공학 석사 학위논문

Corpus-Based Analysis and Pronunciation Modeling of Korean Produced by Chinese Learners

February 2016

Graduate School of

Seoul National University

Interdisciplinary Program in Cognitive Sciences

Seung Hee Yang

Corpus-Based Analysis and Pronunciation Modeling of Korean Produced by Chinese Learners

지도 교수 정 민 화

이 논문을 공학석사 학위논문으로 제출함
2015년 12월

서울대학교 대학원
협동과정 인지과학전공
양 승 희

양승희의 공학석사 학위논문을 인준함
2016년 1월

위 원 장 ______장 병 탁______ (인)

부위원장 ______정 민 화______ (인)

위 원 ______이 호 영______ (인)

Abstract

Learners of second foreign language stand to benefit greatly from Computer-Assisted Pronunciation Training (CAPT) systems which can offer automatic mispronunciation detection and individualized corrective feedback. The increasing demand for learning Korean as a foreign language yields a strong need for a Korean CAPT system development. However, recognition accuracy for non-native speech is often too low to make practical use of Automatic Speech Recognition (ASR) technology in language learning interfaces, and there is limited research on Koran pronunciation produced by non-natives.

As a preliminary research towards developing a CAPT system for Mandarin Chinese learners of Korean, the first part of the study surveys major agreements and disagreements among related works. By conducting a corpus-based experiment, these disagreements are resolved and segmental variations patterns are analyzed, in which flap sounds show the highest variation rate of 35%.

The second part of this study discusses quantitative modeling of these variation patterns for adapting Korean ASR system to Chinese learners. Using context-dependent variation rules describing substitutions, insertions, deletions, and phonological knowledge, extended pronunciation dictionary is generated. With the proposed approach, 21.2% and 1.3% relative WER reduction is obtained from consonantal and vocalic models. The result verifies the corpus-based variation pattern analysis. This study lays the groundwork for Korean CAPT system for various linguistic backgrounds.

……………………………………………

keywords : Computer-Assisted Pronunciation Teaching (CAPT), Korean Produced by Chinese Learners, Non-native speech recognition, Pronunciation Variation Modeling

***Student Number* : 2014 - 20137**

Contents

List of Tables

List of Figures

1. Introduction

An increasing number of people are learning Korean as a foreign language (L2), but only few researches exist in CALL (Computer-Assisted Language Learning) targeting Korean. Statistics[1] as of November 2015 show that 61,940 out of 102,117, which is more than 60%, of foreign students studying in Korea are Chinese students, which cover more than 60%. The second largest group of foreign students is 6,953 from Vietnam, followed by 4,358 from Mongolia. The statistics indicate the increasing demand for Korean language learning, especially by learners whose mother tongue (L1) is Chinese.

Since communication ability is a major purpose of learning a foreign language, precision and fluency in speaking are important goals for language learning. As described in Figure 1, a particular application of CALL named Computer-Aided Pronunciation Training (CAPT) uses Automatic Speech Recognition (ASR) technology to detect mispronunciation, pinpointing where pronunciation error occurs, and provide corrective feedback.

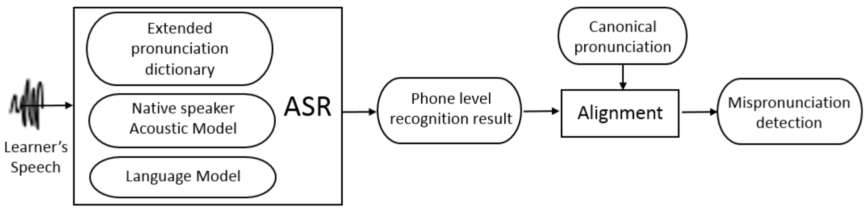
CAPT system benefits learners in various aspects and is to be distinguished from conventional pronunciation teaching methods. While native instructor's feedback in an 1:1 classroom environment is conventional to acquire spoken proficiency, the number of available Korean teachers is too low to meet the learners' needs. This explains why CAPT systems that assist

[1] Published by Statistics Korea http://kostat.go.kr/

or substitute human tutors has been attracting considerable attention in recent years [1-5], which motivates this research. Not only as a substitute of tutors, a CAPT software is also able to offer individualized 1:1 tutoring regardless of constraints in time and place, maximizing learning opportunities at learners' convenience.

Moreover, the feedback generation in CAPT system has an advantage in that it can provide L1-specifc individualized feedback, whereas the instructor in conventional classroom environment cannot address and is not necessarily aware of L1 diversities. This is an important advantage because learners' mother tongue influences the target language production in foreign language learning, indicating that different types of feedbacks are required for each L1, and addressing the L1 influence is a strength of a CAPT system.

Figure 1. CAPT system architecture using ASR technology to automatically detect learner's mispronunciation



1.1. Purpose of Research

The previous section showed that despite the growing interest in learning Korean as a foreign language, no automatic tutoring system using ASR technology has been seriously considered or developed. For English, however, there has been extensive amount of research effort for automatic evaluation of speaking tests. For instance Versant™ English Test [2] is a popular commercialized application and SpeechRater from ETS [3] is an automatic pronunciation scoring with proven its efficiency and accuracy in TOEIC® and TOEFL® tests and is now widely applied in real test conditions [4].

There are possible reasons why CAPT system development for Korean remains under-researched despite the growing demand. Building a CAPT system requires technologies for non-native speech recognition and mispronunciation detection, which remain challenging tasks even with the state-of-the-art technology. First, for the recognition task, pronunciation variations in non-native speech are far more diverse than those observed in native speech [1]. These variations lower the recognition performances for non-native speech. The acoustic-phonetic spaces of native and non-native speakers can be quite difficult and the mismatch causes confusion in speech recognition [5]. Pronunciation and acoustic modeling methods are two major approaches to overcome the challenge, which will be explored in more detail in Chapter 4. Towards this effort, previous studies have analyzed segmental and context-dependent variation patterns in non-native speech [6,7,8].

Second, there are number of issues in mispronunciation detection task. For example, mispronunciation patterns by non-native speakers are unpredictable and poses a challenge in comparing learner's speech with the canonical pronunciation to capture where the mispronunciation occurred. Moreover, resources in non-native speech remain limited to sufficiently model mispronunciation patterns. Related researches have refined conventional GMM-HMM based approach by the DNN training and enhanced the goodness of pronunciation algorithm to minimize mispronunciation detection errors [9].

These twofold challenges involved in CAPT system explain why there has been little research and success in building a system in Korean, while at the same time, motivates this study.

1.2. Research Scope

As a preliminary research towards developing a CAPT system for Korean, this study covers corpus-based discovery of variation patterns and speech recognition modeling. The L1 and L2 languages will be restricted to Mandarin Chinese and Korean.

It is expected that the improved recognition results in this study will enable better mispronunciation detection in future works, and that the methodology employed in this study can be expanded for other L1 languages in CAPT softwares.

1.3. Outline of Research

Chapter 2 surveys previous studies to find the agreements and disagreements among them. It also analyzes Korean and Chinese phonetic inventories. The Chapter 3 discusses corpus-based experiment method and results to identify prominent segmental variation patterns that occur in Korean spoken by Chinese learners. Findings of Chapter 3 are modeled in Chapter 4 in order to obtain improved speech recognition performance. The modeling approach and its method will be described, followed by proposing how the knowledge could be utilized to give corrective feedback to the learners in future work.

2. Survey of Related Works

Learners' L1 and the target L2 both influence L2 production, and analyzing the two phonetic inventories can help predict variation patterns. Contrastive analysis approach assumes that L1 influences foreign language learning and therefore, comparing its similarities and differences with L2 will help predict the learners' variation patterns. Taking a different approach, Speech Learning Model (SLM) [10], claims that the phonemes in L2 that are relatively similar to L1 phonemes are more difficult to acquire than the phonemes that are perceived as more obviously "different" by the learners.

Many previous studies have used a contrastive analysis at the segmental and supra-segmental level to predict what variation patterns will be frequent for Chinese learners of Korean. The following section compares the phonemic inventories and syllable structure of Korean and Chinese, which provides grounds for predicting pronunciation variations in Korean segments produced by Chinese learners. Then, it surveys the predictions made by previous studies. This will be followed by comparing experiment findings in these studies, summarizing the agreements and disagreements among them.

2.1. Contrastive Analysis of Chinese and Korean

Tables 1 and 2 show Korean and Chinese consonants. The Chinese language discussed in this paper refers to Mandarin Chinese. There are 19 phonemes in Korean consonants excluding the approximants /w, j, ɰ / [11], and 19 in

Chinese consonants.

The stops and affricates in Korean can be grouped into lenis, fortis, and aspirated, while in Chinese, they are grouped into aspirated and unaspirated distinctions. The lenis stops /b̥, d̥, g̊/ and lenis affricate /d̥ʑ/ in Korean are slightly aspirated, while the aspirated stops /p^h, t^h, k^h/ and affricate /tɕ^h/ are heavily aspirated. The fortis stops /p$^=$, t$^=$, k$^=$/ and affricate /tɕ$^=$/ are laryngealized and not aspirated. Chinese affricates /tʂ, tʂ^h, ts, ts^h/ do not exist in the Korean counterpart.

Fricatives are grouped into lenis and fortis in Korean, while they are grouped into aspirated and unaspirated in Chinese. The post-alveolar fricative /ʂ/ and dental fricative /f/ do not exist in Korean. In approximants, Korean has /w/, /j/, and /ɰ/ semi-vowels, which do not count as individual phonemes in Chinese.

