음성인식기를 이용한 발음오류 자동분류 결과 분석

강효원* 이상필* 배민영* 이재강** 권철홍*

* 대전대학교 정보통신공학과
** 대전대학교 일어일문학과

Performance Analysis of Automatic Mispronunciation Detection Using Speech Recognizer

Hyowon Kang*, Sangpil Lee*, Minyoung Bae*, Jaekang Lee** and Chulhong Kwon*

* Dept. of Information and Communications Eng., Daejeon Univ.
** Dept. of Japanese Language and Literature, Daejeon Univ.

kanghyowon@hotmail.com, nishot@hanmail.net, missbea97@hotmail.com,
ljgang@dju.ac.kr, chkwon@dju.ac.kr

Abstract

This paper proposes an automatic pronunciation correction system which provides users with correction guidelines for each pronunciation error. For this purpose, we develop an HMM speech recognizer which automatically classifies pronunciation errors when Korean speaks foreign language. And, we collect speech database of native and nonnative speakers using phonetically balanced word lists. We perform analysis of mispronunciation types from the experiment of automatic mispronunciation detection using speech recognizer.

Ⅰ. 서론

본 연구는 국제화 시대에 있어서 외국어 발화의 자동 발음 교정 시스템을 개발하여 발음 교정의 기회를 충분히 제공할 토대를 마련하는 것을 목표로 하고 있다. 이러한 발화 양상의 교정은 극소수의 언어·음성학 전문가들에 의해서만 수행되고 있는 형편이며, 아동의 언어발달과 성인의 외국어 습득을 위한 발음교정은 여러 차원에서 교육 및 재교육이 필요하나, 비용과 접근의 불편함 때문에 그 해결의 중요성에도 불구하고 기회 제공이 원활히 이루어져 오고 있지 않다. 이를 위해서 음성인식 기술을 이용하여 발음상의 미비함, 조음기관 동작의 부정확성을 정확히 인식하고 그 각각의 범주에 따른 교정방안을 컴퓨터를 통하여 제시함으로써 자동화 시스템의 이점, 즉 보다 값싸고 보다 편리한 발음 교정 시스템의 개발을 목표로 한다.

기존의 음성인식엔진을 이용한 영어학습 시스템은 정상인들의 발음 모델을 통계적으로 추출하여 미리 제작해 놓은 후, 임의의 발화음성이 입력되었을 경우 발화했으리라 추측되는 가장 유사한 음소로 인식하는 것을 목표로 한다. 따라서 인식율은 얼마나 정상발음과 가깝게 발화했는지의 여부에 의해서만 결정될 뿐이다. 즉, 학습자의 발음과 원어민의 발음을 특징신호 추출을 통한 유사성의 판별을 통해 발음의 정확도만를 제시해 주는 시스템이다. 본 연구의 오류발음 자동 분류 시스템은 발음 교정을 목표로 하고 있으므로, 발음에 따른 오류 양상을 자동으로 분류하여 그 오류에 정확히 해당하는 교정사항을 제공하기 위한 오류 유형별 발화 모델들을 미리 수집하여 HMM 모델을 통해 훈련하여 미리 만들어 놓는다.

본 논문에서는, 한국인이 일본어 발화시 일본어 음소와 혼동하여 발음하는 한국어 음소들을, 음성학 전문가가 분류하여 유사발음 음소 셋을 선정하고, 이 음소 셋에 기반하여 발음오류를 자동으로 분류하는 HMM 음소인식기에 대한 연구를 수행하였다. 본 논문의 구성은, 2장에서 한국어 및 일본어 폰 셋 및 유사발음 음소 셋의 구성에 대하여 설명한다. 3장에서는 오류발음 자동 분류 음성인식기 구현에 대하여, 음성 DB 구축방법, 실험 방법, 실험 결과 등을 중심으로 기

술한다. 그리고 4장에서 결론 및 향후 연구과제에 대하여 기술한다.

