외국어 발화오류 검출 음성인식기를 위한 스코어링 기법

강효원* 배민영* 이재강** 권철홍*
* 대전대학교 정보통신공학과
** 대전대학교 일어일문학과

Machine scoring method for speech recognizer detecting mispronunciation of foreign language

Hyowon Kang*, Minyoung Bae*, Jaekang Lee** and Chulhong Kwon*
* Department of Information and Communications Engineering, Daejeon Univ.
** Department of Japanese Language and Literature, Daejeon Univ.

kanghyowon@hotmail.com, missbea79@hotmail.com, ljgang@dju.ac.kr, chkwon@dju.ac.kr

Abstract

An automatic pronunciation correction system provides users with correction guidelines for each pronunciation error. For this purpose, we propose a speech recognition system which automatically classifies pronunciation errors when Koreans speak a foreign language. In this paper, we also propose machine scoring methods for automatic assessment of pronunciation quality by the speech recognizer. Scores obtained from an expert human listener are used as the reference to evaluate the different machine scores and to provide targets when training some of algorithms. We use a log-likelihood score and a normalized log-likelihood score as machine scoring methods. Experimental results show that the normalized log-likelihood score had higher correlation with human scores than that obtained using the log-likelihood score.

Ⅰ. 서론

본 논문에서는 음성인식 기술의 응용분야인 외국어 학습기에 초점을 맞추어 한국인의 외국어 발음 교정 시스템을 위한 발음오류유형 자동분류 음성인식기를 다룬다 특히 본 논문에서는, 발음에 따른 오류 양상을 음성인식기로 분류하여 그 오류에 해당하는 교정사항을 제공하기 위하여, 원어민의 발음과 그 오류에 해당하는 유사발음을 정확하게 구분해 낼 수 있도록 음성인식기의 성능을 향상시키는 데 목적이 있다. 이 목적을 위해, 음성인식기가 표준 및 오류 발음을 자동으로 분류하기 위한 학습자의 발음에 스코어를 주는 방법에 대한 연구를 수행하였다. 이 스코어는 전문음성학자의 청취판단과 상관관계가 높은 것이 필요하므로 음성학자의 청취판단과의 결과를 비교한다 즉, 본 연구의 초점은 발음오류유형 자동분류 음소인식기가 인식해 낸 결과와 해당 언어 음성학 전문가의 청취판단 결과가 가능한 한 유사하게 나타내게 하는 데 있다

이를 위해서 언어모델에 가중치를 적용한 스코어를 사용하여 성능개선을 시도한 바 있다[1]. 이 스코어는 언어모델의 확률을 청취판단에 근거하여 가중치를 적용하였다. 실험결과로부터 가중치를 주기 전과 비교하여 11.5%의 성능개선을 이루었다. 이 실험으로부터, 음성인식기의 스코어 방법에 따라 음성학자의 청취판단과 보다 높은 상관관계를 갖는 결과를 얻을 수 있다는 사실을 알 수 있다. 본 논문에서는 언어별 음소인식기에 존재하는 로그유사도의 편차를 보상해 주는 스코어링 방법을 제안하여 추가적인 성능개선을 이루고자 한다

본 논문의 구성은, 2장에서 언어별 음소인식기의 구현 방법과 실험 결과를 소개하고, 언어별 음소인식기에 존재하는 성능의 차이를 알아본다 3장에서는 언어별 음소인식기의 로그 유사도 편차를 보상해 주는 정

규화 로그 유사도 스코어를 제안한다. 4장에서 실험 결과를 기술하고, 그리고 5장에서 결론을 맺는다.

Ⅱ. 언어별 음소인식기

2.1. 언어별 음소인식기 구현

본 절에서는 한국어 및 일본어 음소인식기의 구현 방법에 대하여 설명한다. 언어별 모노폰 셋의 구성, 음성 DB 구축 방법, 음소별 HMM 모델 생성 방법은 다음과 같다.

한국어 모노폰 셋은 자음은 변이음을 고려하여 29개를, 모음은 음성학적 차이를 보이는 모음만을 고려하여 17개를 선정하여[2] 총 46개의 모노폰으로 구성하였다. 일본어 모노폰 셋은 일본음향학회에서 선정한 모노폰 셋을 참조하여[3] 38개의 모노폰을 선정하였다.