Table 1. Korean Consonants [11]

	bilabial	alveolar	palatal	velar	Glottal
Stop	b̥ p^h p$^=$	d̥ t^h t$^=$		g̊ k^h k$^=$	
affricate			d̥ʑ tɕ^h tɕ$^=$		
Fricative		s s$^=$			h
Nasal	m	n		ŋ	
Liquid		l			
semi-vowel	w		j	w ɰ	

Table 2. Chinese Consonants [12]

	Bilabial	labio-dental	dental	post-alveolar	velar
Stop	p p^h		t t^h		k k^h
Affricate			ts ts^h	tʂ tʂ^h	
Fricative		f	s	ʂ	x
Nasal	m		n		ŋ
Liquid			l		
approximant				ɹ	

For vowels, there are eight phonemes in Korean vowels[2] and five in Chinese monophthongs. Tables 3 and 4 show Korean and Chinese vowel inventories. The two inventories share /i/, /u/, and /a/ in common. While /ɯ/, /e/, /o/, and /ɛ/ sounds of Korean do not exist in Chinese at phonemic level, /ə/ sound of Chinese does not exist in Korean. It should be noted that contrastive analysis at allophonic level may yield slightly different results. However, allophones are realized or not realized in certain contexts, and thus not considered for variation prediction purpose.

Table 3. Korean Vowels [11]

	Front	Central	Back
Close	i		ɯ u
Close-mid	e		o
Open-mid	ɛ		
Open		a	ʌ

Table 4. Chinese Vowels [12]

	Front	Central	Back
Close	i y		u
Mid		ə	
Open	a		

A syllable in Korean is composed of (C)V(C), a consonant in the onset, a monophthong in the nucleus, and a consonant in the coda. The onset and coda consonants are optional. A syllable in Chinese is composed of initials and finals, and the former is composed of an optional consonant and the latter is composed of a monophthong or a diphthong, followed by an optional /n/,/ɻ/ or /ŋ/. The differences in syllable structures show that /n/ and /ŋ/ are the only

[2] The distinction between /e/ and / ɛ / are becoming neutralized, especially among Korean youth population.

consonants that can be realized as the syllable coda in Chinese, whereas a Korean syllable allows /g̊, n, d̥, l, m, b̥, ŋ/ as the syllable coda.

2.2. Survey of Previous Studies

Previous studies have conducted experiments to check their predictions on pronunciation variation patterns occurring in Korean produced by Chinese learners. These experiments were conducted by analyzing read speech that is composed of words [13, 14, 15] or sentences [16] from Korean as a foreign language textbooks for Chinese learners of beginner [14], intermediate [13, 14, 15], and advanced [14, 16] levels. These researches come from diverse backgrounds, ranging from linguistics to education studies.

Cui (2002) conducted contrastive analysis of phonemes and allophones of Korean and Chinese, and based on the analysis, proposed effective teaching methods [17]. Cui (2004) undertook an acoustic phonetics approach for the contrastive analysis of 24 Korean and Chinese consonant phonemes and allophones [18]. This study gave an detailed analysis on phonemic relation between the two languages supported by spectrogram comparisons. However, the spectrogram analysis in this study is an observation from few individuals' tendencies and thus not representative. Moreover, a detailed diagnosis of mispronunciations should precede teaching method proposal, which means that the teaching methods in this study needs to be revised.

Han et al. (2003) [19] conducted contrastive analysis between Chinese, Japanese, English and Korean, described intuitive observations from teachers on L1-specific learner tendencies, and suggested teaching methods. Although these comments from teachers are useful resources, they are case-specific, rely on auditory impression, and do not consider the diversity across teachers and students.

Qin (2010) predicted consonantal and vocalic mispronunciations based on contrastive analysis of Chinese and Korean. In order to verify the predictions, learners' read speech recordings were evaluated. Variation rates per phonemes were presented and were organized into different types. Results showed that alveolar fricative and front middle vowels are the most difficult phonemes. Hwang (2012) also conducted recording and evaluation experiments and added listening experiment results to study the correlation among the results. Both studies used their findings to make suggestions on effective teaching methods.

These experiments contribute to previous studies by quantifying thirty (Qin, 2010) and sixty (Hwang, 2012) learners' mispronunciations with experiment data. However, the experiments were not phonemically balanced. That is, all vowels, liquids, and nasals have not been evaluated for Hwang and none of them included closed syllables. This is a limitation not only because Korean is both open and closed-syllable language that allows seven consonant endings, but also because rich phonological rules apply from the consonant endings, which influence the pronunciation of next onset.

Leng (2014) and Kim (2008) conducted experiments covering phonemic contexts in all possible positions in the syllable. These experiments, however, remain limited in the number and scope of participants. In Leng (2014), advanced level learners were tested and almost 90% of the 17 participants were female. The participants in Kim (2008) were all intermediate level and only twelve were tested. Although these studies have larger coverage of phonemes than the others, their experiment data is limited and proficiency levels and gender are not balanced to be able to generalize the results as a generic conclusion.

Many of the experiments in previous studies have been followed by suggestions on teaching method proposal. Cho (2013) surveyed the methods and evaluated effectiveness in a classroom environment [20]. It discovered that in practice, learners' acquisition of affricates and lenition rules are different from what theories say. However, there is a room for improvement in this experiment since it does not rely on a reliable diagnosis of mispronunciation patterns.

The paragraphs above summarize the work done by previous studies. In order to identify variation characteristics of Chinese learners of Korean, the experiment results should be quantifiable, and for the results to be generalizable, it should be designed holistically including diverse linguistic aspects of Korean language. Also, its participants should be sizeable and balanced in gender and proficiency levels. Otherwise, the experiment results cannot reveal representative variation patterns. The paragraphs below summarize the predictions and experiment results in these previous works, mainly focusing in agreements and disagreements among the results.

For plosives, many studies agree in their predictions that lenis will be the articulation that learners will find difficult, since there is no lenis equivalent phonemes in Chinese while fortis and aspirated phonemes exist. However, they disagree on how the mispronunciation is realized. While some argued that aspirated phonemes in Chinese are equivalent to Korean lenis, others viewed that fortis phonemes are the equivalent sounds of Korean lenis. Based on this disagreement, the former group found more confusion between lenis and aspirated phonemes, while the latter group found more confusion with fortis and lenis. For example, in case of velar stop in /g̊⁼i/ (air in Korean), the learners find the pronunciation /g̊/ difficult, and when they do, studies disagree whether it is realized as /kʰi/ (height) or /k⁼i/ (talent).

For affricates, studies agreed that lenis will be a difficult pronunciation, and experiment results showed indeed that lenis shows high variation rates. However, it was not clear whether lenis are realized as aspirated or fortis. For example, while the studies agreed that alveolar stop in /ʥ̥ad̥a/ (to sleep) is difficult to pronounce, they did not agree whether the affricate is realized as /ʨʰad̥a/ (being cold) or /ʨ⁼ad̥a/ (being salty).

For fricatives, some studies predicted that since Chinese /s/ and Korean /s⁼/ are corresponding sounds, learners will substitute Chinese /s/ with Korean /s⁼/. However, other studies predicted that learners will substitute Korean /s⁼/ with Chinese /s/. In the experiment results, some confirmed the former while others found substitution in the other direction. For example, whether the learners pronounce /sad̥a/ (to buy) as /s⁼ad̥a/ (to be cheap), or the other way around is not clearly established.

For liquids, the previous studies agree that flap will have tendency to be realized as [l] or [ɹ] due to L1 influence. Their experiments confirmed the prediction. However, experiment results do not show what is more prominent. For example, they agree that the flap in /naɾa/ (country) is difficult to pronounce, but it is not clear if /nala/ (to carry) or /naɹa/ is more likely to occur as the variation. Table 5 summarizes these agreements and disagreements.

Table 5. A survey of predictions and experimental results in previous works

	Agreements	Disagreements
Plosives	Substitutions found in lenis [13,14,15,16,17,19,20]	Lenis realized as fortis [76,20] or aspirated [14,15]
	Substitutions found in fortis [13,14,16,18,19,20]	Fortis realized as lenis [13, 14,18,19,20] or aspirated [15]
Affricates	Substitutions found in lenis [13,15,16,17,20]	Lenis realized as aspirated [13,16,20] or fortis [13]
Fricatives	Substitutions found in lenis [13,14,15,16,19]	Substitutions found in fortis [14]
Liquids	Substitutions found flap [13,15,16,17]	Flap realized as lateral, retroflex [13,15,16,17] or not mentioned [14,20]
Monoph-thong	Substitutions found in /ɰ/ [13,14,15,16,17,19]	None
	/o/ realized as /ʌ/ [13,15,18]	/ʌ/ realized as /o/ [16,17,19], /o/ realized as /u/ [13,18], or else [20]
	Substitutions found in /ɯ/ [13,15,18]	/ɯ/ realized as /ʌ, u/ [18,13,15] or diphthongs [13]
Diphthong	None	Variations in diphthongs mentioned [13,15] or not mentioned [18]
		Realized as other diphthongs [15], or monophthongs [13,15]
Coda	Variation in all final consonants - deletion, and substitution with other consonants [13,14,15,16,17,19,20]	consonant insertion of /n/,/l/ [15]
		Closed syllables showed less variation than open syllables [16]

According to the disagreed aspects, it is not yet clear for example, whether variations in lenis stops are realized more frequently as fortis or aspirated stops. For example, Hwang and Cin both agree that lenis is a difficult pronunciation. However, it is interesting to note the disagreement where 48.3% of lenis in the former study were realized as aspirated and 33.3% as fortis, while 85.5% of lenis in the latter study were realized as fortis and

14.3% as aspirated. This is interesting because both experiments had included the list with the three-way distinctions of lenis, fortis, and aspirated, yet there is a stark disagreement between the two. The number and nature of disagreements pose difficulty in achieving a consensus over salient variation patterns.

