

혼동음소판별훈련을 통한 영어발음평가정확도개선의 한가지 방법

오영근, 방금현

컴퓨터지원언어학습(CALL: Computer Assisted Language Learning)이 발전하면서 발음교육의 중요성으로부터 음성인식기술을 발음교육에 도입하여 학생들의 발음을 자동적으로 평가하기 위한 연구[1-3]들이 광범히 진행되고있다.

발음평가에서 중요한것은 혼동하기 쉬운 음소들을 잘 분리하여 평가하는것이다. 하지만 음성인식을 위한 음향모형은 최대우도추정방법으로 훈련되므로 혼동하기 쉬운 음소들에 대한 판별능력이 약하다.

논문에서는 최소음소오류/최소단어오류(MPE/MWE)에 기초한 판별훈련을 통하여 음향모형을 갱신하여 발음평가를 보다 정확하게 할수 있는 한가지 방법을 제안하였다.

1. 문 제 설 정

선행연구[1]에서는 중국어에서 혼동하기 쉬운 음소들을 찾고 판별훈련을 통하여 중국어음성인식엔진의 음향모형을 갱신하여 중국어발음평가를 보다 정확히 진행할수 있는 방법을 제기하였다.

선행연구[2]에서는 영어에서 혼동하기 쉬운 음소들을 찾고 모호측도개념을 리용하여 발음을 평가하는 방법을 제기하였지만 평가가 발성단위로 진행되고 정량적으로, 구체적으로 진행되지 못하였다.

논문에서는 혼동음소모임을 정의하고 혼동음소판별훈련을 통하여 영어음성인식엔진의 음향모형을 갱신하여 발음평가의 정확도를 높이기 위한 방법을 제안하였다.

2. 영어에서 혼동하기 쉬운 음소

혼동음소에는 두가지 유형이 있는데 하나는 자음-자음혼동음소와 모음-모음혼동음소이다.

선행연구[2]에서는 중국어를 모국어로 하는 영어학습자들에 대하여 혼동음소를 표 1과 같이 찾았다. 하지만 혼동음소모임은 모국어에 따라 서로 다르다.

또한 선행연구[2]에서는 모음-모음 혼동에 대하여 긴 모음과 짧은 모음도 영어음소표기에 따라 서로 다른 음소로 보고 혼동음소를 정의하였다.

하지만 논문에서 리용하는 음성인

표 1. 선행방법[2]의 혼동음소모임

음소	혼동음소	음소	혼동음소	음소	혼동음소
e	æ	i:	ei	au	o:
	aɪ		i		ɔ
o:	ɔ	ɔ	aʊ	æ	e
	aʊ		ɔ:	ə:	ə:
u:	u	ŋ	n	ə:	æ
aɪ	e	i	i:	u	u:
ʌ	ei	v	w	w	v
s	θ	θ	s	z	ð
ð	z	n	ŋ	ei	i:
					ʌ

식체계에 정의된 영어음소모임은 43개로서 긴 모음과 짧은 모음은 발성연장시간만 다른 같은 음소로 보았기때문에 혼동음소를 표 2와 같이 찾았다. 혼동음소모임은 탐색공간의 동적생성과 발음평가에서 리용된다.

표 2. 논문에서 정의한 혼동음소모임

음소	혼동음소	음소	혼동음소	음소	혼동음소
e	æ	i	ei	aʊ	o
ɔ	aʊ	æ	e	ə	ɔ
	ə				
b	v	f	h	d	ð
ð	d	s	p	θ	s
l	r	r	l	v	b

3. 탐색공간의 동적생성

일반적인 음성인식체계와는 달리 언어학습을 위한 발음평가체계에서는 학생이 발성하는 문장을 이미 알고있다.

따라서 음성인식을 위하여 미리 훈련된 언어모형을 발음평가에 그대로 리용하는것은 속도측면이나 정확성측면에서 다 적합하지 않다.

발음평가에서는 학습자가 발성하게 되는 문장에 기초하여 탐색공간을 동적으로 생성하도록 할수 있다.

