 평가 척도에 대한 연구는 미비하다. 신뢰도 평가를 위해 가장 널리 사용되고 있는 척도는 정규화된 유사율을 사용하는 방법과 정규화된 유사율과 각각의 패턴인식에 적합한 특징을 결합하여 사용하는 방법이다[12].

음소 단위 신뢰도는 음성 인식기의 출력인 음소 확률과 인식된 음소에 대한 반 음소 확률의 비로 정의되며 반 음소 모델의 평균 로그 확률은 다음 식과 같이 나타낸다.

$$\log pr_a = \frac{1}{M} \sum_{i=0}^{M-1} \log pr_{a_i} \quad (5)$$

M은 반 음소 모델의 수를 나타내며 음소 단위 신뢰도는 다음의 식과 같이 나타낸다.

$$\mathrm{r\,m_p} = \frac{\log pr_p - \log pr_a}{|\log pr_p|} \quad (6)$$

단어 구성 음소 모델의 로그 확률을 $\log pr_p$로 나타낸다. 음소 단위 신뢰도는 음성 인식기의 출력인 음소 확률과 인식된 음소에 대한 음소 단위 신뢰도는 다음 식과 같이 나타낸다.

$$\mathrm{r\,m} = \frac{1}{\mathrm{f_{cm}}} \log\left(\frac{\left(\sum_{\mathrm{p}=0}^{\mathrm{n}-1} \exp(f_{cm} \cdot cm_p)\right)}{\mathrm{n_p}}\right) \quad (7)$$

f_{cm}은 음의 값을 갖는 가중치를 나타낸다. 일반적인 반 음소는 전체 음소 집합에서 인식된 음소를 제외한 나머지 음소에 대하여 HMM을 이용한다.

3.2 유사 음소 인식 오류 보정

오류 보정 처리에서는 음소 단위의 군집 모델들에 대해서는 인식을 위해 사용하고 군집에 포함되지 않는 모델들은 분류하여 특징 추출을 수행한다. 인식 오류의 경우에는 비슷한 인식 결과를 포함하고 있어 후보 음절을

생성할 때 초성 자음과 모음에 대해서 리스트 상에 유사한 음소를 사용한다.

오류 보정 처리는 신뢰도와 음소 유사율을 고루 사용한다. 후보 음절을 선정하여 우선순위 어절을 선정하여 오류 보정을 수행한다. 신뢰도가 높은 음절의 경우 다음 후보 음절과의 오류 보정율 차이는 작으며 신뢰도가 낮은 음절의 경우 다음 후보 음절과의 오류 보정율 차이가 크게 나타난다. 따라서 우선순위 어절을 선정할 때 올바른 어절이 낮은 순위로 나타나게 한다.

신뢰도가 높아도 음소 유사율이 높은 비슷한 음소로 대치되어 오류가 발생하는 경우가 존재하며 신뢰도가 낮지만 정확히 인식될 경우도 존재한다. 신뢰도가 낮으며 음소 유사율이 높은 음소를 가지고 있는 음소부터 오류 보정이 수행되어야 한다.

초기값 1로 설정된 클래스 i에 대한 HMM 가중치 훈련은 클래스 오인식 척도를 감소시키기 위해 요구되는 클래스 i에 대한 HMM 가중치에 대한 차분 계수 $\triangle w_i$를 이용하여 오류 보정률을 다음 식을 이용하여 구한다.

$$\triangle w_i = \frac{d_i(X, \Lambda)}{-g_i(X, \Lambda)_k} \tag{7}$$

조절을 위해서 클래스 i에 대한 HMM 가중치가 변화되는 비율을 설정한다. 클래스 i에 대한 HMM 가중치 조절을 위한 차분 계수는 오인식 척도와 훈련 음성의 클래스 i에 대한 분별 함수의 값을 이용한다.