Possible causes of disagreement are differences in the purposes of the experiments, types of data used, and research methods used in phonetics and in education fields. For example, some studies have tested intermediate learners only [15,20], while others also included advanced levels [14,16]. Limiting the scope to intermediate level is justified for educational purposes where the end goal is to identify pronunciation errors distinguished from misreading errors that beginners make, and from habitual reading that advanced learners may show. This type of mispronunciation are to be distinguished from mispronunciations caused by articulation difficulties. Therefore these different scopes could also cause disagreements in the results of the corresponding studies. Experimenting with different learner levels could have caused disagreement in the results. Considering the fact that the degree of L1 interference is more likely to be stronger at the beginner level than at the intermediate level, is after the period when learners have begun to receive pronunciation training and have higher awareness of their weaknesses, different learner levels are likely to exhibit different variation patterns.

Moreover, some studies have used a list of monosyllabic meaningless words consisting of stops and affricates only, while others have used lists of carefully designed meaningful words composed of two or three syllables. For example, there is a difference in the experiment design as Hwang's study tested minimal pairs and Cin's study tested meaningful words. In order to test the three-way distinction in bilabial stop /b̥ pʰ p͈/, Hwang's list included /b̥aŋ/, /pʰaŋ/, /p͈aŋ/, which are minimal pairs that included both meaningful and meaningless words, and Cin's list included meaningful words only, such as /b̥inu/, /pʰad̥o/, /ap͈a/. This difference could have caused a disagreement because the segments of interest were occurring in a fixed position for Hwang's study, while they were occurring in phonemic contexts for Cin's study. The following Chapter begins by describing an improved experimental methodology that overcomes such limitations.

3. Segmental Variations Produced by Chinese Learners

3.1. Experiment Design for Corpus-based Analysis

The disagreements shown in Table 5 demonstrate the necessity for a corpus-based approach. A corpus-based approach uses large scale data to transcribe and statistically analyze a representative and data-driven phenomenon. This study proposes to conduct an experiment with a larger number of learners of all levels, consisting of all disagreed aspects, in order to find out the prominent variation patterns and resolve the disagreements. This corpus-based approach not only resolves inconsistencies among previous studies, but also enables us to examine new patterns that were undiscovered in previous studies. This result can be later used to improve speech recognition performance.

The corpus used for this research is L2KSC (L2 Korean Speech Corpus), a speech corpus for Korean as a foreign language spoken by Chinese learners [21]. The corpus was built to evaluate acquisition of phonetic and phonological sounds in Korean language by foreign learners of various L1 backgrounds.

From L2KSC corpus, this study analyzes a 300 words read-speech produced by 53 male and female Chinese learners. The list of words was built for Korean segmental pronunciation learning as a foreign language, based on the vocabulary used in 8 mainstream textbooks. The gender and proficiency, from beginner to advanced levels, are balanced in the distribution.

Transcribers only use the units available in PLU (phone-like unit) set shown in Tables 6 and 7 to transcribe the speech data [22]. Six Chinese phonemes /tʂ/, /ʂ/, /ts/, /tsʰ/, /f/, and /ɹ/ have been added to the Korean PLU set. The data was force-aligned and canonical transcriptions were generated Auditory transcription results marking the differences against the automatically-generated canonical transcriptions was obtained using the extended PLU set. Three graduate students with knowledge in phonetics from the Department of Linguistics of Seoul National University performed the transcription.

In order to confirm that the transcribers transcribed reliably, correction rate and pair-wise agreement are calculated. Correction rate is calculated by dividing the number of corrected phones by the total number of phonemes in forced-aligned result. Pair-wise agreement is calculated by dividing agreement pairs by the sum of agreement and disagreement pairs between annotators.

$$\text{Correction rate} = \frac{No.of\ Corrections}{Total\ No.of\ Phonemes}$$

$$\text{Pairwise Agreement} = \frac{No.of\ Agreement}{Total\ No.of\ Phonemes}$$

Correction rate and pair-wise agreement are 0.143 and 0.86, respectively. Comparing these figures with those of previous studies' rates, 0.105 and 0.881 [23, 24] respectively, we verify the reliability of transcription results in this study.

Table 6. PLU set for Korean sounds

Vowels							
PLU	IPA	Korean	Example	PLU	IPA	Korean	Example
AA	a	ㅏ	아이	AX	ʌ	ㅓ	언니
OW	o	ㅗ	고모	UW	U	ㅜ	구경
IY	i	ㅣ	기술	WW	ɯ	ㅡ	그녀
EH	ɛ	ㅐ	해	EY	E	ㅔ	꽃게
UI	ɥi	ㅟ	복귀	JA	Ja	ㅑ	야구
JX	jʌ	ㅕ	견제	JH	jɛ	ㅒ	얘기
JE	je	ㅖ	계단	JO	Jo	ㅛ	교통
JU	ju	ㅠ	휴식	WA	wa	ㅘ	과일
WH	wɛ	ㅙ	돼지	WX	wʌ	ㅝ	권장
WE	we	ㅞ	궤양	WI	ɰi	ㅢ	의자
Consonants							
PLU	IPA	Korean	Example	PLU	IPA	Korean	Example
K	g̊	ㄱ	가을	KQ	k˺	ㄱ	색동
KK	k⁼	ㄲ	까닭	KH	k^h	ㅋ	칼
T	d̥	ㄷ	다리	TQ	t˺	ㄷ	받고
TT	t⁼	ㄸ	딸	TH	t^h	ㅌ	탈
P	b̥	ㅂ	바람	PQ	p˺	ㅂ	입술
PP	p⁼	ㅃ	빨래	PH	p^h	ㅍ	파리
Z	ʥ̊	ㅈ	자리	ZZ	ʨ⁼	ㅉ	짜임새
CH	$ʨ^h$	ㅊ	처음	HH	H	ㅎ	하늘
S	s	ㅅ	소리	SS	s⁼	ㅆ	싸리
M	m	ㅁ	마음	MM	m˺	ㅁ	감
N	n	ㄴ	나리	NN	n˺	ㄴ	간
NX	ŋ	ㅇ	둥지	L	L	ㄹ	빨래
R	ɾ	ㄹ	소리				

Table 7. Added PLUs for transcription of Korean sounds spoken by Chinese learners

PLU	CS	SH	ZS	ZH	F	RR
IPA	ts^h	ʂ	ts	tʂ	f	ɹ

3.2. Experiment Results

To quantify the relation between the canonical and the actual pronunciation, a confusion matrix is generated. When the realized segment is different from the canonical, it is considered as a variation. The average variation rates for consonants and vowels are 13.74% and 3.35%, respectively. In order to identify salient variation patterns, the phoneme whose variation rate is larger than the average variation rates is considered as a major variation pattern and shown in the Target Segment column of Table 8.

The Realized Target column shows how the overall variation rate is composed For example, we see that 33% of flap /r/ is realized as lateral /l/, 20.2% and 11.9% of the fortis /p꞊/ and /tɕ꞊/ are realized as their lenis counterparts, and 7.9% of the lenis affricate /dʑ̥/is realized with a Chinese phoneme with similar place of articulation. The rates quantifying realized phonemes are meaningful because they can be used to model the likelihood of variation and where they occur, and thus enabling ASR system to better recognize in Chinese learners’ speech.

Table 8. Salient variation patterns of Korean produced by Chinese (del.=deletion)

	Target Segment	Count	Variation Rate (%)	Realized Target & Rate (%)					
Consonant	R	1,604	36.10	l	33.0	n	1.0		
	p͈	942	27.49	b̥	20.2	pʰ	6.9		
	tɕ͈	613	23.49	dʑ̥	11.9	ts	7.2	tɕʰ	2.9
	dʑ̥	1,694	22.67	ts	7.9	tɕ͈	6.9	tɕʰ	3.7
	t̚	1,792	21.94	del.	15.6	k̚	3.8		
	k͈	1,607	21.22	ɡ̊	15.6	kʰ	4.7		
	d̥	3,584	17.75	t͈	14.0	tʰ	3.2		
	t͈	1,840	16.37	d̥	13.1	tʰ	2.8		
	k̚	2,086	14.96	del.	11.0	l	1.4		
Vowel	ɯi	230	47.16	i	30.1	e	7.9	ɥi	3.5
	jɛ	47	25.53	je	19.1	ɛ	6.4		
	wʌ	240	16.31	u	4.7				
	we	474	14.76	ɥi	9.1				
	wɛ	235	10.74	ɛ	3.4				
	ja	570	9.12	a	5.1	jʌ	3.7		
	je	331	9.06	e	4.2				
	ɥi	143	7.69	we	2.8	u	2.1	ɯi	2.1
	ju	239	7.53	ɥi	2.5	u	2.1	jo	2.1
	jʌ	1746	7.1	i	1.8	jo	1.6		

3.3. Comparison with Related Works

The results in this study help resolve the disagreements among previous studies as well as discover new variation patterns. Regarding the disagreements mentioned earlier, quantitative analysis confirms that lenis is more frequently substituted by fortis than aspirated. This verifies contrastive analysis result of Korean and Chinese in Chapter 2 by showing that Chinese aspirated consonants sound more similar to Korean fortis than aspirated consonants. Regarding vowels, variations in diphthongs are more likely to be realized as monophthongs than other diphthongs.