Ⅱ. 언어별 폰 셋 및 유사발음 음소 셋의 구성

1. 일본어 폰 셋

일본음향학회에서 선정한 음소 셋을 참조하여[1], 자음 27개, 모음 10개, 반모음 2개, short pause, silence를 포함하여 총 41개의 폰으로 구성하였다[표 1]. 표 1을 보면, 첫 번째 열에 자음이 속하는 음절의 예를 들고, 두 번째 열에 자음의 기호를 보여준다. 표 1에서, ax ~ ox 는 장음 모음을, q는 이중 자음을 의미한다

2. 한국어 폰 셋

자음은 변이음을 고려하여 29개를, 모음은 음성학적 차이를 보이는 모음만을 고려하여 17개를 선정하였고[2], short pause, silence를 포함하여 총 48개의 폰으로 구성하였다[표 2, 표 3]. 표 2를 보면, 첫 번째 열에 모음을, 두 번째 열에 기호를 보여준다. 표 3을 보면, 첫 번째 열에 자음을, 두 번째 열에 기호를 보여주고, 세 번째 열에 그 자음이 속한 단어를 예로 들었다. 표 2와 3에서, 일본어와 한국어 음소 표기의 구분을 위해 한국어 음소 표기는 'c' 로 끝난다.

표 1. 일본어 음소 셋

자음	기호	자음	기호	자음	기호	모음	기호
ぴ	**pi**	つ	**tsu**	も	**mo**	あ	**a**
ぴょ	**pyo**	ちょ	**cho**	みゅ	**myu**	い	**i**
て	**te**	は	**ha**	な	**na**	う	**u**
か	**ka**	ひょ	**hyo**	にょ	**nyo**	え	**e**
きゃ	**kya**	さ	**sa**	ん	**N**	お	**o**
べ	**be**	し	**shi**		**q**	あー	**ax**
びゅ	**byu**	ず	**zu**	ファ	**fa**	いー	**ix**
だ	**da**	じゅ	**ju**	반모음	기호	うー	**ux**
が	**ga**	ら	**ra**	わ	**wa**	えー	**ex**
ぎゃ	**gya**	りょ	**ryo**	や	**ya**	おー	**ox**

표 2. 한국어 모음 셋 (17개)

모음	기호	모음	기호	모음	기호	모음	기호
ㅏ	**axc**	ㅣ	**ixc**	ㅑ	**jac**	ㅖ	**jec**
ㅓ	**eoc**	ㅐ, ㅔ	**aec**	ㅕ	**jeoc**	ㅘ	**wac**
ㅗ	**oxc**			ㅛ	**joc**	ㅝ	**woc**
ㅜ	**uxc**			ㅠ	**juc**	ㅞ, ㅙ, ㅚ	**wec**
ㅡ	**euc**			ㅟ	**wic**	ㅢ	**euic**

표 3. 한국어 자음 셋 (29개)

자음	기호	예제	자음	기호	예제	자음	기호	예제
ㄱ	**kc**	가시	ㅋ	**khc**	칼날	ㅅ	**sc**	사장
	gc	대구	ㄲ	**kkc**	까치		**sic**	쉽다
	kqc	독자	ㅌ	**thc**	타자	ㅆ	**ssc**	쓰다
ㄷ	**tc**	도시	ㄸ	**ttc**	따님	ㅎ	**hc**	하지
	dc	사다	ㅍ	**phc**	파도		**hic**	동해
	tqc	받기	ㅃ	**ppc**	뽑다	ㄹ	**lc**	사슬
ㅂ	**pc**	바지	ㅊ	**chc**	차례		**rc**	다리
	bc	나방	ㅈ	**zc**	자다	ㅇ	**ngc**	강사
	pqc	입다		**zhc**	이제	ㅁ	**mc**	감사
			ㅉ	**zzc**	짜다	ㄴ	**nc**	간소