오류 보정 알고리즘은 다음과 같이 진행된다.

```
Begin
        Initialize the non-recognition
Job1 : Get Input Vector Xl
        Process the likehood(ALL n∈CL)
        data collection response-selective and
        response variable Y to division data K
        Add the NL to Evaluation(n)
        Process the phone transition and
        sum log f(n)
        Add the Expand(n) to NL
        Insert BP to BPframe
        Get the frame of word list BP for transition
        Process word transition
        Replace CL, NL
Job2 : Create the word graph list having
        max likehood of BP table
```

음절의 신뢰도에 따른 임계값을 다음 식과 같이 주어지며 적절한 오류 보정율을 갖게 한다.

$$R = \sum_{k=1}^{n} \alpha_k \cdot \frac{1}{n} \sum_{k=1}^{n} (1-\alpha_k) b_k \tag{8}$$

음소의 신뢰도가 낮고 음소와 음소 유사율이 높은 음소는 오류 보정율이 크며 음소의 신뢰도가 높거나 음소와 음소 유사율이 낮은 음소는 오류 보정률이 작아진다.

4. 실험 결과

본 논문에서 제안한 음소 유사율과 신뢰도 측정을 이용한 음성 인식 오류 보정 방법의 성능 검증을 위하여 인식 실험을 수행하였다. 음성 인식 목록은 서울 시내의 지역명 50개, 지하철명 50개로 구성하였다. 인식 실험은 [13]을 참조하여 실험하였으며 실험에 참가한 3명의 화자가 어휘 목록을 3회 발음하여 총 900단어를 대상으로 실험을 수행하였다.

제안한 시스템의 성능 평가를 위하여 기존 방식과 비교 실험을 하였다. 실험은 화자 종속형과 화자 독립형으로 구분하여 실험하였으며 화자 종속형은 음성 모델을 만들 때 참가하였던 화자가 직접 인식에 참여하여 실험한 것이며 화자 독립형은 음성 모델에 참가하지 않았던 화자가 인식을 한 실험을 나타낸다.

<표 1>은 기존의 방식인 Error Pattern, Semantic 그리고 제안한 방법의 음성 인식률에 대한 실험을 나타낸다.

〈Table 1〉 Recognition Rate

Speech	Recognition Rate (%)		
	Error Pattern	Semantic	Proposed Method
Speech Dependent	96.3	97.1	97.5
	97.8	97.2	98.5
	97.5	97.6	98.9
Speech Independent	97.5	96.9	98.3
	97.2	96.5	98.5
	97.1	96.3	98.3

<표 2>는 Error Pattern, Semantic 그리고 제안한 방법의 오류 보정률에 대한 실험을 나타낸다.

실험 결과 에러 패턴 학습을 이용한 방법의 음성 인식률 평균 97.2%로 나타났으며 의미 기반의 방법을 이용한 음성 인식률 평균 96.9%의 인식률 나타내었고 제안방법의 인식률 평균 98.3%를 나타내었다.

〈Table 2〉 Error Correction Rate

Speech	Error Correction Rate (%)		
	Error Pattern	Semantic	Proposed Method
Speech Dependent	92.3	93.1	95.1
	92.9	92.8	95.8
	94.8	90.6	94.5
Speech Independent	91.6	96.1	97.3
	92.7	94.2	96.3
	91.8	93.2	93.7

또한 에러 패턴 학습을 이용한 방법의 음성 인식 오류 보정률 평균 92.7%로 나타났으며 의미 기반의 방법을 이용한 음성 인식 오류 보정률 평균 93.3%의 인식률 나타내었고 제안 방법의 음성 인식 오류 보정률 평균 95.5%를 나타내었다.

5. 결 론

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 단점을 개선하기 위해 본 논문에서는 음소 유사율과 신뢰도 측정을 이용한 음성 인식 오류 보정 방법을 제안하여 성능을 평가하였다.

부정확한 어휘의 입력으로부터 특징을 추출하여 인식할 경우 유사한 음소로 인식하거나 오인식 오류로 나타나게 되므로 음소 유사율과 신뢰도를 측정하여 오류 보정을 수행하므로 인식률을 향상시켰다. 음소 유사율은 가우시안 분포를 이용하여 구하였으며 신뢰도 측정은 후보군을 확보하여 확률적 계산을 이용하여 구하였으며 후보군에서 오류 보정을 실시하였다.