The experiment results not only resolved the disagreements, but also found the following characteristics regarding Korean spoken by Chinese learners. Variations in phonological rule application and under-emphasized substitution patterns will be discussed below.

First, what is noticeably different from the previous studies is that variation rates for fortis is higher than lenis. Indeed, four of the top ten most frequent variations were fortis sounds. However, it should be noted that many of the cases when fortis is realized as lenis are found in words where phonological rules apply. For example, 52% cases of fortis bilabial stop realized as lenis are misapplication or non-application of phonological rules in Korean, depending on whether or not the learner has the knowledge about the rule. For instance, the canonical pronunciation of the bilabial stop in the word candle, /tɕʰot̚p͈ul/, is pronounced as /tɕʰot̚b̥ul/ as it is read according to the underlying form of the word. It is only when the phonological rule to fortify /b/ in front of /t̚/ applies, it is pronounced as fortis. Not only for the bilabial stop but also for other fortis /tɕ͈, k͈, t͈/ sounds, the same misapplication of phonological rule occurred.

Although other studies have not mentioned high variation rates for fortis, which is different from the current experiment result, the difference is not surprising. List of words in previous experiments did not include words in which phonological rules apply, and therefore, lenis have higher variation rates. This has a twofold meaning. First, results from the current study do not contradict the previous works in that lenis pronunciation is difficult for the learners. Second, Korean vocabulary is rich in phonological rule, such as fortification rule, and is an area that requires deliberate teaching.

Another major characteristic is substitution patterns between lateral and flap sounds, which was underemphasized in previous studies. Liquid sounds in Korean are pronounced as a flap when they are positioned at the onset position of the syllable. This pattern is consistent with the contrastive analysis result that flap sounds do not exist in Chinese, and due to L1 influence, they are often realized as lateral or retroflex /ɻ/.

Other characteristics include deletion of final consonants, substitutions in diphthongs by monophthongs, and final consonant insertions. Deletion could be explained by contrastive analysis of syllable structures where Korean consists of three parts, onset, nucleus, and coda, and Chinese has two parts, initial and final. This means that there is no sound in Chinese that is precisely equivalent to stop sounds at the coda of a Korean syllable, and that deletion occurs as learners adapt to the difference by omitting pronunciations that do not exist in L1.

The points below summarizes the comparison

- Substitution patterns between /r/ and /l/, which shows the highest variation rate are underemphasized in related works.
- L1 retroflex influence is underemphasized in related works.
- Consonant insertion is a new finding that was not discovered in related works.

3.4. Analysis of Frequent Segmental Variations

For the liquid sound, the variation with the highest frequency, this section analyzes the pattern with context. The paragraphs below present in more detail the variation patterns in /r/ and /l/ by describing where and how many times they occurred. Table 9 presents context-dependent patterns with examples. L1, L4, L5, L6, Ins1, and Ins2 patterns are new findings in this study that were not discovered in previous studies.

Context dependent variation patterns Table 9 are followed by possible explanations why they occurred. While the difference in syllable structures can explain why deletion patterns occur, it does not suggest a clear explanation for consonant insertion patterns. According to contrastive analysis, only /n/,/ɻ/ or /ŋ/, which are the possible stop sounds at the final in Chinese, can influence the pronunciation at coda. This characteristic occurs more often when the next syllable starts with a stop sound instead of a vowel. It is not clearly explicable, by contrastive analysis alone, why final consonant insertion patterns are observed at open syllables.

Other kinds of explanations are possible. For example, the transcribers were instructed to mark only substitution, deletion, and insertion, and no other linguistic phenomenon. It is possible that other linguistic phenomenon in Chinese such as juncture between the syllables could have influenced Korean production, which was marked in this study as "insertion" by the transcribers. Another possible explanation could be that rate of speech was slower, lenition between syllables were not realized, or similar sounds were introduced in order to enable easier articulation. It would be interesting to conduct more detailed analysis in the future work.

Table 9. Context-dependent variation patterns for liquids produced by Korean learners of Chinese. The patterns are represented as [canonical phoneme→realized phoneme / left context description__right context description], where Ins=insertion.

No.	Variation Pattern	Rate (Freq.)	Example
R1	R→L/ vowel__vowel	33.0 (529)	K AX R IY →K AX L IY
L2	L→ NN/ vowel__l	5.0 (244)	K WX L L IY → K WX NN L IY
L3	L→-/vowel__L	6.6 (243)	N OW L L AA T AA →N OW – L AA T AA
L4	L→ RR /vowel__L	1.42 (73)	M AX L L IY →M AX L RR IY
L5	L→RR/vowel__	1.17 (60)	JX NN PH IY L → JX NN PH IY RR
L6	L→RR/vowel__ T	0.58 (30)	PH A L T AA → PH A RR T AA
Ins7	- →RR/vowel__ R	45.14 (739)	N AA R AA → N AA RR R AA
Ins8	- →L/vowel__ R	2.5 (53)	JO R IY → JO L L IY

- R1: Since Chinese inventory does not have Korean flap sound, it is realized as lateral.

- L2: Korean phonological rule assimilating final nasal with the lateral sequence is not realized.

- L3: Since there is no lateral final in Chinese, it is omitted in pronunciation.
- L4,L5,L6: Chinese retroflex influences the Korean lateral pronunciation, which is not a possible final in Chinese.
- Ins7, Ins8: Since Chinese inventory does not have Korean flap sound, retroflex and lateral are inserted before realizing the flap, possibly to enable easier articulation by introducing similar sounds of mother tongue in between.

4. Modeling Pronunciation Variation for Speech Recognition

Non-native speech shows more pronunciation variations compared to native speech. To be able to automatically evaluate a pronunciation in a CAPT system, automatic recognition of the input speech is necessary, and non-native speech is an obstacle for ASR systems that are trained with native speech data [25, 26].

According to related works, there are two different approaches to improve recognition accuracy for non-native speech; pronunciation modelling and acoustic modelling. The former approach works by adding likely variation patterns in the pronunciation dictionary [26]. Quantified analysis of language-dependent variation patterns provides useful knowledge in this regard. The second approach, acoustic modelling, involves combining native and non-native acoustic models or adapting the native model to the non-native speech in order to improve the recognition accuracy. This approach requires sufficient amount of corpus in both languages [25].

This study models the pronunciation variations by using the experiment results discussed in the previous chapter, which identify the major characteristics of Korean segments produced by L1 Mandarin Chinese learners.

The next section describes how variation patterns were grouped into substitution, deletion, insertion, and phonological rule categories. The first three describe articulation-related characteristics, which are separated from phonological rule-related characteristics. What also affects pronunciation variations is learners' knowledge of phonological rules in Korean. Phonological rules dictate how an underlying phoneme is added, substituted, or deleted in its phonetic representation. For example, by the fortis rule the Korean pronunciation for the word "school" is /hak̚k⁼jo/ and not /hak̚kjo/, which means that the canonical pronunciation is different from the underlying representation. This often results in learners' mispronunciations depending on whether or not they have the knowledge of such rules, and has correctly applied them.

4.1. Rule Derivation

This section describes how 158 pronunciation variation rules are derived from 48 major variation patterns, followed by generation of 1,737 pronunciations for 300 words by using these rules. It is necessary for the pronunciation variation rules to consider left and right phonemic context because the variation patterns can occur or not occur depending on the phonemic context. At most one phoneme for each left and right context are considered for defining variation patterns. For example, for the substitution pattern in which canonical pronunciation /PP/ is realized as /P/, variation rules with differing contexts can be generated, such as [TQ $ PP AA > TQ $ P AA] and [AA $ PP AA > AA $ P AA], where $ denotes syllable boundary. In order to generate the condition and realized output of the rules, we use the PLU (Phone-like Unit) set shown in Table 1, adapted from [22].

Because including rare patterns in the pronunciation dictionary can cause confusion in the search space and mislead the recognition result [26], not all variation patterns are used to derive pronunciation variation rules that actually affect the pronunciation lexicon.

Table 10 shows some of these variation rules. Each rule consists of Condition Input and Rule Output to describe the phonemic context where the variation occurs and how it is realized. The rules are then categorized into substitution, deletion, insertion, and phonological knowledge, as shown in Table 11. Phonological knowledge refers to cases of non-application or mis-application of phonological rules of Korean. In the examples paragraph above, the former where /PP/ is preceded by /TQ/ is a substitution rule whereas the latter, where /PP/ is preceded by /AA/, is a rule concerning phonological rule.