음성 DB는 연구실 수준의 조용한 방에서 1명씩 동시에 3곳에서 녹음하였다. 음성 데이터를 PC에서 수집하였고, 사운드카드는 Soundblaster Audigy를, 마이크는 SHURE 565SD를 사용하였다. 한국어 음성 DB는 SITEC(음성정보기술 산업지원센터)에서 작성한 한국어 PBW(Phonetically Balanced Words) 452개 단어를 대전대학교 대학생 70명을 대상으로 수집하였다. 일본어 음성 DB는 ATR(일본 자동통역 연구소)에서 작성한 PBW 216개 단어를 고려대학교에서 한국어 연수 중인 일본인 70명을 대상으로 녹음하였다. 각 언어별 음성 DB에서 50명의 음성 DB를 HMM 모델 생성을 위한 훈련용으로, 20명분을 인식기의 성능을 테스트하기 위한 음성 DB로 사용하였다.

한국어 및 일본어 음소인식기를 다음과 같은 방법으로 구현하여 성능을 평가하였다. 음성신호의 분석은 매 10msec 마다 25msec의 Hamming 창함수를 사용하여 MFCC 39차를 추출하였고, HMM의 구조는 3 state left-to-right continuous HMM을 사용하여 언어별 음소별 음향모델을 생성하였다.

2.2. 언어별 음소인식기 실험결과

언어별 음소인식기 실험 결과, Mixture 수가 1인 경우 한국어 및 일본어 음소인식기의 인식률은 각각 52.2%, 75.4%로 나타나, 일본어 음소인식기의 성능이 한국어인 경우보다 23.2% 더 좋은 인식률을 보여 주었다. Mixture 수에 따른 언어별 인식 성능의 변화를 보여 주는 그림 1을 살펴보면, Mixture 수가 증가함에 따라 인식 성능의 큰 향상을 볼 수 있는데, Mixture 수가 10인 경우부터 인식률이 포화되고 있음을 알 수 있다 일본어인 경우 Mixture 15인 경우 인식률이 84.4%로 Mixture 1인 경우보다 9.0% 향상되었고, 특히 한국어인 경우는 71.3%로 19.1%의 큰 성능 개선이 이루어졌다. 따라서 두 언어의 인식 성능 차이는, Mixture 1인 경우 약 23.2%에서 Mixture 15인 경우 약 13.1%로 차이가 좁혀짐을 알 수 있다. 그러나 여전히 두 언어의 인식 성능은 큰 차이를 보이고 있다. 이와 같은 실험 결과는 일본어보다 한국어 음소끼리의 음성 자질이 유사하다는 사실을 반영한 것으로 생각된다.

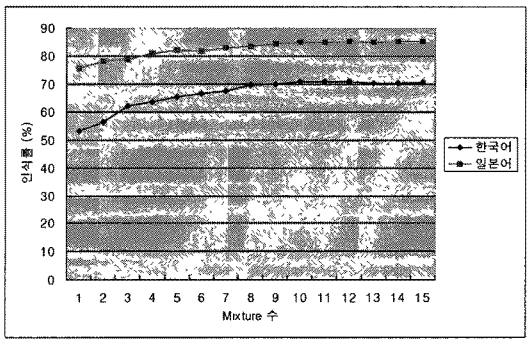

그림 1. Mixture 수에 따른 언어별 인식률의 변화

Ⅲ. 정규화 로그유사도 스코어

음성인식기에서 사용하는 일반적인 음향모델 로그 유사도 스코어와 청취판단(human expert ratings) 사이의 상관성은 일본어 음소인식기와 한국어 음소인식기의 로그 유사도의 차이를 정규화 함으로써 향상될 수 있다. 정규화는 각 언어별로 평균 로그 유사도에 대한 차이를 각 프레임에 대하여 보상해 주는 방법을 사용한다.

언어별 로그 유사도를 정규화 하는 이유는, 학습자가 현재 음성 세그먼트를 발음한 것이 한국어 음소에 가깝게 발음했음에도 불구하고, 로그 유사도 값이 원어민의 발화 모델과 유사도가 높을 경우 한국어 음소로 발음한 것을 일본어로 오인식하는 문제를 해결할 수 있다는데 있다[4]. 이것은 2절에서 살펴보았듯이 일본어 음소인식기의 성능이 한국어 보다 우수하다는 사실로부터 예측 가능하다. 따라서 한국어 음소인식기와 일본어 음소인식기의 로그 유사도의 전체 평균의 차이를 한국어 음소에 대하여 프레임별로 더해 줌으로써 한국어와 일본어 음소인식기의 로그 유사도의 차이로 인한 오인식을 줄일 수 있다

각 음소별 평균 로그 유사도는 음소별 로그 유사도의 합을 음소 유지기간 즉 프레임 숫자로 나누어 평균

을 구하였고, 언어별 평균 로그 유사도는 음소별 평균 로그 유사도의 합을 음소 수로 나누어 평균을 구하였다. 실험 결과 일본어 평균 로그 유사도는 -70.71이고 한국어인 경우는 -72.73으로, 일본어의 평균 로그 유사도 값이 2.02[-70 71-(-72 73)] 더 좋음을 알 수 있다.

