An accent reduction course for native speakers of American English learning Korean[*]

Jong-mi Kim
(Kangwon National University)

Kim, Jong-mi. 2012. An accent reduction course for native speakers of American English learning Korean. *Studies in Phonetics, Phonology and Morphology* 18.1, 27-50. An accent reduction course was offered for native speakers of American English learning Korean to see if such a course would affect accent reduction as measured by several major phonological elements. Thirty-two adult American English native speakers at various levels of learning Korean as a second language took a five-week, one session per week course in Korean pronunciation during the 2010-2011 Academic Year at the University of Hawai`i at Manoa. Each session looked at one of the following five major phonological elements in Korean: 1) consonants, 2) vowels, 3) phonological adjustments, 4) intonation, and 5) rhythm.

The learners took a set of listening comprehension and speech production pre- and post-tests, and received mid-term feedback on their pre-test results. A control group of nine Korean native speakers took the same test, so that the test results could be compared. All test results were evaluated by another group of 56 native Korean speakers in terms of 1) accuracy for the given phonological elements, 2) fluency of the embedded utterances, and 3) the relative degree of a non-native accent apparent in each utterance before and after the instruction. The results show that 1) accent reduction was demonstrated, despite what has been written in the literature (e.g. Morales Pech and Izquierdo 2011 for pros and Ducate and Lomicka 2009 for cons), and 2) learners attain fluency, but retain accent on the comparable phonological elements in line with the previous literature by Moyer (2004). **(Kangwon National University)**

1. Introduction

1.1 What is an accent reduction course?

Accent reduction is a systematic approach used to learn or adopt a new accent. It is the process of learning the sound quality and the sound system of a language or dialect. The methodology involves several steps, including identifying deviations in the person's current speech pattern from those of the desired accent in terms of using consonants and vowels as well as syllables, stress, and intonation. Clients learn to change the way they use

---

their mouth, teeth, and tongue to form vowel and consonant sounds; to modify intonation and stress patterns; and to alter the speech rhythm between syllables and stress. For example, second language learners can alter their speech to resemble the accent of what is considered standard in the language they are learning, thus enhancing clarity.

Accent reduction has often taken the form of classroom instruction or speech therapy. The term varies between "accent reduction" or "accent modification," depending on whether the focus is on losing the accent or gaining the accent, and between "training" or "therapy," depending on whether the intervention is offered by a teacher or a therapist. For further details, refer to an overview of the accent reduction therapy in Kim (2010: 56-57, 65-68). For purposes of this research, an accent reduction course was offered to adult Korean language learners whose native language was American English. The speech of these learners was evaluated in terms of various phonological properties in the learner's non-native accent.

1.2 Background

While accent reduction is offered in classroom instruction (Morales Pech and Izquierdo 2011), internet markets (Blommaert 2009), and speech therapy practices (Kohnert et al. 2003, Schmidt 1997, Stockman et al. 2008, Wolfe et al. 2003), non-native sounding accents are known to persist regardless of whether there is some treatment efficacy (e.g. Morales Pech and Izquierdo 2011) or not (e.g. Ducate and Lomicka 2009). Longitudinal results are particularly important, because non-native accents among adult learners due to age constraints (Scovel 1988) can prohibit the attainment of native-like L2 pronunciation (Lenneberg 1967). That is, studies have shown that although many learners attain fluency, many more retain a noticeable accent (Moyer 2005: 45), and there is correlation with Voice Onset Time and sounding non-native (Schmidt and Flege 1995).

However, research has not yet shown statistically significant comprehensive outcomes on phonological properties despite the numerous accent reduction courses taught by teachers, use of accent reduction software, or accent reduction work done by speech therapists. This is because there have not yet been enough pre- and post-test results accumulated for reliable analysis which validate the accent reduction studies. Most importantly, to my knowledge, there has been no study of pronunciation improvement that comprehensively tests major phonological elements. The purpose of this study is to investigate if there is instruction efficacy in an accent reduction course, as attested to by an evaluation of the overall phonological aspects of a person's speech.

1.3 Research question

In order to investigate if there is instructional efficacy in an accent reduction course, a post-test given after as little as five hours of instruction was analyzed to see if results were significantly better than pre-test results given before instruction commenced. If the results were significantly better, they would support that instruction efficacy occurred. If results were not significantly better, then instruction efficacy would not have occurred. To determine this, this study measured the major phonological properties of the English non-native speakers learning the Korean language.

(1) Research question

Given an accent reduction course with as little as five hours of instruction, can learners after instruction receive a significantly higher score in listening and pronunciation tests than before instruction?

To test the efficacy of an accent reduction course on non-native speakers, an experimental pronunciation class was offered that was specifically designed to test the improvement of learner speech with respect to the following major phonological elements: 1) consonants, 2) vowels, 3) phonological adjustments, 4) intonation, and 5) rhythm. To control the practice effect on the post-test, learner performance was compared with the performance of a control group made up of native speakers. To increase the validity of the results, non-native accent properties were studied in terms of accuracy, fluency, and heavy or light foreign accent.

2. Method used in the accent reduction course

This section describes the participants, speech materials, and procedure used in the accent reduction course.

2.1 Participants

There were 100 research participants, excluding the model speakers in the textbooks. These participants included one teacher (the author), two newly recruited model speakers of both genders to add variety to the model voice used in the testing, 32 American English speakers who were Korean language learners at various levels of proficiency, a control group of nine native Korean speakers who took the same proficiency pre- and post-tests, and 56 Korean native speakers who evaluated learner speech patterns.