탐색공간을 생성할 때 위에서 언급한 혼동하기 쉬운 음소모임을 참고한다.

실례로 문장이 《This is a book》인 경우 음소전개를 진행하면 ðis iz ə buk로 되며 여기에 혼동음소모임까지 고려하면 this는 8가지, is는 두가지, a는 두가지, book는 두가지의 서로 다른 이행경로가 존재할수 있다.

《This is a book》라는 문장에 대하여 이렇게 혼동음소모임까지 고려하여 탐색공간을 만들면 그림과 같은 상태이행 그래프로 표기할수 있다. SIL은 무음에 대한 음소표기이다.

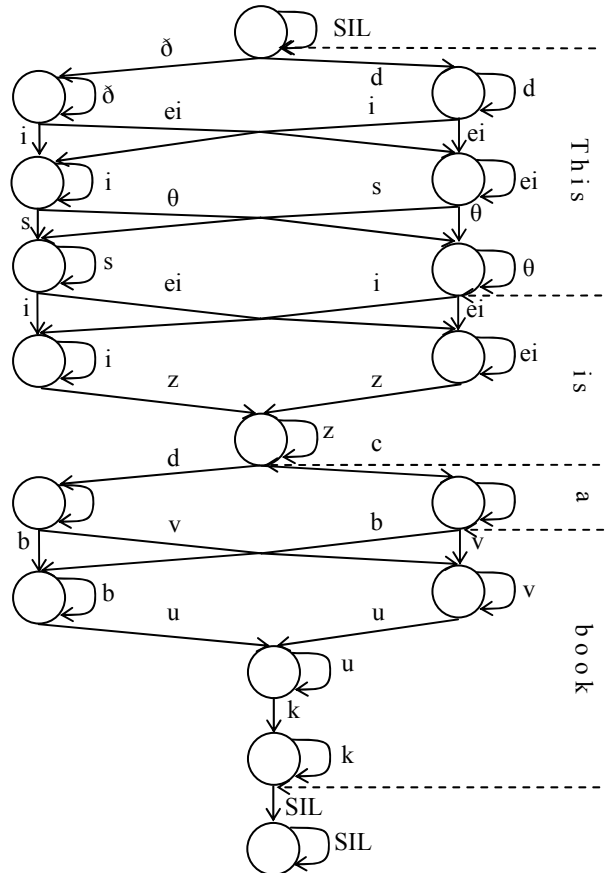


그림. 실례문장 《This is a book》에 대하여 동적으로 생성된 탐색공간

4. MPE/MWE에 기초한 음향모형의 판별훈련

음소모임이 l 개의 음소를 가지고 때 음소가 s 개의 상태를 가진 숨은마르코프모형(HMM) θ_i 로 표시된다고 하자.

그러면 음향모형 θ 는 다음과 같이 표시할수 있다.

$$\theta = \{\theta_1, \theta_2, \dots, \theta_l\}$$

이때 MPE/MWE의 목표는 다음과 같이 θ 를 조정하여 음소오유/단어오유를 최소화하고 정확한 음소/단어의 수를 최대화하는것이다.

$$F(\theta) = \sum_{W \in M} P_{\theta}(W|O)\alpha(W, W_r) \quad (1)$$

$$\theta = \arg \max_{\theta} F(\theta)$$

여기서 $P_{\theta}(W|O)$ 는 단어사후확률로서 다음과 같이 계산할수 있다.

$$P_{\theta}(W|O) = \frac{P_{\theta}(O|W)P(W)}{\sum_{W' \in M} P_{\theta}(O|W')P(W')} \quad (2)$$

식 (1)과 (2)에서 W 는 인식된 단어열, W_r 는 참고단어열이다.

$\alpha(W, W_r)$ 는 인식된 단어열 W 에 대한 정확도로서 다음과 같이 계산할수 있다.

$$\alpha(W, W_r) = \max_z \begin{cases} -1 + 2e(q, z), & q = z \\ -1 + e(q, z), & q \neq z \end{cases} \quad (3)$$

여기서 q 는 인식된 단어열 W 안의 음소/단어, z 는 참고단어열 W_r 의 q 에 대응하는 음소/단어, $e(q, z)$ 는 q 와 z 의 겹침률이다.