IV. 실험 결과

4.1. 실험방법

실험을 위한 발음 네트워크는 기본적으로 그림 2와 같은 네트워크로, 일본어 음소 [k]에 대하여 유사발음 일본어 음소(ky, g, gy)와 한국어 오류음소(/ㅋ/khc, /ㄲ/kkc, /무성 ㄱ/kc, /유성 ㄱ/gc)를 통합한 네트워크를 사용하였다. 일본어 음소 모델은 일본인이 발화한 일본어 음성 DB로, 한국어 음소 모델은 한국인이 발화한 한국어 음성 DB로 생성하였고, Mixture 수는 일본어, 한국어 음소모델 모두 15를 사용하였다.

발음오류 자동분류 음성인식기의 성능을 검증하기 위하여 비전공 한국인 대학생 남성 15명, 여성 15명 등 30명이 일본어 PBW 단어 24개를 발화한 음성 데이터를 수집하였다 이들을 대상으로 인식 실험을 수행하여 유사음소별 인식결과를 구했다

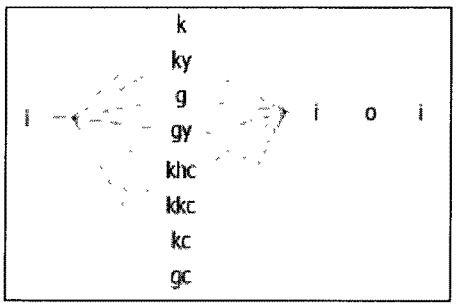

그림 2. 오류발음 검출 발음 네트워크
(일본어 단어 i-k-i-o-i 인 경우)

4.2. 일반적인 로그 유사도를 적용한 스코어링에 따른 인식률 및 정확도 분석

이 절에서는 일반적인 로그 유사도 스코어를 적용하여 실험을 하였다. 실험결과 어두 음소와 어중 음소에서 각 음소별로 크게 차이를 보이므로 실험결과를 어두와 어중으로 나누어 분석하였다. 표 1, 2에 어두와 어중인 경우 청취판단과의 일치도를 보여주고, 표 3에 음소별 결과가 보인다. 어중의 /ㄲ/ 음소를 제외한 나머지 경우에 성능 개선의 필요성이 있음을 알 수 있다

표 1. 어두 [k] 음소의 청취판단 결과와의 정확도 비교

음소	청취판단 개수	청취판단과 일치한 인식 개수	정확도(%)
ㄲ(kkc)	4	1	25.0
ㅋ(khc)	67	13	19.4
ㄱ(kc/gc)	139	37	26.6

표 2. 어중 [k] 음소의 청취판단 결과와의 정확도 비교

음소	청취판단 개수	청취판단과 일치한 인식 개수	정확도(%)
ㄲ(kkc)	318	191	60.1
ㅋ(khc)	130	47	36.2
ㄱ(kc/gc)	62	6	9 7

표 3. 음소별 청취판단 결과와의 정확도 비교

음소	청취판단 개수	청취판단과 일치한 인식 개수	정확도(%)
ㄲ(kkc)	322	192	59.6
ㅋ(khc)	197	60	30 5
ㄱ(kc/gc)	201	43	21.4

4.3. 정규화 로그 유사도를 적용한 스코어링에 따른 인식률 및 정확도 분석

음향모델 스코어는 3절에서 설명한 것과 같이 각 언어별 음소인식기의 성능 차이를 보상해 줌으로써 각 언어의 평균 유사도의 차이로 인한 오인식을 막아 줄 수 있다.