The 32 American English speaking participants were all volunteers, who responded to an advertisement at the University of Hawaii. There were four novice learners, 14 low proficiency learners (L2 Low, henceforth), and 14 high proficiency learners (L2 High, henceforth). Table 1 describes

the information for both these learners and the native speaker control group in terms of proficiency, heritage, age, and gender.

Table 1. Speaker Information (Proficiency, Heritage, Age, Gender) by Group

	Native Speaker (N=9)	L2 High (N=14)	L2 Low (N=14)	Novice (N=4)
Listening [a]	97% (3.5%)	80% (3.6%)	62% (8.9%)	Incapable [b]
Reading [c]	Native	Good	Capable [d]	Incapable
PronGood [e]	6.9 (0.1)	3.6 (0.5)	2.8 (0.5)	2.8 (0.2)
HertgSpkr [f]	9 (100%)	2 (14%)	0 (0%)	0 (0%)
AGE [g]	31.2(14.7)	21.5 (5.4)	29.2 (14.5)	29.3 (11.6)
Gender	5f,4m	7f, 7m	10f, 4m	2f, 2m

a. Pre-test listening score (standard deviation in parenthesis); *b*. Novice speakers were incapable of taking the test, as they do not understand the text (Chance level score is 48%); *c*. Reading knowledge of Korean text in Hangeul (Korean alphabet); *d*. All learners could read the Korean text, but six of these learners preferred to read the Romanized transcript (four people in the pretest, six people in the post-test) or were able to read along with the model speaker (two people in the pretest, nobody in post-test); *e*. Average rating of good pronunciation in terms of accuracy and fluency on the pre-test measured by a 7-point scale from bad to good (standard deviation in parenthesis); *f*. Heritage speakers (percent of all speakers in parenthesis); *g*. Current age (standard deviation in parenthesis).

In Table 1, the listening scores of the pre-test are differentiated by group in that the native speakers correctly answered 97% of the time, high proficiency learners (L2 High) 80% of the time, and low proficiency learners (L2 Low) 62% of the time. Novice learners who had no background in the Korean language were not able to take the listening or pronunciation pretests in Korean. However, these novice learners did take the post-tests in both components. The test results will be shown later in Figure 5. Low-level learners who were able to read the Korean alphabet, had listening pretest scores as low as chance level ($48 \pm 5\%$) to as high as 72%. The high-level learners who were able to comprehend listening and reading Korean well, had listening test score between 75% to 86%. This group included two heritage speakers, whose parents are both Korean. The Korean native speakers were born and raised in Korea during childhood, and speak Korean more than English in their daily lives. Their listening test scores were 93% to 100%.

In addition to the experimental subjects described in Table 1, there were 56 evaluators rating the speech of the experimental subjects. Of these evaluators, 23 evaluated the accuracy and fluency of the pronunciation pre- and post-tests of both the learners and native speakers, and 33 evaluated the non-native pronunciation accents on the pre- and post-tests of both the learners and the native speakers.

Neither the native speaker control group nor the learners had noticeable regional accents in their own native languages. The learners spoke Mainstream American English, while the native speakers spoke Standard Korean as the native language.[1] All participants were college educated.

2.2 Speech materials

There were two types of speech materials utilized in this study: recording materials for pronunciation and listening pre- and post- tests and teaching materials for class instruction.

2.2.1 Speech materials for pre-tests and post-tests

The test elements were chosen based on the greatest variety of phonological elements offered from an open source (Choo and O'Grady 2003) to facilitate potential cross-checking of data by other researchers. The listening test was composed of 104 questions taken from the first two comprehension exercise items in Choo and O'Grady (2003). Each question consisted of a set of two or more answers: the correct answer, a distractor word or phrase taken from the distractor condition (the foil answer), and sometimes other plausible answers. The correct answers in this research were randomly chosen and were not the same as in the textbook. The pronunciation test consisted of these utterances along with the correct answer. Of these, 25 questions were selected from all sections in the textbook. The remaining 79 questions were fillers so that learners would not know which items would actually be used in this experiment.

The content validity and face validity of this research was reinforced by the way the speech items were chosen. Content validity was enhanced by covering major aspects in Korean pronunciation. Face validity, was assured by choosing the items from a textbook authored by an established faculty at the learners' university.[2]

2.2.2 Speech materials for pronunciation and listening comprehension

Before and after instruction, the same set of 104 test items were used for both listening and speech production. The post-test on listening comprehension was the same set of questions and possible answers, but the correct answers were in random order. That is, the correct answers were possibly the same or different from the pretest out of the same choice sets. The pronunciation test materials were the same as the answers to the listening pretest, so that learner production could be compared across items.

---

[1] Although the learners were from various parts of the United States, many had changed their residence and none had noticeable regional accents.
[2] Face validity occurs when learners feel the test is valid, and keeping their own motivation high, because people are likely to try harder when the test seems reasonable.

Romanization of the Korean alphabet was provided to subjects because some learners were not able to read the Korean Alphabet well.