Table 10. Examples of pronunciation variation rules used for pronunciation modeling (shown in sequences of left phoneme, phoneme of interest, and right phoneme, where “$”= syllable boundary, “_”= insertion, “-“ = deletion)

Condition Input (PLU)	Rule Output (PLU)	Examples (English meaning)
Vowel $ r (Vowel $ R)	Vowel $ l (Vowel $ L)	nara (country) → nala
t˺ $ p͈ Vowel (TQ $ PP Vowel)	t˺ $ p Vowel (TQ $ P Vowel)	tɕʰot˺p͈ul (candle) → tɕʰot˺pul
p˺$ t͈ Vowel (PQ $ TT Vowel)	p˺$ d̥ Vowel (PQ $ T Vowel)	d̥ʌp˺t͈a (to be hot) → d̥ʌp˺d̥a
n˺$ t͈ Vowel (NN $ TT Vowel)	n˺$ d̥ Vowel (NN $ T Vowel)	an˺t͈ a (to sit) → an˺t͈ a
Vowel $ d̥ʑ (Vowel $ Z)	Vowel_k˺ $ d̥ʑ (Vowel_KQ $ Z)	g̊ad̥ʑok˺ (family) → g̊ak˺d̥ʑok˺
Vowel $ p͈ (Vowel $ PP)	Vowel_p˺ $ p͈ (Vowel_PQ $ PP)	ap͈a (father) → ap˺p͈a
Vowel k˺ $ s͈ (Vowel KQ $ SS)	Vowel - $ s͈ (Vowel - $ SS)	nak˺s͈i (fishing) → nas͈i

Table 11. Number of variation rules and pronunciation variations generated from 300 words (Some rules generate more than others depending on the phoneme context of the list of words).

Type	Abbrevi-ation	Category	Variation Rules	Pronunciation Variations
Con-son-antal	Del.	Deletion	26	71
	Phon.	Phonological Knowledge	28	73
	Sub.	Substitution	53	577
	Ins.	Insertion	51	1016
Voc-alic	Mono.	Monophthong Substitution with Monophthongs	80	67
	Diph 1.	Diphthong Substitution with Diphthongs	16	5
	Diph 2.	Diphthong Substitution with Monophthongs	31	23

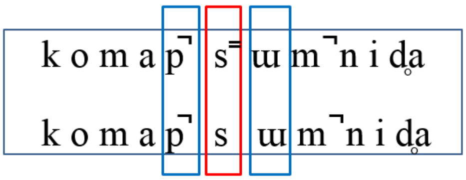
4.2 Dictionary Generation

One sequence per word is defined as the canonical pronunciation, according to the definitions available in The National Institute of the Korean Language [27]. The pronunciation model of the baseline system consists of the canonical pronunciation and their possible realizations by Korean native [28].

To model the pronunciation variations, the canonical and transcribed phonemes are aligned. Figure 2 illustrates an example alignment of the canonical pronunciation and transcribed result for the word "thank you" in Korean. The phonemes where a variation rule applies are highlighted in boxes, with left and right context. We can see that the condition part of the substitution rule, [PQ SS Vowel], matches the canonical pronunciation, and the realized output of the rule, [PP S Vowel], matches the transcribed result. From this case, the training data learns to increase the generation probability for the phonological knowledge rule [PQ SS Vowel > PQ S Vowel]. The learning algorithm can be expressed as the equation below.

$$p(Variation\ Rule) = \frac{\mathrm{p(Canonical\ \&\ Realized\ Pronunciation)}}{\mathrm{p(Canonical\ Pronunciation)}}$$

Figure 2. Alignment of canonical and transcribed result for the word k o m a p̚ s⁼ɯ m̚ n i d̥ a (thank you), where NF=No Final



4.3 Speech Recognition Experiment

This study use L2KSC (L2 Korean Speech Corpus), a speech corpus built for Korean as a foreign language [21]. The corpus was built to evaluate acquisition of phonetic and phonological sounds in Korean language by learners of various L1 backgrounds. The list of 300 words in the corpus is based on the vocabulary found in 8 mainstream textbooks, including frequently used nouns, compound nouns, noun phrases, and verbs. The gender and proficiency are balanced in the distribution, with 20, 19, 14 students in beginner, intermediate, and advanced levels, and 25 male and 28 female students, each respectively.

For the baseline, Korean native speech corpus is used for the list of 300 words in L2KSC. This corpus, consisting of 51 native speakers, was built for the purpose of comparative analysis of speech produced by non-natives with speech produced by natives. One third of the words is chosen as the test set, and the remaining data is used as training and development sets.

For the training set, L2KSC corpus spoken by 53 Chinese learners is randomly partitioned into training and test sets, in which 100 words are assigned as test corpus, and the remaining data are used as train and development sets. For the 200 words in the training set, generation probabilities are learnt from the transcribed results. MLLR and MAP are used for speaker adaptation.

In the pronunciation dictionary of the trained system, pronunciation variations are generated for each word in the test set. In order to optimize the pronunciation generation for the pronunciation dictionary, pruning threshold option is used. For different categories of rules, the option cuts off all candidates with higher $-log_2(\text{p})$ value than the set threshold. This allows to check that the pronunciation variations are generated in a way that minimizes the confusability in the search space.

4.4 Speech Recognition Result and Discussion

The acoustic model trained on native corpus with native pronunciation dictionary serves as a baseline. Pronunciation variation rules are used to generate extended pronunciations from the canonical pronunciation input. After checking how each type of rule affects recognition accuracy, combinations of rules were tested to further reduce the error rate. Table 4 shows the recognition results in terms of WER (word error rate). The first column of the table refers to the rule abbreviations in Table 12.

In general, the accuracies are higher for advanced learners than for beginners, and intermediate learners show closer tendency with the advanced than the beginners. The best result for consonantal model is achieved when the pronunciation model includes variation rules regarding substitution, insertion, and phonological knowledge. Compared to the baseline system, 3.3% absolute WER reduction that is 21.2% relative WER reduction is obtained. For the vocalic model, the best result is obtained when the model only includes diphthong substitution with other diphthongs. Compared to the baseline, absolute and relative reduction rates are 0.13% and 1.3%.

Table 12. Recognition results for consonantal pronunciation models for each speaker level (in WER)

Pronunciation model	Beginner	Intermediate	Advanced	Average
Baseline	23.79	13.65	8.6	15.49
Del.	23.85	13.65	8.37	15.43
Phon.	23.85	13.49	8.31	15.37
Sub.	22.08	11.72	7.31	13.87
Ins.	20.44	11.72	7.13	13.21
Sub.&Ins.	19.85	10.1	6.54	12.34
Phon.& Sub.&Ins.	**19.67**	**10.02**	**6.42**	**12.21**
Del.&Phon&Sub.&Ins	19.91	**10.02**	6.48	12.32

Table 13. Recognition results by pruning threshold levels for each speaker level (in WER)

Pruning Threshold	Beginner	Intermediate	Advanced	Average
0	23.79	13.65	8.6	15.49
2	20.79	10.72	6.89	12.98
4	20.20	10.56	6.72	12.66
6	20.38	**9.87**	6.78	12.55
8	19.96	**9.87**	6.84	12.43
10	**19.67**	10.02	**6.42**	**12.21**

Table 14. Recognition results for vocalic pronunciation models for each speaker level (in WER)

Pronunciation model	Average
Baseline	15.49
Mono.	15.65
Diph 1.	**15.31**
Diph 2.	15.61
Mono.&Diph 1.&Diph 2.	15.36

Among the seven different categories of variation rules, insertion rules contribute the most in the recognition performance improvement regardless of learner levels. As shown in Table 12, the variation rules by categories are incrementally applied. Deletion rules increase the performance only for the advanced learners by a small amount. The case is similar for phonological knowledge rules. In contrast, when substitution and insertion rules are added, the recognition accuracy improves for all levels of learners.

The degree of improvement is higher for insertion than for substitution rules. In order to confirm that this is not resulting from the larger number of pronunciation variations generated for insertion rules than substitution rules, additional test was conducted under the condition that the same number of pronunciation variations were generated for these two categories. The results for insertion rules still influenced recognition accuracy more positively than substitution rules, by 1.24% in average, which verifies the observation that insertion rules have contributed the most for the error reduction.

This in turn confirm the new findings observed in the corpus-based analysis. As summarized earlier, surveying previous studies on Korean spoken by Chinese learners only mention substitution, deletion, and phonological knowledge, and do not mention that consonant insertion is a prominent pattern. By analyzing the auditory transcriptions, however, consonant insertion was also found to be a frequent characteristic. This finding is, in fact, the most important category of rules in our experiment that improved ASR performance for learners of all levels. This means that final consonant insertion deserves more attention to analyze the phenomenon and explain why it occurs and was previously left undiscovered. This calls for interesting future work.

Table 13 shows recognition results for average number of pronunciation variations generated per word. The number of pronunciation variations is controlled by setting pruning threshold options. Within the rules we have set, 10 is the maximum possible average number of pronunciation variations that can be generated. Our results show that the higher the pruning option, the lower the number of pronunciation variations. That is, the error rate decreases as more pronunciation variations are allowed.

Table 14 shows recognition results when vocalic pronunciation variations were included in the dictionary. WER decreased from 15.49 to 15.36, which is 1.3% of relative improvement. This is minor compared to 21.2% relative improvement observed when the same experiment was conducted to include consonantal variation patterns. There are two possible reasons why there is a major difference in this degree of improvement. First, the number of variations generated is far less for vocalic than for consonantal patterns. Pronunciation variations are generated by applying the variation rules on the canonical pronunciation strings. For vocalic variation rules, 95 variations were generated, while 1737 were generated for consonantal variation rules. For example, although the phoneme /ɰ/ shows high variation rate of 47.16%, the phoneme only occurs for only two words in the corpus, reducing the possible number of phonemic contexts where the rule can apply. Moreover, the vocalic variation rates are lower than those of consonants in general, whose averages are 13.74% and 3.35%, respectively. This means that less recognition errors were caused by vocalic variations and therefore, the room for improvement is lower for vocalic variations.