표 4. 유사발음 음소 셋

일본어 음소	일본어 유사발음 음소	발음오류 한국어 유사음소
p	py, b, by	phc, ppc, pc, bc
py	p, b, by	phc, ppc, pc, bc
b	p, py, by	phc, ppc, pc, bc
by	p, py, b	phc, ppc, pc, bc
t	d, ts, ch	thc, ttc, tc, dc
d	t, ts, ch	thc, ttc, tc, dc
k	ky, g, gy	khc, kkc, kc, gc
ky	k, g, gy	khc, kkc, kc, gc
g	k, ky, gy	khc, kkc, kc, gc
gy	k, ky, g	khc, kkc, kc, gc
h	hy	hc, hic
hy	h	hc, hic
s	sh, z, j	sc, ssc, sic, zc, zhc
sh	s, z, j	sc, ssc, sic, zc, zhc
z	j, s, sh	zc, zhc, sc
j	z, s, sh	zc, zhc, sc
r	ry	rc, lc
ry	r	rc, lc
ts	ch, z, j, t	chc, zc, zhc, thc
ch	ts, z, j, t	chc, zc, zhc, thc
a	ax	axc
i	ix	ixc
u	ux	uxc, euc
e	ex	aex
o	ox	oxc
ax	a	axc
ix	i	ixc
ux	u	uxc, euc
ex	e	aec
ox	o	oxc

3. 유사발음 음소 셋

본 연구에서 수집한 음성 데이터를 음성학 전문가가 청취하여, 한국인이 일본어 발화시 나타나는 음소별 발음오류 양상을 분석하여 유사발음 음소 셋을 선정하였다. 표 4에서 유사발음 음소 셋은, 자음인 경우 일본어 자음 20개와 유사 발음 한국어 자음 23개로 구성되고, 모음인 경우 일본어 모음 10개와 유사발음 한국어 자음 6개로 구성하였다. 표 4의 두 번째 행을 보면, 일본어 [p]를 일본어 [py, b, by]로 발화하거나, 한국어 [phc /ㅍ/, ppc /ㅃ/, pc /무성 ㅂ/, bc /유성 ㅂ/]로 발화하는 오류가 발생함을 의미한다.

Ⅲ. 오류발음 자동분류 음소인식기

1. 음성 DB 구축

1.1 녹음 환경

연구실 수준의 조용한 방에서 1명씩, 동시에 3곳에서 녹음하였다. 발화 데이터를 PC에서 수집하였고, 사운드카드는 Soundblaster Audigy를, 마이크는 SHURE 565SD를 사용하였다.

1.2 훈련 음성 DB

일본어 폰 모델 생성을 위하여 수집한 원어민 일본어 음성 DB는, ATR(일본 자동통역 연구소)에서 작성한 PBW 216개 단어를[3] 고려대학교에서 한국어 연수중인 일본인 70명을 대상으로 녹음하였다. 유사발음 한국어 폰 모델 구성을 위한 한국어 음성 DB는, SITEC에서 작성한 한국어 PBW 452개 단어를[4] 대전대학교 대학생 70명을 대상으로 녹음하였다.

1.3 테스트 음성 DB

오류발음 자동 분류 음성인식기를 테스트하기 위한 음성 DB는, 일본어 전공자 120명과 비전공자 120명 등 한국인 대학생 240명이 일본어 PBW 216개 단어를 발성한 것을 수집하였다.

2. 실험 방법

음성신호를 매 10msec 마다 25msec의 Hamming 창함수를 사용하여 MFCC 계수 39차를 추출하였고, HMM의 구조는 3 state left-to-right continuous HMM을 사용하였다. 언어별 음소인식기에서 가장 나은 성능을 보여주는 실험 방법으로 발음오류 분류 음소인식 실험을 수행하였다. 즉, 특징벡터는 MFCC를 채택하였고, Mixture 수는 10으로 하여, Bootstrap 방식으로 HMM 모델을 훈련하였다. 학습자의 오류 발음을 검출하기 위한 유사발음 음소 셋을 이용한 발음 네트워크는 그림 1과 같다.

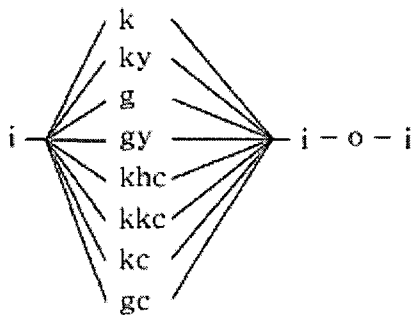

그림 1. 유사발음 음소 셋을 이용한 발음 네트워크
(단어 「いきおい」 [i k i o i] 인 경우)

3. 실험 결과

표 5에서 오류발음을 음성인식기로 자동 분류한 결과를 보여준다.