2.2.3 Speech materials for classroom instruction

The teaching materials in class were mainly from the textbook (Shin 2008) and Korean pronunciation in the SORIDA database (Kim 2001)[3], supplemented by all 104 questions from Choo and O'Grady (2003) that were used for the pre- and post-tests. All the students were given the textbook and the accompanying audio CD, and another audio CD with the pronunciation of five model native speakers for all the possible answers to the 104 test items. Classroom materials with audio files were also given to learners, and were downloadable from the class homepage. The five model speakers were: two newly recruited speakers of both genders for this class, the two speakers of both genders in the book by Choo and O'Grady (2003) and the teacher.

The 25 actual research items included five items from each of the major phonological aspects of the Korean language: consonants, vowels, segmental rules, intonation, and rhythm. The classroom activities mainly worked on the difference between the native and target languages. These activities consisted of using minimal word pairs in a natural context whenever possible. Students listened to and tested on some of these classroom materials on their listening test, and read and recorded the same item on the pronunciation test.

2.2.4 Consonant data

The data displayed below were selected for the Korean consonant recording list as marked by the italicized Korean letter in parentheses. The words in parentheses in non-italic letters are the competing items on the listening tests. The two words in parentheses are minimal word pairs, and were put into meaningful contexts whenever possible. The selection criteria included maximum consonant variety in the learner's speech. For each item, a transcript in the Korean alphabet was given first, and then the Romanization followed. The last line is the translation into English. Learners were given all three representations for their tests.

(2) Recording list for consonant data
    a. (*벼*, 뼈)가 부러졌어요. (byeo, ppyeo)-ga bureojeosseoyo.
       (rice plant, bone) has been broken.
    b. 누구 (*턱,* 덕)이야? nugu (teok, deok)-iya.
       (Treat, Thanks) to whom is that?
    c. (*까지,* 가지) 마세요. (kkaji, gaji) maseyo.

Please don't (*peel it, go*).
d. 밤에는 (*자*, 자)요. bameneun (cha, ja)-yo.
   At night, it is (cold, sleeping).
e. 여기서 (*서*, 써)요. yeogiseo (seo, sseo)-yo.
   Please (stop, buy) here.

The list above contains consonants articulated in different glottal states: two plain consonants in (2a) and (2e), two aspirated consonants in (2b) and (2d), and one tense consonant in (2c). This list also contains consonants articulated in different places and manners: three bilabial, alveolar, and velar stop positions in (2a), (2b), and (3c) stops, respectively; one affricate consonant in (2d), and one fricative consonant in (2e). This list also contains consonants in various positions in a sentence (i.e., intonation phrasing): two sentence initial positions in (2a) and (2c) and three sentence medial positions in (2b), (2d), and (2e).

2.2.5 Vowel data

The data below were selected for the vowel recordings for the tests. The answers are in italicized Korean letters in parentheses. The criteria for selection included the maximum variety of vowel sounds in the learner's speech. The two words in parentheses are minimal word pairs, put into meaningful context whenever possible. The learners were told that the answers might not be that which they expected.

(3) Recording list for vowel data
   a.(*극*, 국)이 싱거워요. (geuk, guk)-i singgeoweoyo.
      (pole, soup) is not salty enough.
   b. (*굴*, 글)을 씁니다. (gul, geul)-eul sseumnida.
      I'm writing/using (oyster, novel).
   c.(*먹*, 목)이 짧아요. (meok, mok)-i jjalbayo.
      (ink stick, neck) is short.
   d. (*볼*, 벌)에 쏘였어요. (bol, beol)-e ssoyeosseoyo.
      I got stung (on the cheek, by a bee).
   e. 같은 (*반*, 번)이에요. gatheun (ban, beon)-ieyo.
      We're in the same (number, class).

The data above includes the maximum variety of vowel sounds such as the high back unrounded vowel "으" [ɯ] in (3a), the high back rounded vowel "우" [u] in (3b), the mid back unrounded vowel "어" [ʌ] in (3c), the mid back rounded vowel "오" [o] in (3d), and low back unrounded vowel "아" [ɑ] in (3e). All these vowels occur in the initial positions of CVC syllables in mono-morphemic words. The data contains various suffixes that function as: two nominative case markers in (3a) and (3c); one accusative case marker in (3b); one locative in (3d); and one copula in (3e).

The data also contains vowels in the sentence-medial position as in (3e) and the sentence initial positions as in (3a), (3b), (3c) and (3d).

2.2.6 Data in phonological adjustments

The data below were selected for recording phonological adjustments as marked by italicized Korean letter in parentheses. The selection criteria included the maximum amount of phonological adjustment in learner speech. The two words in parentheses are also minimal word pairs, although only some of which were made into meaningful contexts: (4a), (4b), and (4d). Only consonant changes are shown in these data, because, to our surprise, a suitable vowel adjustment rule in Korean could not be found.