표 4와 5는 정규화 로그 유사도를 스코어로 사용한 인식기의 결과와 청취판단과의 일치도를 보여 준다. 표 4, 5를 보면 표 1, 2와 비교해서 모든 경우에서 청취판단과의 일치도가 증가함을 알 수 있다. 표 1과 4를 보면, 어두 [k] 음소인 경우 /ㅋ/ 음소는 19 4%에서 26.9%로 7 5%, /ㄱ/ 음소는 26.6%에서 40 3%로 13 7%가 향상되었다 표 2와 5를 보면, 어중 [k] 음소인 경우 /ㄲ/ 음소는 60 1%에서 75.8%로 15.7%, /ㅋ/ 음소는 36.2%에서 57 7%로 21.50%가 향상되었다. 음소별로 살펴보면, /ㄲ/ 음소인 경우 59 6%에서 75.5%로 15 9%, /ㅋ/ 음소는 30.5%에서 47.2%로 16.7%, /ㄱ/ 음소는 21.4%에서 31.3%로 9.9%가 향상되었다(표 3과 6 비교). 청취판단과의 비교대상 전체 발화 데이터 720개 중에서, 로그 유사도 정규화 전 방식은 41 0%(295

개)의 일치도를 보여 주었고, 정규화 후 방식은 55.4%(399개)의 일치도를 보여 주어 14.4%의 성능 개선을 이루었다 이와 같은 결과로부터 본 논문에서 제안한 정규화 로그 유사도를 스코어로 사용한 방법이 상당한 성과를 얻었다는 것을 알 수 있다.

표 4. 로그 유사도 정규화 후 어두 [k] 음소의 청취판단 결과와의 정확도 비교

음소	청취판단 개수	청취판단과 일치한 인식 개수	정확도(%)
ㄲ(kkc)	4	2	50 0
ㅋ(khc)	67	18	26.9
ㄱ(kc/gc)	139	56	40.3

표 5. 로그 유사도 정규화 후 어중 [k] 음소의 청취판단 결과와의 정확도 비교

음소	청취판단 개수	청취판단과 일치한 인식 개수	정확도(%)
ㄲ(kkc)	318	241	75 8
ㅋ(khc)	130	75	57.7
ㄱ(kc/gc)	62	7	11.3

표 6. 로그 유사도 정규화 후 음소별 청취판단 결과와의 정확도 비교

음소	청취판단 개수	청취판단과 일치한 인식 개수	정확도(%)
ㄲ(kkc)	322	243	75.5
ㅋ(khc)	197	93	47.2
ㄱ(kc/gc)	201	63	31 3

V. 결론

본 논문에서는 음성인식 기술의 응용분야인 외국어 학습기에 초점을 맞추어 발음 교정 시스템의 발음오류 유형 자동분류 음성 인식기를 구현하였다. 음성인식기는 표준 및 오류 발음을 자동으로 분류하기 위하여 학습자의 발음에 스코어를 주는데, 본 논문에서는 학습자의 개개발음 음소에 대하여 발음 품질을 자동으로 평가하는 측정 방식에 대한 연구를 수행하였다. 이 스코어는 전문음성학자의 판단과 상관관계가 높은 것이 필요한데, 두개의 스코어를 실험하여 어느 것이 음성학자의 청취판단과 더 가까운가를 결정하였다.

로그 유사도 스코어와 각 언어별 로그 유사도의 편차를 보상해 주는 정규화 로그 유사도 스코어 등 두 가지 스코어링 방법을 적용하여 실험한 결과, 언어별로 정규화 된 스코어가 음성학자의 판단과 더 가까운 결과를 얻었다. 이 결과는 발화자의 다양성에 강인하도록 정규화 시킨 효과를 보았다고 할 수 있다.

앞으로의 과제는 혼동되기 쉬운 음소들에 대해 HMM 최적화 방법들을 연구하는데 있어서 앞에서 행했던 실험들에 대한 파라미터들을 조정해가면서 어떤 파라미터 값을 취했을 때 음소인식기가 전문가의 청취판단과 가장 근접한 인식을 하는가를 조사하고, 또한 제시한 스코어링 방법의 알고리즘에 관한 연구를 지속적으로 수행하여 발음교정 시스템에 최적화된 음성인식 엔진을 구현하는데 있다.

참고문헌

[1] 강효원, 권철홍, "외국어 발화오류 검출 음성인식기의 성능개선을 위한 스코어링 기법", 말소리 49호, pp. 1-11, 2004.

[2] 권철홍, 강효원, 이상필, "음성인식기를 이용한 한국인의 외국어 발화오류 자동검출", 말소리 48호, pp. 15-23, 2003.

[3] T. Kawahara et al., "Sharable software repository for Japanese large vocabulary continuous speech recognition", Proc. ICSLP 98, pp.3257-3260, Sydney, 1998.

[4] Y. Kim, H. Franco, and L. Neumeyer, "Automatic pronunciation scoring of specific phone segments for language instruction," Proc. of Eurospeech 97, pp 645-648, 1997.