The results can be interpreted for each learner level. For beginners, variations rules for deletions and phonological knowledge do not improve the recognition result, whereas substitution and insertion rules improve the recognition rate. For intermediate and advanced learners, adding substitution, insertion, and phonological knowledge rules in the pronunciation dictionary improves the recognition result. In all cases, adding insertion rules are most efficient in recognition improvement.

The experiment results confirm the hypothesis that modeling pronunciation variation by using data-driven analysis improves the recognition result. Pronunciation modeling approach used in this study can enable a CAPT system to recognize learner's speech.

The methodology can be further used in a CAPT system to provide corrective feedback, For a CAPT system, it is not only important to improve the overall ASR performance, but it is equally important to know exactly what kinds of features are more useful than others so that the CAPT system can also provide a corrective feedback by comparing the pattern with the canonical pronunciation. Previous study also support this method of using context-sensitive rules to diagnose mispronunciation [29].

For example, for a given recognition result "S AA KQ $ K WA" for the canonical pronunciation "S AA $ K WA" (apple), the system knows that the variation rule [Vowel $ K > Vowel KQ $ K] in the insertion category was used to detect the word. The system can then generate corrective feedback that consonant insertion occurred at the syllable ending. Moreover, the use of deletion rules in generating lexical entries does more harm than it helps for beginner level, while it is helpful for advanced levels. In this way, variation pattern modeling can be used for providing corrective feedback to the learners specific to their learning levels.

5. Conclusion

For developing a CAPT system for Mandarin Chinese learners of Korean, this study identified L1 specific variation patterns and used the quantitative analysis result for pronunciation modeling to improve speech recognition performance.

Surveying of the related works on variation pattern with agreements and disagreements revealed the need for a corpus-based experiment. Based on the results, the characteristics of variation patterns of Korean produced by Chinese learners were summarized. These findings allowed this study to resolve the disagreements among the previous studies, found that substitutions in liquid sounds had been under-emphasized, and discovered new variation patterns such as consonantal insertion that were previously left undiscussed.

From this, major context-dependent rules from the variation patterns were used to generate pronunciation variations from canonical pronunciation. In order to find the most effective way of designing a CAPT system, different types of pronunciation dictionaries were incrementally tested for different levels of speakers. Using this model, 21.2% relative WER reduction was obtained for consonants and 1.3% for vowels. The most effective pronunciation variation model was insertion, followed by substitution, phonological knowledge, deletion, substitution of diphthongs with other diphthongs, with monophthongs, and substitution of monophthongs with other monophthongs, in the order of importance.

This pronunciation modeling approach can be further utilized in a CAPT system for feedback generation. Moreover, insertion patterns that were newly discovered in this study is shown to be the most influential category for improving recognition, which is an interesting observation and calls for a future work on a closer inspection on this phenomenon.

References

[1] M. Eskenazi. “An overview of spoken language technology for education,” Speech Communication, 51, 832-844. 2009.

[2] Bernstein J., Van Moere, A., and Cheng J. (2010). Validating automated speaking tests. *Language Testing,* 27 (3), 255-277.

[3] D. Higgins, K. Zechner, X. Xi, and D. Wiliamson, "A Three-Stage Approach to the Automated Scoring of Spontaneous Spoken Responses," Computer Speech and Language, 25(2), pp. 282-306, 2011.

[4] K. Zechner, D. Higgins, X. Xi, and D. Wiliamson, "Automatic Scoring of Non-native Spontaneous Speech in Tests of Spoken English," Speech Communications, 51(10), pp. 883-895, 2009.

[5] Qian, Xiaojun, F. Soong, and H. Meng. "Discriminatively Trained Acoustic Model for Improving Mispronunciation Detection and Diagnosis in Computer Aided Pronunciation Training (CAPT)," Proc. of Interspeech. 2010.

[6] H. Hong, S. Kim, and M. Chung, “A Corpus-based Analysis of Korean Segments Produced by Japanese Learners,” in Proc. SLaTE 2013, Grenoble, France, pp. 189-192, 2013.

[7] N. Chen, V. Shivakumar, M. Harikumar, B. Ma, H. Li, “Large-Scale Characterization of Mandarin Pronunciation Errors Made by Native Speakers of European Languages,” in Proc. Interspeech 2013, Lyon, France, pp. 2370-2374, 2013.

[8] Lee, Ann, and James Glass. "Context-dependent Pronunciation Error Pattern Discovery with Limited Annotations." Fifteenth Annual Conference of the International Speech Communication Association. 2014.

[9] Hu, Wenping, et al. "Improved mispronunciation detection with deep neural network trained acoustic models and transfer learning based logistic regression classifiers." Speech Communication 67: 154-166. 2015

[10] J. E. Flege, “Second language speech learning: Theory, findings, and problems”, in W. Strange [Ed], Speech Perception and Linguistic Experience: Issues in cross-language research, 233-277, MD: York Press, 1995.

[11] H.Y. Lee, Korean Phonetics, Taehaksa, Seoul, Korea, 1996.

[12] Y.H. Lin, Sounds of Chinese, Cambridge University Press, New York,

U.S.A., 2007.

[13] Y. Qin, “A Study on the Plans for Teaching Korean Pronunciation to Chinese Learners,” M.A. thesis, Konkuk University, Seoul, Korea, 2010.

[14] M. Hwang, “A Study on Errors in Consonant Pronunciation by and an Educational Approach for Korean Language Learners from Chinese-Speaking Region,” Ph.D. dissertation, Chungnam National University, Daejeon, Korea, 2012.

[15] K. Kim, “A Study on Teaching Korean Pronunciation to Chinese Speakers,” Ph.D. dissertation, Dankook University, Seoul, Korea, 2008.

[16] L. Leng, “Chinese Pronunciation of Consonants Education of Korea Students,” Ph.D. dissertation, Chungnam National University, Daejeon, Korea, 2014.

[17] J. Cui, “An Approach to Teaching Consonants to Chinese Learners of the Korean Language,” Keimyung Korean Studies, vol.31, pp. 215-232, 2004.

[18] J. Cui, “Teaching Korean Pronunciation for Chinese Learners,” Bilingual Research, vol. 20, pp.309-343, 2002. (in Korean)

[19] J. Han, et. al, Teaching Korean Pronunciation, Hollym, Seoul, 2003.

[20] H. Cho, “A study on the teaching method of Korean pronunciation for foreigners - Focused on Chinese-speaking learners of Korean,” Korean Journal of General Education, vol. 7, no. 6, pp. 531-559, 2013.

[21] S. Lee, J. Kim, and J. Chang., “Design and Construction of Speech Corpus for Korean as a Foreign Language (L2KSC),” The Journal of Chinese Language and Literature, vol. 33, pp. 35-53, 2005.

[22] H. Hong and M. Chung, “Improving phone recognition performance via phonetically-motivated units,” in Proc. Interspeech 2009, Brightoon, UK, pp.1663-1666, 2009.

[23] H. Ryu, K. Lee, S. Kim, and M. Chung, “Improving transcription agreement of non-native English speech corpus transcribed by non-natives,” in Proc. SLaTE, 2011, Venice, Italy, pp. 61-64, 2011.

[24] S. Goddijn, S. Binnenpoorte, "Assessing Manually Corrected Broad Phonetic Transcriptions in the Spoken Dutch Corpus," in Proc. 15th ICPhS 2003, Barcelona, Spain, pp. 2981-2984, 2003.

[25] Y. R. Oh, J. S. Yoon, and H. K. Kim, “Acoustic model adaptation based on pronunciation variability analysis for non-native speech recognition,” Speech Communication, vol. 49, no. 1, pp. 59-70, 2007.

[26] H. Strik and C. Cucchiarini. "Modeling pronunciation variation for ASR: A survey of the literature," Speech Communication, volume 29, pp. 225-246, 1999.

[27] Retrieved from http://www.korean.go.kr/

[28] K.N. Lee and M. Chung, "Morpheme-based modeling of pronunciation variation for large vocabulary continuous speech recognition in Korean." IEICE transactions on information and systems 90.7 pp. 1063-1072, 2007.

[29] A. Harrison, W.Y. Lau, H. Meng, and L. Wang, “Improving mispronunciation detection and diagnosis of learners’ speech with context-sensitive phonological rules based on language transfer,” Proc. of Interspeech, 2008.

Appendix

A. Confusion matrix generated for transcription results for consonants, where the x and y axes refer to transcription and canonical results. The units are the PLU set defined in Table 6.