(4) Recording list for adjustment data
  a. Korean <u>allophonic</u> rule of /s/-palatalization in 실 [ʃil] that will bring a <u>phonemic</u> change in English; and English <u>allophonic</u> rule of Velarization in *wool* [velar l] and 을 [velar l] that will bring an <u>allophonic</u> change in Korean.
  (*실리학을*, 심리학을) 공부합니다.
  (sillihak, simnihak)-eul gongbuhamnida.
  I study (utilitarianism, psychology).
  b. Korean <u>allophonic</u> rule of Voicing as in 간 [voiceless k] and 격 [voiced g] that will bring a <u>phonemic</u> change in English; and English <u>phonemic</u> rule of Place Assimilation as co<u>n</u>crete [ŋk] and *간격* [ŋk] that will bring a <u>phonemic</u> change in Korean.
  (*간격*, 감격)이 커요.
  (gangyeok, gamgyeok)-i keoyo.
  (The space in between, thrill) is big.
  c. Korean <u>phonemic</u> rule of /n/-insertion as in 일 [nil] that will bring a <u>phonemic</u> change in English.
  (웬일[*웬닐*], 웨닐)이에요?
  (wennil, wenil)-ieyo?
  What's up?
  d. Korean <u>phonemic</u> rule of Obstruent Tensing in 비 [tense p] that will bring an <u>allophonic</u> change in English; and English <u>phonemic</u> rule of Geminate Simplification as in *summer*[single m] applied to *잡비* [double p] in Korean that will bring a <u>phonemic</u> change in Korean.
  (*잡비*, 자비)가 들어요.
  (japbi, jabi)-ga deureoyo.
  It requires (miscellaneous expenses, mercy).
  e. Korean <u>phonemic</u> rule of Nasal Assimilation in 단 [dan] that will bring a <u>phonemic</u> change in English.
  바닷물[*바단물*, 바닫물]은 짜죠.

(bada***n***mul, badatmul)-eun jjajo.
Sea water is salty, of course.

The data above included the maximum variety of phonological rules: Palatalization and Velarization in (4a), Voicing and Place Assimilation in (4b), /n/-insertion in (4c), Obstruent Tensing and Geminate Simplification in (4d), and Nasal Assimilation in (4e). They also include both Korean and English phonological rules: five Koran phonological rules in (4a), (4b), (4c), (4d), and (4e), and three English phonological rules in (4a), (4b), and (4d). The data also contains various combinations of phonemic and allophonic rules: Korean-phonemic and English-phonemic adjustments in (4b), (4c), (4d) and (4e); Korean-phonemic and English-allophonic adjustments in (4d); Korean-allophonic and English-phonemic adjustments in (4a) and (4b); and Korean-allophonic and English-allophonic adjustments in (4a).

2.2.7 Intonation data

The data below were selected for recording differences in intonation. The selection criteria included the maximum variety of intonation patterns in learner speech: (5a) for focus and emphasis, (5b) for Yes-No question intonation, (5c) for Wh-questions, (5d) for exclamations and statements, and (5e) for commands.

(5) Recording list for intonation data
 a. Focus and emphasis
  A: 도서관에 책 빌리러 가요? B: 아뇨, 도서관에 *일하러* 가요.
  A: Doseogwone chaek bilireo gayo? B: Anyo, Doseogwane ilhareo gayo.
  A:*Are you going to the library to borrow books?* B: *No, I'm going to the library to work.*
 b. Yes-no question
  저 영화 몇 번 봤어?
  jeo yeonghwa myeotbeon bwasseo?
  Have you seen the movie?
 c.Wh question
  오늘 어디 가요? oneul eodi gayo?
  Wh: *Where are we going today?*
 d. Exclamation and statement
  정말 맛있다.jeongmal masitda.
  It's really delicious.
 e. Command
  여기 서요.yeogi seoyo.
  Stop/stand here.

The expected intonation of the data in (5) is the following. Intonation of (5a) is to test focus and emphasis, and learners are expected to place prominence (High tone or loudness) on the emphatic information *ilhareo* 'to work.' Placing prominence on a different word would be pragmatically wrong and hamper communication. The intonation of (5b) is to test Yes-no questioning, and the learners are expected to place a high tone at the intonational phrase boundary (H%), i.e., at the end of the sentence (Jun 2000). Notice that a different tonal pattern may be misunderstood as a Wh-question instead, because (5b) includes a Wh-marker *myeotbeon* 'how many times.' The intonation in (5c) is to test a Wh-question, and learners are expected to place a high tone followed by a low tone at the end of the sentence (HL%), as outlined in Jun (2000). Placing a different tonal pattern here may incorrectly be interpreted as a Yes-no question or a command[4]. The intonation in (5d) is to test exclamation and statement, and the learners are expected to use a falling boundary tone 'HL%' with proper H tone on the emphatic word *jeongmal* 'really.' The intonation in (5e) is to test a command, and the learners are expected to place a break right in front of the word *seoyo* "stop" or "stand". If the break is not placed properly, then the meaning of the utterance will be misunderstood as "Here." The tonal pattern would be a falling-rising boundary tone (HLH%) is used when a speaker is expecting listeners' agreement (Jun 2000).

2.2.8 Rhythm data

The data below were selected for recording rhythm. Due to the difference in speech rhythm between the syllable-timed Korean language and the stress-timed English language (Kim et al. 2007), a native speaker of English learning Korean may sound unnatural by alternating strong and weak syllables when speaking Korean. Data in (6) are marked by the expected stress-timed rhythm for the alternation of strong (S) and weak (w) syllables. The selection criteria included the maximum rhythm patterns expected in stress-timed learner speech: (6a) for 'Sw', (6b) for 'SwSwS', (6c) for 'SwSwSw', (6d) for 'SwwSww', and (6e) for 'SwSwSwSww'.

(6) Recording list for rhythm data
    a. Sw(w/s)
      비싸요. bissayo.
      It's expensive.
    b. SwSwS
      금리 비싸죠? (geumni) bissajo?
      Is (interest) high?
    c. SwSwSw

---

[4] These intonation data in 5(b) and 5(c) turned out to be too difficult for native Korean speakers as well. Nearly 20% of the native speaker ratings were invalidated by too much difference (3 ranks or more) from the average.

자고나서 할게. jagonaseo halge.
I will do it after sleeping.
d. SwwSww
후추 좀 주세요. huchu jom juseyo.
Please give me the black pepper.
e. SwSwSwSww
실리학을 공부합니다. (silli)hageul gongbuhamnida.
I study utilitarianism.

The correct rhythm for the data above is syllable-timed rhythm, while the less correct rhythm includes alternating stress. Learners with the poorest command of rhythm have no rhythm and are somewhat monotonous and hesitant when reading.

2.3 Procedure

The experimental procedure used on the data consisted of nine stages: 1) the pre-test on listening comprehension, 2) the pre-test on pronunciation, 3) lectures on Korean pronunciation, 4) feedback given to the students on the pre-test during second and third week of the five-week course, 5) stopping of lectures on Korean pronunciation, 6) the post-test on listening comprehension, 7) the post-test on pronunciation, 8) the assessment of the listening comprehension test results, and, 9) the native speaker evaluation of pronunciation tests in terms of accuracy, fluency, and native-likeness.

The learners, first took the listening pre-tests within one week prior to instruction. Learners chose the correct answers on test sheets for the listening comprehension test as outlined previously. For intonation, learners listened to the recorded speech and chose the correct meaning on the answer sheet. Novice learners did not take the listening pre-test, as they could not read and answer in written Korean, even in Romanized form. The average time spent on the pre-test listening was 20 minutes, as the learners were tested on 104 question items.

Learners, then took the pronunciation pre-test. The test-materials were the same as the listening pre-tests, and the learners were given the correct answers to the listening pre-tests and asked to read them. Using the same material in the listening test increased the reliability of the test as learners were tested on the target speech they had listened to at least once. Learners were told that they could repeat their utterances and that the last utterance would be used as test data. This is done to reduce the anxiety of the leaners and, thereby, enhancing the reliability of the research. Novice learners took a different set of pronunciation test in Romanized form from the test set that rest of the speakers took. The novice test set was easier and shorter than the regular test set.

The digital recording was in a 16 *KHz*, 16-bit format, and produced in a quiet room using an Audio-Technica AT8033 microphone. The distance

between the microphone and the mouth of the speaker was adjusted by the amplitude of the voice displayed through a waveform signal. Utterances with errors or hesitations were allowed to repeat. All recordings by learners were completed within a week of instruction.

After these recording were completed, the lectures began. Each of the five sessions of the course covered one of the five aspects of phonology: consonants, vowels, segmental rules, intonation rules, and rhythm rules. The class webpage was open to learners to download class materials.

The class size was under 20 students, and two classes were offered. Some effort was made to classify students in terms of their proficiency levels, with one of the two classes for learners with some Korean language background, and the other class for students at varied L2 language levels including all novice learners. Because these students were volunteers and had different daily schedules, it was not possible to further divide the classes.

During the instruction, the teacher explained the phonological differences in English and Korean, and the learners listened to and repeated after the modeled sound. Further, the model sound was distributed via a CD, and available through web-downloads for optional practice at home.

Learners then received written feedback on their pronunciation pre-tests. The feedback included what sounds were incorrectly pronounced and how learners could improve their pronunciation. The learners were invited to the instructor's office to get personal tutoring on the given pronunciation. The classroom instruction on pronunciation lasted for five weeks.

At the end of five weeks, instruction stopped. Ninety-four percent[5] of the learner volunteer participants remained and attended at least four out of five lectures, and successfully completed the four component pre- and post- tests on pronunciation and listening comprehension. The two missing volunteer were learners who could not schedule individual test-times due to a series of holidays and vacations around the post-test period. These two missing volunteers were excluded from this study.

After instruction stopped, participants were tested again on their listening comprehension by choosing correct answers to a series of recorded questions. All novice learners took the post-test, although they could not take the pre-test because of their lack of knowledge of the Korean language. The listening comprehension post-test was the same set of questions and possible answers that was on the pre-test, but the correct answers were randomized. That is, the correct answers was the same or different from the pretest, based on the same choice sets. The model speakers used to record the listening materials were different from those used in the pre-tests. This was done to increase the reliability of the post-test so learners would not rely on their memory of a particular speech

---

[5] Thirty four learners showed up at the second session of the class after completing pre-tests. Of these, only two learners were disqualified as they did not take the post-test pronunciation.

pattern, but would rely on their phonological ability to comprehend a new language.

Following this, participants were asked to read and record the same test materials as the pre-test again using the same recording rooms. All these post-test recordings were completed within a week after instruction was stopped.

In contrast with the procedure described above, the native speakers in this study did not have any lessons or training sessions in English. They took the listening pre-test and pronunciation pre-test immediately followed by the listening post-test and pronunciation post-test immediately. The native speaker control group only took the 32 listening items and 25 recording items used for the experiment, instead of the whole list of 104 listening and pronunciation test items that the learners took. All of the native speakers acknowledged that they understood the meanings of all recorded sentences as well as their tasks.

The data acquisition procedure was carefully designed to enhance the reliability of this study in that the tests were not affected by unwanted influences induced from examinee, examiner-scorer, test content, situation, and time. Examinee-incurred influence occurs when learners are motivated to feel that they are ready, thereby they want to do their best on the test. This was achieved by letting the learners know that the more they improved, the better my research result would be. All lectures given in the two classes and the test administration were done by one teacher to prevent examiner-scorer influence. As mentioned previously test content influence was avoided by choosing different speakers to record the pre-test, post-test and in class material. Situation-induced reliability problem was reduced by supplementing all reading material on the test and in class with Romanization so that low level learners would not fear the reading difficulty of the Korean alphabet, Hangeul. In addition, the listen and repeat method was available for pronunciation tests for those who felt insecure about their reading ability. The instructor sat outside of the testing room rather than inside for the pronunciation test, and for those learners who opted to self-record without any other person around. Additionally, cheating was not allowed as examinees were asked whether or not they studied the book before the pre-test. Finally, time-influenced reliability was assured by strict control of time for all the pre- and post-tests.

The researcher acquired a large corpus, 2,496 comprehension test records, and 2,073 speech production recordings from all learners and nine native speakers of Korean.[6] There was some missing data from not responding or not recording.

---

[6] 2,496 comprehension test records came from [32 items * ((28 learners + 9 NS controls) * 2 pre- and post-tests + 4 novices * 1 post-tests))]. On the other hand, 2,073 speech production recordings came from [5 items * 5 phonological elements * (28 learners + 9 NS controls + 4 Novices) * 2 pre-post tests + 1extra post-test recording of read speech after the recording of modeled speech – 2 missing data from not recording.).

3. Analysis of the test data on learner improvement

The listening comprehension tests were evaluated by research assistants in terms of correct identification among perceptually similar sounds, while the pronunciation tests were rated by 56 Korean native speakers in terms of 1) accuracy, 2) fluency, and 3) non-native accentedness.

The study employed a number of measures to ensure the reliable evaluation of the learner pronunciation quality. First, the study normalized the amplitude of all learner speech data so that the evaluators would not consider a loud voice as being more proficient. Recordings were done individually, and loudness differed depending on the speaker, time, and situation. Secondly, the learners' pronunciation was intermixed with native speaker pronunciation and randomized. Thirdly, all speech data were randomized in terms of speakers and pre- and post-test recordings so that the raters would be blind to the speaker information as well as pre- or post-test recordings. Fourth, the raters worked on the same recording list in the same test set taken by the speaker groups of NS, L2 High, and L2 Low (Korean native speakers and the learners of high and low proficiency). These raters did not rate the utterances by novice leaners who took a different test set. Instead, one trained phonetician (the author) rated all the novice speech data. Fifth, the duration of each rating session was strictly controlled for 10-15 minutes, to lessen the cognitive load on the raters. Break time between each session lasted for at least three or five minutes, with an additional 15 minutes' break after every hour, followed by another 60 minutes' break after three to four hours. No rater worked for more than seven hours during a day, including these breaks. Sixth, the raters took the baseline test to see if they were capable of rating the learner speech data in a reliable manner. The baseline test used different sets of sound material from the actual ratings of learner speech data: Novice recordings for accuracy and fluency ratings, and dog-barks for accentedness ratings. Seventh, we prevented invalid data from being entered into a cell, by not responding, not reading, or rating the opposite to the rest of the raters.[7]

3.1 Evaluation of listening comprehension

Listening comprehension tests were evaluated by grading correct answers from the answer choices. For expositional convenience, a sample question was reproduced from (3d). The correct answer was marked by an asterisk (*).

---

[7] The difference of 3 ranks or more from the average marks the complete opposite judgment to the rest of the raters as the maximum possible difference is 6 ranks in the 7 rank scale. This is to invalidate rating mistakes by losing concentration or marking in a wrong cell. We used the average score as a reference tool as the number of raters are sizable: 23 native speaker raters for L2 accuracy and fluency ratings, and 33 native speaker raters for non-native accentedness.

(7) Sample answer sheet for listening comprehension test
   (*벌, 볼)에 쏘였어요. (*beol, bol)e ssoyeosseoyo.
   *I was bit by/on the (*bee, cheek)*

In (7) the answer is *beol* 'bee,' and the evaluation of the listening comprehension is done by counting the numbers of correct answers. We gathered 2496 data from the 41 speakers listening to 28 question items.[8]

3.2 Evaluation of pronunciation accuracy

The recorded speech data was randomized and rated by Korean native speakers in terms of 1) accuracy, 2) fluency, and 3) non-native accentedness. 23 raters evaluated on the accurate production of phonological elements in terms of consonants, vowels, segmental adjustments, intonation, and rhythm. Six research assistants were included among the raters. The raters ranked each phonological element on a 7-point scale with, 1 equal to non-native novice level, and 7 equal to native-like level. The rating scale was actually the following: 1. "definitely incorrect" non-native, 2. "somewhat incorrect" non-native, 3. "borderline correct and incorrect" non-native, 4 "somewhat correct" non-native, 5. "definitely correct" non-native, 6. "probably" native, 7. "definitely" native.

For evaluation purposes, learner pronunciation data were intermixed with native speaker pronunciation data and randomized without pairing pre- and post-test of the same speaker. The randomization resulted in playing utterances in the order that any pre- or post-test speech data of any random speaker followed any other speech data.

The procedure used for rating the accuracy of the recorded speech data using the base-line test with two pairs of novice utterances, then utilizing 25 rating sessions that lasted 9-15 minutes each, for rating a total of 1873 data from the speaker groups of NS, L2 High, and L2 Low.

3.3 Evaluation of speech fluency

The same 23 raters evaluated the overall fluency of all recorded utterances. To increase the reliability of the rating results, the methods and procedures used by these same raters were kept as similar as possible for both accuracy and fluency ratings. The only difference was the targeting evaluation content: one being the accurate production of specific phonological features, the other being the fluency of overall utterance. For each utterance, the raters evaluated first the accuracy of the given phonological feature and then the fluency of the overall utterances.

The raters ranked the learners' fluency on another 7-point scale with, 1

---

[8] The 41 speakers consisted of nine native speakers, 14 L2 High, 14 L2 Lo, and 4 novices. The 28 listening items correspond to 25 pronunciation items with additional three to make the question format coherent.

equal to non-native novice level, and 7 equal to definitely native level. The rating scale was actually the following: 1. "novice" non-native, 2. "low" non-native, 3. "intermediate" non-native, 4 "high" non-native, 5. "excellent" non-native, 6. "probably" native, 7. "definitely" native.

The procedure used for randomizing and rating the speech data for fluency was the same as described in the accuracy evaluation above. The same 23 raters took the base-line test with two pairs of novice utterances, then 25 rating sessions that lasted 9-15 minutes each to rate a total of 1,873 data sets from NS, L2 High, and L2 Low speaker groups.

3.4 Evaluation of non-native accentedness in pronunciation

A different set of 33 Korean native speaker raters were recruited to rate the degree of apparent non-native accent. These raters ranked the degree of apparent non-native accent in the whole utterance by answering two questions: Is one more natural than the other? (Yes / No); If so, which one? (1st / 2nd). The raters were told that the naturalness of speech was related to heavy or light non-native accent of the learner. They were also told that the more natural meant more native-like with less non-native accents. For evaluation purposes, learner pronunciation data were intermixed with native speaker pronunciation data and randomized in two ways: first within the pairs of the pre- and post-test of the same speaker, and second among the pairs by different speakers. The randomization resulted in playing pairs of utterances of a same speaker in a row, but mixing the pre- and post-test sequencing.

The rating procedure was slightly different from accuracy and fluency ratings: Base-line testing with two dog-barks, with 10 rating sessions that lasted 9-15 minutes each rating 948 pairs of data from the NS, L2 High, and L2 Low speaker groups.
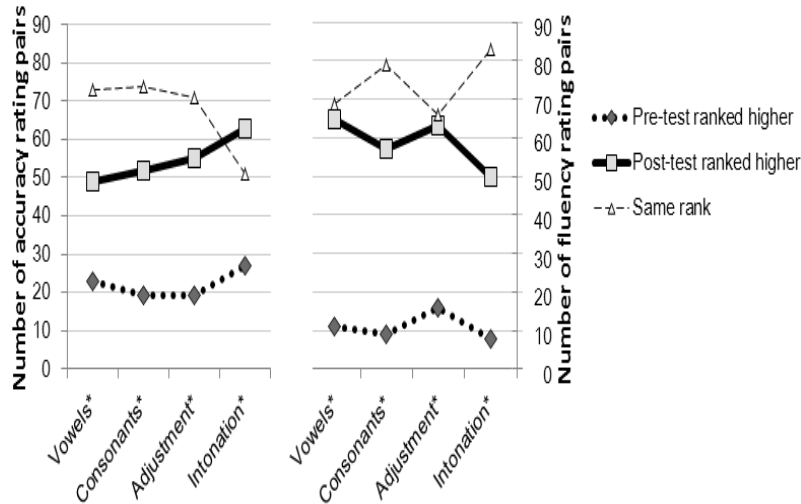
4. Results

The results show that when comparing learner speech to native speaker speech before instruction, learner speech had a significantly lower score in listening and pronunciation than after instruction.

4.1 Phonological accuracy versus overall fluency

Figure 1 shows the pronunciation accuracy and fluency rating results for learner speech on the pre- and post- tests. The results are shown in terms of different phonological elements: consonants, vowels, adjustments, intonation, rhythm, and fluency.
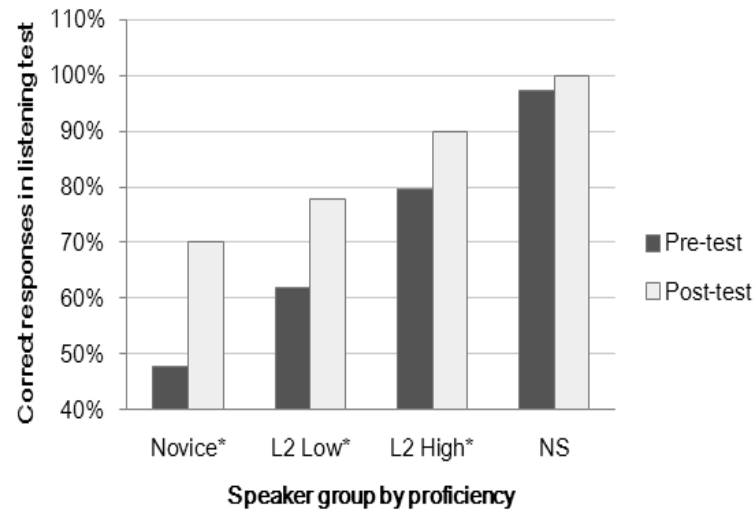
Figure 1. Number of rating pairs for pronunciation accuracy and fluency in pre- and post- tests of learner speech show that the raters gave higher points on the post-test than the pre-test (n=34,454 pairs; *p<0.05). Fluency improved significantly more than Accuracy (33.2%>22.7%; p=0.01).