5. 실험 및 결과분석

논문에서는 판별훈련전의 음향모형과 판별훈련후의 음향모형에 대한 음소판별실험과 논문에서 제안한 방법과 선행방법[2]에 대한 발음평가실험을 진행하였다.

음향특징벡토르는 평균정규화된 멜케프스트라姆결수(MFCC)와 그것의 1계, 2계시간도함수를 포함한 MFCC_0_D_A_Z 39차원파라미터이다.

1) 음소인식실험

실험을 위하여 영어를 모국어로 하는 사람들이 발성한 음소가 정확하다는 가정밑에서 각이한 혼동음소들을 포함한 400개 문장의 남, 녀음성을 준비하였으며 판별훈련전의 음향모형과 판별훈련후의 음향모형을 가진 음성인식엔진으로 인식하도록 하였다.(표 3)

표 3. 음향모형의 음소인식정확도 비교

음향모형	음소인식정확도/%	
	전 체	혼 동
판별훈련전	88.7	65.1
판별훈련후	92.3	72.4

2) 발음평가실험

실험을 위하여 영어를 학습하는 남녀학생들의 100문장분량의 음성을 준비하고 전문가가 미리 평가하였다.

논문에서는 선행방법[2]으로 평가된 점수와 논문에서 제안한 방법으로 갱신된 모형을 리용하여 평가된 점수를 전문가에 의해 미리 평가된 점수사이에 다음과 같은 교차상관을 계산하여 정확도를 평가하였다.

$$CC = \frac{\sum_{r=1}^R [(s_r - E(s_r))(\tilde{s}_r - E(\tilde{s}_r))]}{\sqrt{\sum_{r=1}^R (s_r - E(s_r))^2 \times \sum_{r=1}^R (\tilde{s}_r - E(\tilde{s}_r))^2}} \quad (4)$$

여기서 R 는 전체 문장수, s_r 는 r 번째 문장에 대한 전문가의 평가점수, \tilde{s}_r 는 컴퓨터에 의한 평가점수이다.

$$\tilde{s}_r = \frac{1}{N_r} \sum_{n=1}^{N_r} \ln P(\theta_i | O_r^n) \quad (5)$$

식 (5)에서 N_r 는 r 번째 문장의 음소개수이다.

또한 상대개선률을 계산하였다.

$$RI(\%) = \frac{CC_{\text{제안}} - CC_{[2]}}{CC_{[2]}} \times 100 \quad (6)$$

실험결과는 표 4와 같다.

표 4. 발음평가정확도

평가방법	교차상관
제안한 방법	0.674
선행방법[2]	0.656
상대개선률/%	2.74

3) 결과분석

음소인식실험에서는 인식정확도가 전체 음소에 대해서는 3.6%, 혼동음소에 대해서는 6.7% 개선되었으며 발음평가실험에서는 제안한 방법이 선행방법[2]보다 전문가에 의한 점수와의 교차상관이 커진것으로 즉 보다 정확한것으로 나타났다.

맺 는 말

영어음성인식을 위하여 훈련된 음향모형에 대하여 혼동음소판별훈련을 진행하여 음소인식정확도가 보다 높은 음향모형을 얻었다. 이 음향모형을 리용하면 영어발음평가의 정확도를 더 높일수 있다는것을 확인하였다.

참 고 문 헌

- [1] Ke Yan, Shu Gong; Information Technology and Computer Science, 2, 17, 2011.
- [2] Xin-Guang Li et al.; 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, 2189, 2016.
- [3] Cini Kurian; Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Technique, 7, 2016.

A Method to Improve Correctness of English Pronunciation Evaluation through Discriminative Training of HDP

O Yong Gun, Pang Kum Hyon

In this paper, we suggested a method which could evaluate the pronunciation more precisely using an acoustic model updated through the discriminative training based on MPE/MWE(Minimum Phoneme Error/Minimum Word Error) and verified its accuracy via several tests.

Keywords: pronunciation evaluation, discriminative training, hard to be distinguished phoneme