1) 일본어 (k, t, p)인 경우, 비전공 대학생은 (kkc /ㄲ/, ttc /ㄸ/, ppc /ㅃ/) 등 한국어 경음으로 발음하는 경우가 (khc /ㅋ/, thc /ㅌ/, phc /ㅍ/) 등 한국어 격음으로 발음하는 경우가 보다 약간 많았다. 전공 대학생은 비전공 대학생 보다 경음으로 발음하는 경향이 더 컸다. 이와 같은 실험결과는 일본어 (k, t, p) 발음은 한국어의 경음과 유사한 발음이라는 사실과 일치한다[5].

2) 일본어 (ky, gy, by)인 경우, 전공, 비전공 대학생 모두 인식결과가 (ky, gy, by)로 나타나는 비율이 제일 컸는데, 이는 /y/라는 음가가 존재하기에 다른 한국어 발음과 구분되었기 때문이라고 생각된다. 그리고, 전공 대학생의 실험결과 (ky, gy, by)로 인식되는 비율(40%, 47%, 74%)이 비전공 대학생의 비율(28%, 40%, 62%) 보다 더 큰 사실로부터 전공 대학생의 발음이 보다 일본어에 가깝다는 것을 알 수 있다.

3) 일본어 (g, d, b)인 경우, 비전공 대학생은 한국어 유성 자음 (gc /ㄱ/, dc /ㄷ/, bc /ㅂ/)으로 발음하는 경향을 보여 주었고, 전공 대학생은 일본어 (g, d, b)와 한국어 유성 자음(gc, dc, bc) 발음이 비슷한 분포를 보여 주었다.

4) 일본어 모음 (a, i, e, o)인 경우, 비전공 대학생은 한국어 모음 (axc /ㅏ/, ixc /ㅣ/, aec /ㅐ,ㅔ/, oxc /ㅗ/)으로 발음하는 경향을, 전공 대학생은 일본어 모음 발음과 한국어 모음 발음이 비슷한 분포를 나타냈다. 그리고, 일본어 모음 (u)인 경우, 비전공 대학생은 한국어 모음 (uxc /ㅜ/), (euc /ㅡ/), 일본어 모음 (u) 순으로, 전공 대학생은 일본어 모음 (u), 한국어 모음 (euc), (uxc) 순으로 발음하는 경향을 보여 주었다. 이와 같은 실험결과는 일본어 모음 (u)가 한국어 모음 (ㅜ)와 (ㅡ)의 중간 발음이라는 사실과 일치한다[5].

5) 일본어 장음 모음 (ax, ix, ux, ex, ox)은 비전공 대학생인 경우 한국어 모음 (axc, ixc, uxc, exc, oxc)으로 발음하는 경향이 가장 컸고, 그 다음으로 일본어 장음 모음으로 발음하는 비율이 일본어 단음 모음으로 발음하는 비율 보다 더 커서, 비전공 대학생도 일본어 모음의 장, 단음을 구분한다는 사실을 알 수 있다. 전공 대학생은 일본어 장음 모음으로 발음하는 비율이 한국어 모음으로 발음하는 비율 보다 더 큰 사실을 발견할 수 있어, 보다 일본어에 가까운 발음을 한다는 사실을 알 수 있다.

1)에서의 '실험결과가 일본어 (k, t, p) 발음은 한국어의 경음과 유사한 발음이라는 사실과 일치'와, 2)에서의 '실험결과가 일본어 모음 (u)가 한국어 모음 (ㅜ)와 (ㅡ)의 중간 발음이라는 사실과 일치'라는 결론, 그리고 1), 2), 4), 5)의 실험 결과에서 보듯이 비전공 대학생의 발음에 비해 전공 대학생의 발음을 보다 일본어 음소에 가깝게 인식하는 등, 본 연구에서 구현한 유사발음 자동 분류 음성인식기의 자동 분류가 제대로 수행되었음을 알 수 있다.

Ⅳ. 결론

본 과제에서 음성인식기의 역할은 학습자의 발음 품질을 평가하여 표준 및 오류 발음을 자동으로 분류하는 것이다. 현재까지의 연구 결과 자동 분류가 어느 정도의 성과를 보여 주었다고 판단할 수 있으나, 차년도 연구에서는 이의 정확도를 높여 자동분류 결과와 음성학 전문가의 청취판단을 좀 더 일치시키는 연구를 수행할 계획이다. 이를 위하여, 오류발음의 자동 검출을 위한 발음 네트워크의 구성에 관한 연구와, 학습자의 발음에 대한 자동 스코어링(scoring)에 관한 연구를 진행할 계획이다.

참고문헌

[1]. T. Kawahara et al., "Sharable software repository for Japanese large vocabulary continuous speech recognition", *Proc. of ICSLP* 98, pp. 3257-3260, Sydney, 1998. 11

[2] 신지영, "우리말 소리의 이해", *대한음성학회 창립25주년기념 학술대회발표논문집*, pp. 15-23, 2002. 11

[3] ATR, Spoken Language Translation Research Laboratories, http://www.slt.atr.co.jp

[4] SITEC, 음성정보기술산업지원센터, http://www.sitec.or.kr

[5] 민광준, *일본어 음성학 입문*, 건국대학교출판부, 2002

표 5. 발음오류 분류 음소인식기 실험결과 (%)

	비전공 한국인			전공 한국인		
k	k(7)	kkc(39)	khc(34)	k(16)	kkc(39)	khc(23)
ky	ky(28)	kc(29)	khc(19)	ky(40)	kc(21)	khc(20)
g	g(19)	gc(46)	kkc(13)	g(37)	gc(32)	kc(8)
gy	gy(40)	gc(20)	ky(15)	gy(47)	ky(16)	gc(14)
t	t(14)	ttc(46)	thc(19)	t(19)	ttc(47)	thc(18)
d	d(16)	dc(37)	ttc(19)	d(26)	dc(29)	t(16)
p	p(10)	ppc(34)	phc(29)	p(22)	ppc(36)	phc(26)
py	py(13)	phc(73)	ppc(5)	py(27)	phc(54)	ppc(9)
b	b(19)	bc(51)	pc(10)	b(32)	bc(39)	p(8)
by	by(62)	bc(15)	py(5)	by(74)	py(8)	bc(6)
ts	ts(18)	chc(47)	zc(22)	ts(39)	chc(22)	ch(15)
ch	ch(18)	chc(53)	zc(15)	ch(31)	chc(40)	zc(11)
h	h(26)	hc(37)	hic(30)	h(33)	hic(32)	hc(30)
hy	hy(43)	hc(48)	hic(6)	hy(61)	hc(33)	hic(4)
s	s(13)	sc(44)	ssc(36)	s(24)	sc(43)	zc(22)
sh	sh(23)	sc(59)	zc(8)	sh(38)	sc(49)	zc(6)
z	z(19)	zhc(47)	zc(21)	z(35)	zhc(27)	zc(15)
j	j(14)	zhc(42)	zc(33)	j(25)	zhc(32)	zc(26)
r	r(25)	rc(71)	lc(2)	r(42)	rc(53)	ry(2)
ry	ry(62)	rc(29)	r(7)	ry(79)	rc(14)	r(5)
a	a(27)	axc(67)	ax(4)	a(46)	axc(47)	ax(6)
i	i(27)	ixc(65)	ix(6)	i(43)	ixc(43)	ix(12)
u	u(24)	uxc(43)	euc(30)	u(39)	euc(31)	uxc(25)
e	e(30)	aec(63)	ex(5)	e(43)	aec(41)	ex(9)
o	o(25)	oxc(65)	ox(8)	o(40)	oxc(46)	ox(12)
ax	ax(24)	axc(64)	a(11)	ax(38)	axc(42)	a(19)
ix	ix(25)	ixc(64)	i(10)	ix(55)	ixc(39)	i(4)
ux	ux(29)	uxc(47)	euc(17)	ux(55)	euc(21)	uxc(15)
ex	ex(27)	aec(59)	e(12)	ex(44)	aec(41)	e(13)