	CH	HH	K	KH	KK	KQ	L	M	MM	N	NN	NX	P	PH	PP	PQ	R	S	SS	T	TH	TQ	TT	Z	ZZ	CS	F	RR	SH	ZH	ZS	del	Sum
CH	679	3	1	4		1												7	2	4	11		5	18	4	45					3		787
HH		1573		4			19	5		62							15		1		2											21	1702
K		2	3218	118	193	2	9			1		1					11	1		4	4		1	6							1	15	3587
KH		36	45	911	30	4	3		1			1						1			3											7	1042
KK	1		250	76	1266		2			1										4	2		4	1									1607
KQ			3			1774	29		2		7	9				5						20						6				230	2085
L						8	3225		3	193	244	41				46	727			5	2	27						205				243	4969
M				1			3	1983	2		1	2	4		1																	9	2006
MM		1				4	20	27	1787	2	18	18			2	54						7		1				9				31	1981
N			2				199	12		2743	4			1	1	1	55	4	1	4	1		1	2				1				87	3119
NN						10	25		21		2524	122										41						5				140	2888
NX		1				110	4		16	3	135	2492										2						2				116	2881
P		1						6	1		2		1151	74	60	10				1												8	1314
PH		24		4									10	620							1						1					1	661
PP				1				1					190	65	683	1			1														942
PQ						8	5	2	17		2	3	1	3		1141						4						4				34	1224
R			2		1		529			16							1025	1		5	4							13				8	1604
S	3	1	1							3	1							2293	176	1						2			15		1	5	2502
SS	5																	117	1046				1	4	1	2			6		1	1	1184
T		1	5				2			2							3			2948	114		502	2							2	3	3584
TH	1	9	1	2		1	2													36	809		39									1	901
TQ	1					68	6		2	1	9	6				11		1	1			1398						8				279	1791
TT			3				3	1		1										240	52		1538									1	1839
Z	62		3	2			1			1		1						9	1	18	2		4	1310	117	17				8	134	4	1694
ZZ	18																		1					73	469	8					44		613
ins		1				178	89	6	28	5	120	38	5		1	72	3	2		1	2	213						873					1637

B. Confusion matrix generated for transcription results for vowels, where the x and y axes refer to canonical and transcription results. The units are the PLU set defined in Table 6.

	AA	AX	EH	EY	IY	JA	JE	JH	JO	JU	JX	OW	UI	UW	WA	WE	WH	WI	WW	WX	del	Sum
AA	11509	22	11	5	14	2			2		2	2		3	1				3		3	11579
AX	27	3113	5	2	5				6		19	94		8					25	2	5	3311
EH	19	14	1212	5	11		3	1									1		8			1274
EY	1	4	5	505	1		1												2			519
IY	7	4	2	2	5679		1			3	14		2						3		28	5745
JA	29					518	1				21											569
JE			6	14	5		301				5											331
JH			3				9	35														47
JO						1			1199	20	2	6					1					1229
JU		1			1				5	221			6	5								239
JX	1	18	1	1	31	19	11		27	7	1622	3	1			3						1745
OW	2	22			1				2	1		3063		60			1		12	1		3165
UI			1										132	3		4		3				143
UW		5	1							1		13		2944					10	7	5	2986
WA	8	2				1							1	1	540	7	3		2	7		572
WE		2	2	1			1	1				2	43	5	5	404	6	1	1			474
WH			8	1	1							1	3	1	3	6	210			1		235
WI			6	18	69						1		8			3		121	1		2	229
WW	2	33		1	9	1					1	9		17					1991		8	2072
WX		1											3	11	6	6	6		5	195		233
ins	5	5	1	3	37	1	1		2		1	6		5	1			1	13	6		88

C. List of read-speech words in the experiment described in Chapter 3

1. 앉다
2. 콧물
3. 여덟
4. 밖
5. 건강
6. 씨앗
7. 난로
8. 늦었다
9. 앞치마
10. 부탁했어요
11. 특히
12. 벌레
13. 거짓말
14. 썼다
15. 왕
16. 처음
17. 더럽다
18. 요리
19. 아버지
20. 근처
21. 좋겠다
22. 관광
23. 사과
24. 두더지
25. 삐약삐약
26. 규칙
27. 꽃
28. 너구리
29. 뻐꾸기
30. 기악
31. 등산로
32. 그렇지만
33. 토요일
34. 같아요
35. 가득하다
36. 신라호텔
37. 신선하다
38. 압력
39. 치약
40. 라디오
41. 궤도
42. 일년
43. 축하
44. 곤란합니다
45. 듣습니다
46. 복잡하다
47. 냇물
48. 결단력
49. 된장
50. 사람
51. 공부
52. 문화
53. 까마귀
54. 싸움
55. 입원료
56. 노래
57. 허리띠
58. 빨갛다
59. 끼리끼리
60. 숭늉
61. 안녕
62. 옆 집
63. 아프다
64. 넣다
65. 실내
66. 끼우다
67. 낚시
68. 전화
69. 나라
70. 핑계
71. 동화책
72. 한라산
73. 감다
74. 웃겼다
75. 극장
76. 빨래
77. 가지
78. 명동
79. 선로
80. 쑥
81. 안경
82. 텃밭
83. 기억
84. 끝났군요
85. 따뜻하다
86. 수탉
87. 덥다
88. 빨라요
89. 넉넉하다
90. 멀다
91. 깨끗하다
92. 신문로
93. 즐겁다
94. 학년
95. 쓰다
96. 문을 닫는다.

97. 삯
98. 고맙습니다
99. 다섯 명
100. 낮
101. 선물
102. 벗
103. 개미
104. 디딤돌
105. 돼지
106. 옷깃
107. 닫히다
108. 연락하세요
109. 달립니다
110. 오빠
111. 4000 원입니다.
112. 송아지
113. 졸업식
114. 옛날
115. 아빠
116. 없다
117. 예습
118. 빵집
119. 혓바닥
120. 회의
121. 읽기
122. 김밥
123. 냉큼
124. 멀리
125. 횃불
126. 한 냥
127. 꾀꼬리
128. 멋있다
129. 하숙집
130. 경복궁
131. 꺼칠하다
132. 뚜벅뚜벅
133. 꽈리고추
134. 밤낮
135. 21 일입니다
136. 낱개
137. 읊다
138. 조약돌
139. 떼쓰다
140. 한국말
141. 책
142. 미국
143. 연필
144. 꾸지람
145. 누나
146. 통화
147. 동생
148. 닮았지요
149. 먹다
150. 무엇입니까
151. 부엌
152. 저녁식사
153. 10 월
154. 젊다
155. 압구정동
156. 불렀어요
157. 나뭇잎
158. 힘줄
159. 넓다
160. 밥
161. 민주주의의 의의
162. 들어가세요
163. 있어요
164. 생각했습니다
165. 크다
166. 파랑
167. 얌냠
168. 폐렴
169. 어머니
170. 주었다
171. 푼돈
172. 길러요
173. 꽃집
174. 자장면
175. 거리
176. 뿌리
177. 돈을 찾아요
178. 단어
179. 귀
180. 칼날
181. 낮잠
182. 편리합니다
183. 놀라다
184. 수영하다
185. 북녘
186. 값도
187. 갑자기
188. 어른
189. 짬뽕
190. 게시판
191. 장독
192. 삿갓
193 목욕
194. 승리
195. 괜찮습니다
196. 냉면
197. 벼 이삭
198. 젓가락

199. 옷
200. 저녁
201. 괘종시계
202. 달라요
203. 고장
204. 티끌
205. 부르다
206. 달걀
207. 무쇠
208. 그림
209. 꽤 많다
210. 울다
211. 핥다
212. 촛불
213. 느티나무
214. 선생님
215. 골라요
216. 계란
217. 전라도
218. 교회
219. 우유
220. 묶
221. 늦잠
222. 싫군요
223. 꿩
224. 신라면
225. 초록색
226. 맵시
227. 숙녀
228. 삶다
229. 춥다
230. 국화
231. 독립
232. 팔다
233. 캄캄하다
234. 사랑하다
235. 좌회전
236. 껴안다
237. 월요일
238. 두뇌
239. 공항
240. 연습하기
241. 봄
242. 갔다
243. 얘기
244. 예쁘다
245. 국제
246. 밟다
247. 가족
248. 갸륵하다
249. 한강
250. 할 일
251. 키읔
252. 과자
253. 좋습니다
254. 코 골아요
255. 밝아요
256. 권리
257. 결혼
258. 못했어요
259. 땅콩
260. 효도
261. 괴물
262. 달나라
263. 흐린 날
264. 백 년
265. 빼앗다
266. 세상
267. 맛있다
268. 몇 년
269. 물약
270. 모양
271. 국물
272. 음악
273. 번호
274. 토끼
275. 잉어
276. 신발
277. 생산량
278. 함께
279. 선릉역
280. 해돋이
281. 길어요
282. 컴퓨터
283. 겨울
284. 튀김
285. 맞는다
286. 답답하다
287. 애기
288. 꿰매다
289. 휴게실
290. 옮다
291. 아홉시
292. 당뇨병
293. 냇가
294. 줄넘기
295. 돌솥 비빔밥
296. 십 년
297. 입학
298. 숨다
299. 이쑤시개
300. 우산

D. Instructions provided to transcribers, which were handed out along with the speech data and canonical transcription results.

Transcription Instructions for Korean Sounds spoken by Chinese

I. Purpose of Transcription
The purpose of this transcription is task is to record the actual Korean sound spoken by Chinese learners.

II. Data for Transcription
The speech data for transcription are composed of 300 Korean meaningful words spoken by 53 male and female speakers.

III. Transcription

1. Mark the difference when the realized segment is different from the provided canonical transcription.

2. Use only 50 given PLU (Phoneme-like Unit) set, in which six phonemes are unique to Chinese language.

3. Substitution, Insertion, Deletions are marked.

   - Substitution: Mark the PLU of the substituted phoneme.