In Figure 1, The Korean native speaker raters gave higher points for learner speech on post-tests than pre-tests. Large improvement occurred in fluency, but not so much in accuracy of the vowels, consonants, and phonological adjustments. The accuracy and fluency ratings in post-tests (the thick line with rectangles) are significantly more higher ranked than the ratings in pre-test (the thin dotted line with diamonds). The results positively answer our research question in (1), because learners after instruction received a significantly higher score in accuracy and fluency ratings than before instruction. We did not report the rhythm data in Figure 1, as many NS raters showed inconsistent ratings. The rhythm results, however, are coherent to Figure 1. Rhythm ratings are reported in all other results from Figures 2 to 5.

Moreover, the researcher examined the results of the listening test. Figure 2 shows the improvement in listening comprehension by learners when compared with native speakers.
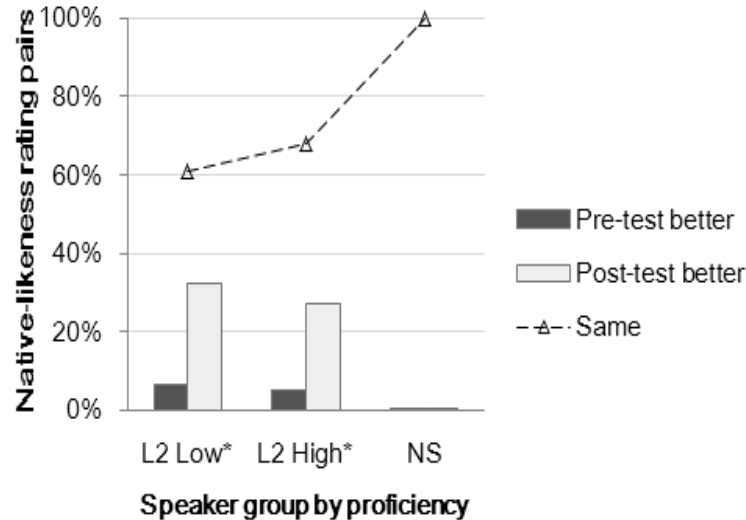
Figure 2. Correct responses for the listening comprehension test results before and after the instruction (n=2,496, *p<0.05). Learners at all levels show significantly more correct responses to the post-test than the pre-test. (*Notes:* NS= Native Speaker; L2 High = Learners with high proficiency; L2 Low = Learners with low proficiency; Novice= Learners with no background in the Korean language.)

In Figure 2, the listening comprehension test results after instruction were significantly higher at all learner levels than those before instruction. The pretest results of the novice group were calculated by the chance level of accuracy. The native speaker group also showed some improvement, which reflects the practice effect. The results positively answer our research question in (1), because learners after instruction received a significantly higher score on the listening comprehension test than before instruction.

To further investigate course efficacy, the researcher also examined the rating results of the degree of native-likeness of learner speech.

4.2 Accent reduction in learner speech

Figure 3 shows that the native-likeness rating results of learner speech in post-test is better than those in the pre-test.
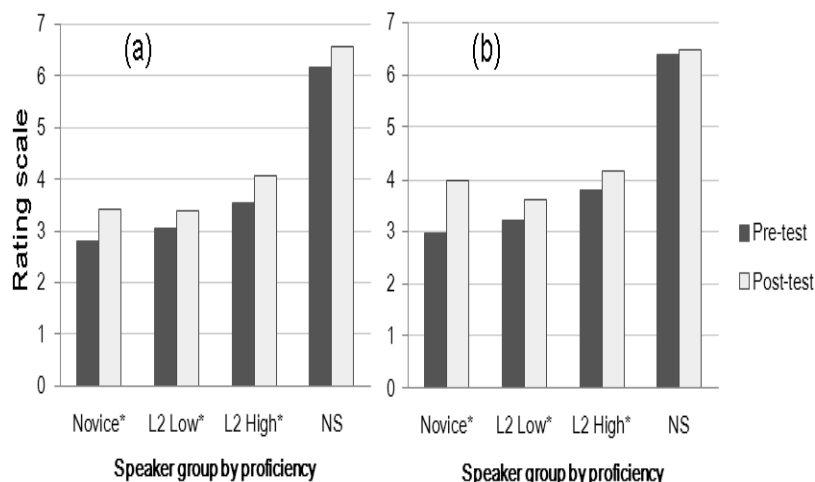
Figure 3. Native-likeness rating results of learner and native speech show that the raters considered larger proportion of post-test pronunciation more natural than the pre-test pronunciation *(n*=948 pairs; *p<0.05). (*Notes:* L2 High = Learners with high proficiency; L2 Low = Learners with low proficiency)

Figure 3 above shows that native speaker raters chose a significantly larger portion of learner speech to be more natural in the post-test recordings than the pre-test recordings. The raters accurately detected native speaker speech as opposed to learner speech in that native speech does not differ much in native-likeness in their post-test data and pre-test data (the dotted line with diamonds). The inter-rater reliability was good, as the average standard deviation among raters for the L2 High, L2 Low and NS groups was 0.47 (*n*=32,184). The results positively answer our research question in (1), because learners after instruction received a significantly higher score in native-like speech ratings than before instruction. To further investigate the course efficacy, the research examined the fluency ratings.

4.3 Improvement range of Accuracy and Fluency in learner speech

Figure 4 shows the accuracy and fluency ratings of native and learner speech in the scale of seven ranks.

Figure 4. