

영어 발음교정시스템을 위한 발음사전 구축

김효숙, 김선주

(주) 언어과학 부설 언어공학연구소

Pronunciation Dictionary for English Pronunciation Tutoring System

Kim, Hyosook & Kim, Sunju

Center for Linguistic Engineering, Eoneo Inc.

e-mail:{hyosook, sunju}@eoneo.co.kr

Abstract

This study is about modeling pronunciation dictionary necessary for PLU(phoneme like unit) level word recognition. The recognition of nonnative speakers' pronunciation enables an automatic diagnosis and an error detection which are the core of English pronunciation tutoring system. The above system needs two pronunciation dictionaries. One is for representing standard English pronunciation. The other is for representing Korean speakers' English Pronunciation. Both dictionaries are integrated to generate pronunciation networks for variants.

I. 서론

1-1. 영어 발음교정시스템

영어 발음교정시스템은 한국어를 모국어로 하는 화자들의 영어 발음을 교정하기 위한 것이다. 이 시스템은 다음 두 가지 하위 부문을 필요로 한다. 첫째, 표준 영어 발음을 모방하는 한국 화자들의 발음을 진단하고 평가하는 부문이 필요하다. 둘째, 한국 화자의 발음을 진단하고 평가한 결과에 따라 최적의 교정정보를 제공하는 부문이 필요하다. 본고의 영어 발음교정시스템은 위의 첫 번째 부문을 위한 방법론으로 음성인식을 채택하였다. 본고는 음성인식을 이용하여 오류 유형을 자동으로 검출하는데 필요한 표준 발음 사전 및 오류 발음사전 구축에 관한 내용으로 이루어져 있다.

1-2. 음성인식의 범위

영어 발음교정 시스템에 이용되는 음성인식의 범위는 유사 음소단위(PLU:Phoneme Like Unit) 수준의 고립 단어 인식에 한정하였다. 한국인의 영어 문장 발화는 원어민의 문장 발화처럼 유창하지 못하기 때문에 문장 내의 단어 사이에서 나타나는 양상이 원어민의 경우와 다르다. 즉, 원어민 화자들이 끊어 읽는 부분에서 제대로 끊어 읽지 못할 수도 있고 음소의 삽입, 삭제 등으로 인해 원어민 화자의 발화와 다른 음운 현상이 나타나기도 한다[1]. 이러한 이유로 문장 단위에서 유사 음소의 오류를 자동으로 검출하여 발음교정시스템을 구성할 수 있을 정도의 인식률을 얻기는 어렵다[2]. 따라서 영어 발음교정시스템의 오류 자동 검출의 범위를 고립단어에서 나타나는 유사 음소 수준의 오류에 한정하였다.

1-3. 발음사전 및 음소네트워크

영어 발음교정시스템을 위한 발음사전은 두 종류이다. 첫 번째는 영어 철자에 영어 음운 규칙을 적용한 표준 영어발음사전이다. 두 번째는 표준영어발음에 한국어 음운 목록과 음절구조의 영향에 의한 변이를 포함한 한국인 영어발음사전이다. 각각의 사전을 구성하기 위해서는 표준영어 음소셋과 한국인의 영어 발음 음소셋이 필요하다. II와 III에서는 각각의 사전을 구축하기 위해 제작된 음성 DB와 음성 DB의 분석을 통해 도출해낸 여러 가지 변이음들을 포함한 음소셋의 목록을 알아보도록 한다. 그리고 IV에서는 II와 III에서 구축된 발음사전을 통합하여 한국 화자의 발음을 입력으로 받아 오류 유형을 자동으로 검출하기 위해 구성한 음소네트워크에 대해서 알아보았다. 그리고 V에서는 본 연구의 의의와 향후 과제에 관해서 살펴보았다.

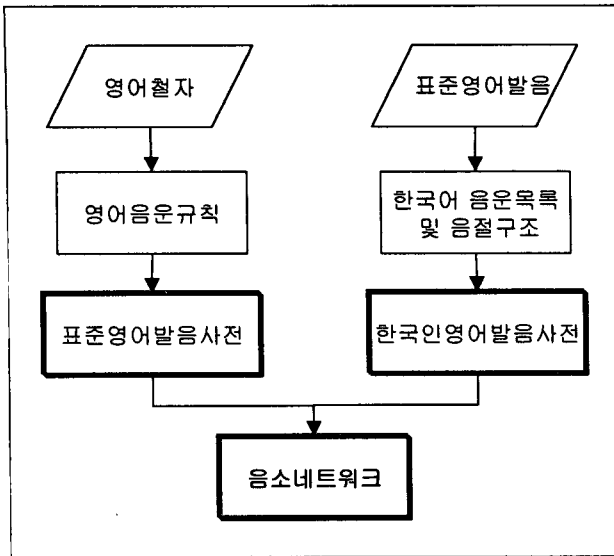


그림 1. 통합된 발음사전과 음소 네트워크

II. 표준 영어 발음사전

2-1. 음성 DB

초등학교 5, 6학년과 같은 연령대의 미국 원어민 화자 254명(남: 121명, 여: 133명)에게 2,145단어를 500여개로 나누어서 10회씩 발화하도록 하였다. 사무실 환경에서 16kHz/16bit로 핸드마이크와 Labtec 헤드셋을 사용하여 2채널로 녹음하였다.

위의 254명 분의 자료 가운데에서 59명 분(남: 28명, 여: 31명)의 자료가 표준 영어 발음사전을 위한 변이음 규칙 선정에 이용되었다.

2-2. 폰셋(Phone set)

고립단어 인식은 관측열 O 가 주어졌을 때 단어의 확률이 최대가 되는 단어를 찾는 작업이다.

$$(1) w' = \text{argmax}_i P(wi|O)$$

위의 확률은 Bayesian 규칙에 의하여 다음과 같이 표현된다.

$$(2) P(wi|O) = P(O|wi)P(wi)$$

즉, 단어의 인식 확률은 음향모델($P(O|wi)$)과 언어모델($P(wi)$)의 곱으로 이루어진다. 단어의 인식률을 높이기 위해서는 인식하고자 하는 단어에 대해 한 가지의 발음형을 표시하는 것 보다는 가능한 발음형을 모두 표시하는 것이 효과적이다[1][2].

표준영어발음 사전의 구축을 위해서 본고에서는 공개된 CMU(Carnegie Mellon University) 사전(v.0.6)을 목적에 맞게 가공하였다. CMU 폰셋은 총 39개의 음소로 이루어져 있으며 강세의 차이에 따라 모음을 따로 구분하지 않았다. CMU 발음사전은 39개의 음소셋(자

음 24개, 모음 15개)을 선정하고 있다. 이 39개의 폰셋에는 변이음을 나타내는 폰셋이 포함되지 않았으므로 여기에 다음과 같이 단어 내의 변이음 규칙을 적용시킨 후에 나타나는 변이음까지 포함하기 위해 16개의 폰셋을 추가하였다. 다음은 본고에서 고려한 변이음 규칙의 일부이다[3].

- (1) 무성파열음 /p,t,k/는 음절 초 /s/다음에서는 무기음화된다. 변이음 : [p=, t=, k=]
- (2) 치조음 /d,t/는 모음 사이에 오며 두번째 모음에 강세가 없을 때 voiced tap이 되기도 한다. 변이음 : [r]
- (3) 설측음 /l/은 모음 뒤나 단어 끝의 자음 앞에서 연구개음화된다. 변이음: [ɫ]

모음의 경우 CMU의 15개 모음에 강세 구분을 주어 총 30개의 폰셋을 정의하였다.

III. 한국인 화자의 영어 발음사전

3-1. 음성 DB

1) 화자 및 인원 수

초등학교 5, 6학년 411명(남 163 /여 248)에게 163 단어를 2회씩 발화시켜 얻은 데이터 가운데 70 명(남 22 /여 48) 자료를 음향 분석하였다.

2) 분석대상 자료 수

한국 학생 120명 분의 자료 가운데 녹음 상태가 좋지 못한 것 등을 제외하고 최종 선택한 단어 갯수는 20160개이다(남 8400 단어/ 여 11760 단어).

3) 검출한 오류의 개수

20160 개 단어를 2명의 음성학 전문가가 수작업으로 분절하였고 그 가운데 17460 개의 음소는 두 명의 작업 결과를 교차시켜 최종 확인한 것이다. 다음은 교차 확인을 마친 17460 개의 음소를 자음과 모음으로 나누어 정오 빈도 및 오류 유형별 빈도를 분석한 것이다.

4) 오류 빈도

다음은 한국 화자의 영어 단어 발음을 분석하여 표준 영어의 음소와 비교한 결과의 일부이다.

표 1. 자음 오류 빈도

음소	정/오(대표오류)	음소 개수
f	정	197
	오([p])	47
v	정	190
	오([b], [b])	53
s	정	712
	오([ʃ])	23
z	정	100
	오([ʒ])	132

표 2. 모음 오류 빈도

음소	정/오(대표오류)	음소 개수
i	정	461
	오([i:])	118
i:	정	470
	오([i])	30
e	정	592
	오([ei])	82
æ	정	216
	오([e])	270

5) 오류 유형[4][5]

(1) 대치

대치 유형은 표준 영어음소 대신 표준 영어음소와 가장 가까운 한국어 음소로 발음하는 경우를 말한다. 영어의 순치마찰음 /f, v/는 양순파열음 /p, b/로 흔히 대치되었고 치간무성마찰음 /θ/는 치조마찰음 /s/나 치조파열음 /t'/ 등으로 대치되었다. 치간유성마찰음 /ð/는 치조 파열음 /d/ 또는 우리말 /ㄷ/ 등으로 대치되는 예가 많았다.

그리고 영어의 유성파열음 /b, d, g/가 우리말 /ㅂ, ㄷ, ㄱ/로 흔히 대치되었다. 우리말 /ㅂ, ㄷ, ㄱ/는 유성음 사이에서는 유성화가 이루어지지만 영어의 /b, d, g/와는 다르다. 무성파열음 /p, t, k/가 주로 /s/다음에 올 때 기식이 거의 없는 소리가 되는데 한국어 화자들은 많은 경우 기식이 있는 소리로 그냥 발음하였다. 또 모음의 경우 전설모음 /i:/와 /u/는 흔히 혼동되었고 후설모음 /u:/와 /u/도 흔히 혼동되었다. 우리말 /예, 애/의 구분이 모호해지는 특성이 영어에도 그대로 반영되어 /æ/가 흔히 /e/로 대치되었다. 이중모음 가운데 /ou, ei/ 등은 많은 경우 /o, e/처럼 단순모음으로 대치되기도 하였다.

(2) 삼입

영어는 한국어와 달리 음절머리 또는 음절꼬리에 자음군을 허용하고 있다. 'strike'라는 단어는 한국인이 발음하면 흔히 [스트라이크]가 된다. 음절머리에 /str/처럼 세 개의 자음이 연이어 오는 현상이 한국어에 없기 때문에 [으] 모음이 흔히 삼입되는 것이다.

그리고 'sports'와 같이 음절꼬리에 오는 자음군에도 [으]를 삼입하여 [스포츠]처럼 흔히 발음한다.

(3) 삭제

단어 끝에서 자음 [r]이 삭제되는 경우가 있었으며, /q/ 다음에 오는 반자음 [w]이 삭제되는 경우도 있었다.

3-2. 폰셋

위에서 나타난 대치, 삼입, 삭제 가운데에서 대치와 삼

입 유형을 발음사전에 표기하기 위해서 한국어 음소 목록을 추가하였다. 추가된 한국어 음소 목록은 다음과 같다.

- (1)/ㅂ, ㄷ, ㄱ/ : /b, d, g/와 흔히 대치되는 한국어 음소
- (2)/ㅅ/ : /s/와 흔히 대치되는 한국어 음소
- (3)/ㅈ/ : /z, dʒ/와 흔히 대치되는 한국어 음소
- (3)/으/ : 자음군에서 흔히 삼입되는 음소

이외에 위의 한국어 음소 목록과는 별개로 표준 영어 발음을 잘 모방한 한국 화자의 발음이 원어민 화자의 발음과 얼마나 유사한지를 평가하기 위해 한국 화자의 발음은 표준 영어음소에 구분기호를 넣어 이후의 분석에 이용되도록 하였다.

IV. 음소 네트워크

위에서 각각 구성된 표준발음 사전과 오류 발음사전은 통합되어 다음과 같은 음소 네트워크를 구성하게 된다 [6][7][8].

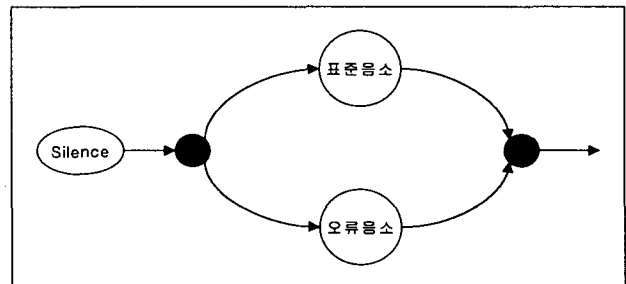


그림 2. 음소 대치 네트워크

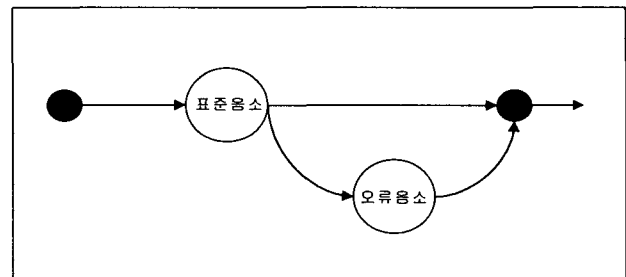


그림 3. 음소 삼입 네트워크

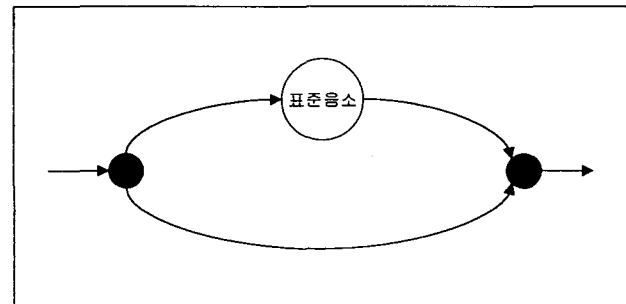


그림 4. 음소 삭제 네트워크

V. 결론

이상에서는 영어발음교정 시스템 개발을 위한 발음구축에 관해서 살펴보았다. 발음교정의 출발점은 표준 발음과 사용자 발음의 차이를 추출하는데 있다. 표준 발음은 기존의 발음 사전을 토대로 다양한 원어민 화자의 발음을 분석하여 용인될 수 있는 변이를 확정지음으로써 구성되었다. 오류 발음사전은 교정의 대상이 되는 화자들의 발음을 분석하여 발생 가능한 오류를 발음사전에 포함시키는 것으로 구성되었다. 발음교정의 정확도는 바로 후자의 오류 발음사전을 얼마나 정확하게 구성했는가에 달려있다고 할 수 있다.

이상과 같은 단어 수준의 발음교정 방법론 일부는 구단위의 발음교정으로 확장될 수 있다. 발음교정의 궁극의 목표는 영어로 원활한 의사소통을 하는데 있으므로 단어 단위 이상의 발음교정에 관한 연구가 계속 이루어져야 한다.

참고문헌

- [1] Cremelie Nick and Jean-Pierre Martens, "On the use of pronunciation rules for improved word recognition," in *Proceedings Eurospeech-95*, pp. 1747-1750, 1995
- [2] ISLE D3.3 "Recognition of Learner Speech," 1999
- [3] Ladefoged, P., *A course in phonetics(4th edition)*, Harcourt College Publishers, 2001
- [4] 김효숙, "한국인을 위한 영어발음교정 시스템 'Dr.Speaking' 소개," *대한음성학회 창립25주년 기념 학술대회*, pp. 47-50, 2002
- [5] Bonaventura, Patrizia et al., "Phonetic annotation of a non-native speech corpus," in *Proceedings Workshop Integrating Speech Technology in the (Language) Learning and Assistive Interface*, pp. 10-17, Dundee (UK), InStil 2000
- [6] Sproat, Richard, "Pmtools: A pronunciation modeling toolkit," in *Proceedings of the Fourth ISCA Tutorial and Research Workshop on Speech Synthesis*, Blair Atholl, Scotland, 2001.
- [7] Pastor, M. and F. Casacuberta, "Automatic learning of finite state automata for pronunciation modeling," in *Proceedings Eurospeech2001*
- [8] E. Fosler-Lussier and G. Williams, "Not just what, but also when: Guided automatic pronunciation modeling for Broadcast News," In *DARPA Broadcast News Workshop*, Herndon, Virginia, March 1999.