   - Deletion: There may be cases of Korean syllable final deletions. The deleted segment should be marked by the hyphen "-" symbol in place of the deleted segment. E.g. 안녕→안녀 (aa nn n jx -)

   - Insertion: The inserted vowels or consonant segment(s) should be marked by the underscore "_" symbol before the inserted segment. E.g. There may be cases of vowel epenthesis, lengthened vowels, or other kinds of insertions. 늦었다→느저쓰다 (n ww z ax ss_ww t aa), 호텔→호우텔 (h ow_uw th ey l]

4. Abnormal Sound Files: There may be cases when the speaker did not produce a word, or when the recording quality is unacceptable. Such file names should be separately marked and left untranscribed, along with a brief comment on the reason why. The list should be submitted together with the transcribed results.

IV. Checkpoints for Transcription

1. Phone set: Only the 50 PLU's and the two symbols "_" and "-" are allowed. Make sure not to use anything else to mark, such as a space, or tab.

2. Praat TextGrid Boundary: All segments are separated by time

boundaries. Do not edit the time stamps. If it was moved or edited by accident, exit the program and restart the program.

PLU set for Korean sounds

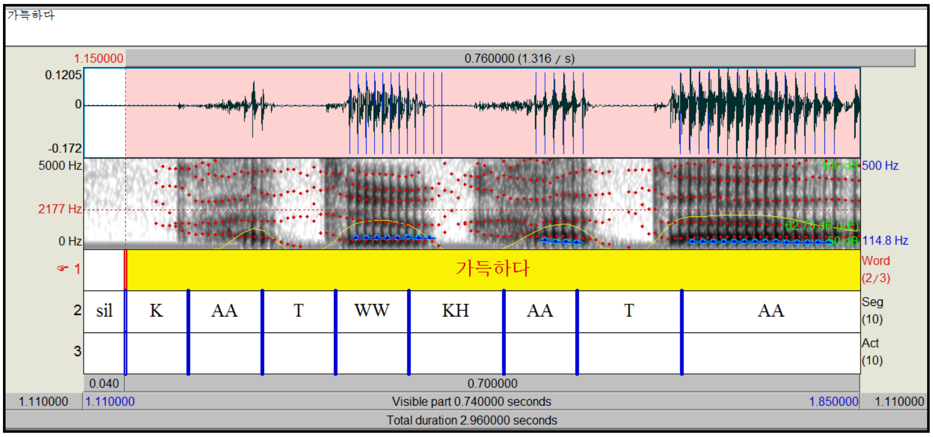
Vowels							
PLU	IPA	Korean	Example	PLU	IPA	Korean	Example
AA	a	ㅏ	아이	AX	ʌ	ㅓ	언니
OW	o	ㅗ	고모	UW	u	ㅜ	구경
IY	i	ㅣ	기술	WW	ɯ	ㅡ	그녀
EH	ɛ	ㅐ	해	EY	E	ㅔ	꽃게
UI	ɥi	ㅟ	복귀	JA	ja	ㅑ	야구
JX	jʌ	ㅕ	견제	JH	jɛ	ㅒ	얘기
JE	je	ㅖ	계단	JO	jo	ㅛ	교통
JU	ju	ㅠ	휴식	WA	wa	ㅘ	과일
WH	wɛ	ㅙ	돼지	WX	wʌ	ㅝ	권장
WE	we	ㅞ	궤양	WI	ɯi	ㅢ	의자
Consonants							
PLU	IPA	Korean	Example	PLU	IPA	Korean	Example
K	ɡ̊	ㄱ	가을	KQ	k̚	ㄱ	색동
KK	k˭	ㄲ	까닭	KH	kʰ	ㅋ	칼
T	d̥	ㄷ	다리	TQ	t̚	ㄷ	받고
TT	t˭	ㄸ	딸	TH	tʰ	ㅌ	탈
P	b̥	ㅂ	바람	PQ	p̚	ㅂ	입술
PP	p˭	ㅃ	빨래	PH	pʰ	ㅍ	파리
Z	d̥ʑ	ㅈ	자리	ZZ	tɕ˭	ㅉ	짜임새
CH	tɕʰ	ㅊ	처음	HH	H	ㅎ	하늘
S	s	ㅅ	소리	SS	s˭	ㅆ	싸리
M	m	ㅁ	마음	MM	m̚	ㅁ	감
N	n	ㄴ	나리	NN	n̚	ㄴ	간
NX	ŋ	ㅇ	둥지	L	l	ㄹ	빨래
R	ɾ	ㄹ	소리				

PLU set for Chinese sounds

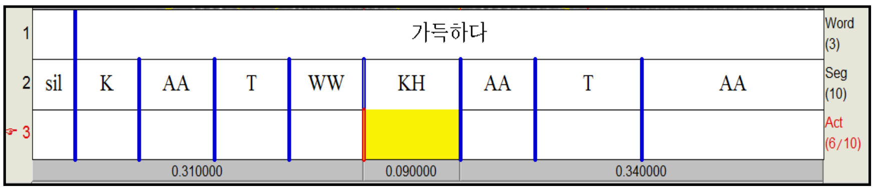
Category	Chinese Segment		Mispronunciation example (canonical→realized)		
	IPA	PLU	IPA	PLU	Korean
Dental-alveolar	ts	ZS	ʥaraŋ→ tsraraŋ	Z AA RAA NX→ZS AA RAA NX	자랑
	tsʰ	CS	tɕʰɛgim→ tsʰɛgim	CH EH K IM →CS EH K IM	책임
Retroflex	ʂ	SH	saŋ → ʂaŋ	S AA NG→ SH AA NG	상
	tʂ	ZH	ʥaraŋ → tʂraraŋ	Z AA RAA NX→ZH AA R AA NX	자랑
	ɹ	RR	baɾam → baɹam	P AA R AA M → P AA RR L AA M	바람
Labio-dental fricative	f	F	pʰado → fado	PH AA T OW => F AA T OW	파도

VI. Transcription Sample Screen

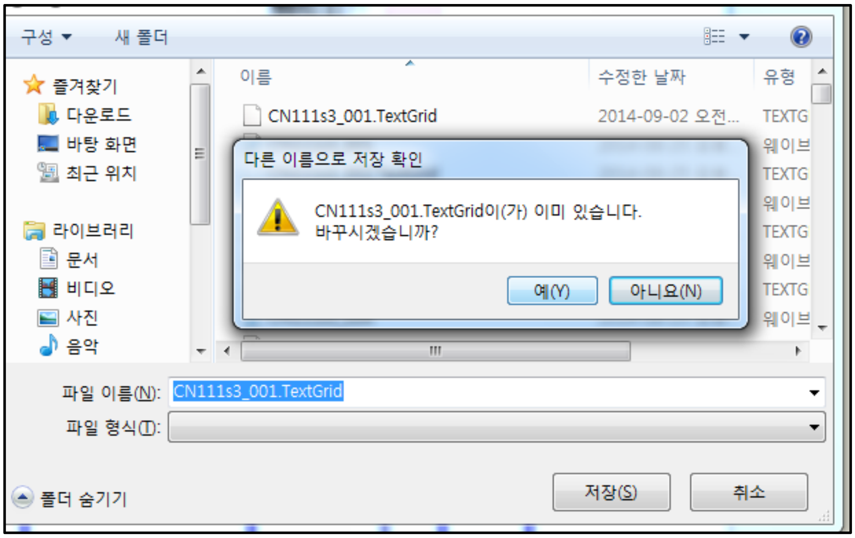
1. Overall View



2. Use the bottom tier (Tier 3) when the auditory result is different from the canonical string and mark accordingly.



3. When the auditory result has been marked, save the TextGrid file.



중국인 학습자의 한국어 분절음 변이양상과 발음모델링

양승희
협동과정 인지과학 전공
서울대학교 대학원
Seoul National University

컴퓨터기반 발음교육 (Computer-Assisted Pronunciation Training, CAPT) 시스템은 학습자의 발음오류를 자동으로 인식 및 검출하고 개인화된 교정 피드백을 제공함으로써, 효율적인 제 2외국어 학습을 제공한다. 최근 증가하는 한국어 교육에 대한 수요는 한국어 CAPT 시스템 개발의 필요성을 의미한다. 그러나, 비원어민 한국어 발화에 대한 연구는 부족한 실정이며, 학습자의 발음에 대한 비원어민 음성인식 성능이 향상됨으로써 학습 인터페이스에 구현이 가능할 것이다.

중국인 학습자를 위한 한국어 CAPT 시스템 개발의 초기 연구로써, 본고의 초반부는 선행연구 실험결과의 일치 및 불일치 양상을 정리한다. 그리고, 코퍼스 기반 실험을 진행하여 불일치 양상을 해소하여 주요 변이양상을 분석한다. 본 연구의 실험결과 탄설음의 설측음화 변이율이 35%로 가장 높게 나타났다.

본고의 후반부는 발음모델링과 음성인식 실험결과를 제시한다. 정량화된 변이양상의 문맥을 고려하여 대치, 삽입, 삭제, 음운변동 패턴

규칙으로 구분하였다. 이 규칙을 사용하여 확장된 발음열을 구축하여 인식 실험을 진행하였다. 이 방법으로 자음 및 모음 모델에서 상대 인식률 21.2%과 1.3%의 개선이 각각 확인되었다. 인식실험 결과는 코퍼스 기반 변이양상 분석 결과와 발음모델링 기법을 검증한다.

Keywords : (6단어 이내) **컴퓨터기반 발음교육** (CAPT), **중국인 학습자의 한국어 발음, 비원어민 음성인식, 발음모델링**
Student Number : 2014-20137