음소 유사율과 신뢰도를 이용하여 오류 보정률을 구하였으며, 어휘 인식 과정에서 오류로 판명된 어휘에 대하여 오류 보정을 수행하였다. 시스템 성능 평가 결과 평균 98.3%의 인식률과 평균 95.5%의 오류 보정율을 나타내었다.

ACKNOWLEDGMENTS

This work was supported by the Gachon University research fund of 2013."(GCU-2013-R139)

REFERENCES

[1] Chan-Shik Ahn, Sang-Yeob Oh. Key-word Error Correction System using Syllable Restoration Algorithm. Journal of the Korea Society of Computer and Information. Vol. 15, No. 10, pp. 165-172, 2010.

[2] Chan-Shik Ahn, Sang-Yeob Oh. Gaussian Model Optimization using Configuration Thread Control In CHMM Vocabulary Recognition. The Journal of Digital Policy and Management. Vol. 10, No. 7, pp. 167-172, 2012.

[3] Chan-Shik Ahn, Sang-Yeob Oh. Vocabulary Recognition Post-Processing System using Phoneme Similarity Error Correction. Journal of the Korea Society of Computer and Information. Vol. 15, No. 7, pp. 83-90, 2010.

[4] Tae-Hee Kwon, Han-Seok Ko. Performance Improvement in Speech Recognition by Weighting HMM Likelihood. The Journal of the Acoustical Society of Korea. Vol. 22, No. 2, pp. 145-152, 2003.

[5] C. Miyajima, K. Tokuda, T. Kitamura. Minimum classification error training for speaker identification using gaussian mixture models based on multi-space probability distribution. EUROSPEECH, Vol. 4, pp. 2837-2840, 2001.

[6] S. Chu, S. Narayanan, C. C. Jay Kuo. Environmental Sound Recognition With Time-Frequency Audio Features. IEEE Trans. on Audio, Speech, and Language Processing, Vol. 17, No. 6, pp. 1-16, 2009.

[7] M. Cowling, R. Sitte. Comparison of techniques for

environmental sound recognition. Pattern Recognition Letters, Vol. 24, No. 15, pp. 2895-2907, 2003.
[8] Yusuke Kida, Hiroyoshi Yamamoto. Minimum classification error interactive training for speaker Identification. IEEE International conference, Acoustic, Speech and Signal processing, Vol. 1, pp. 641-644, 2005.
[9] Tariquzzaman, Md, Min, So-Hui, Kim, Jin-Yeong, Na, Seung-Yu. Modified HMM Decoder based on Observation Confidence for Speaker Identification. Proceedings of the Korean Institute of Intelligent Systems Conference. pp. 443-446. 2007.
[10] Dong-Jo Han, Ki-Ho Choi. A Study on Error Correction Using Phoneme Similarity in Post-Processing of Speech Recognition. The Journal of The Korea Institute of Intelligent Transport Systems. Vol. 6, No. 3, pp. 77-86, 2007.
[11] Ji-Eun Kim, In-Sung Lee. Speech/Mixed Content Signal Classification Based on GMM Using MFCC. Journal of the Institute of Electronics Engineers of Korea. Vol. 50, No. 2, pp. 185-192, 2013.
[12] Jong-Young Ahn, Sang-Bum Kim, Su-Hoon Kim, Kang-In Hur. A study on Voice Recognition using Model Adaptation HMM for Mobile Environment. The Journal of the Institute of Webcasting, Internet and Telecommunication. Vol. 11, No. 3, pp. 175-179, 2011.
[13] Chan-Shik Ahn, Sang-Yeob Oh. Phoneme Similarity Error Correction System using Bhattacharyya Distance Measurement Method. Journal of the Korea Society of Computer and Information. Vol. 15, No. 6, pp. 73-80, 2010.

오 상 엽(Oh, Sang Yeob)

・1991년 2월 : 광운대학교 대학원 전자계산학과 (이학석사)
・1999년 2월 : 광운대학교 대학원 전자계산학과 (이학박사)
・2007년 2월 ~ 현재 : 가천대학교 IT대학 인터랙티브미디어학과 교수
・관심분야 : 버전관리, 형상관리, 음성/음향 신호 처리, 차량 통신
・E-Mail : syoh1234@gmail.com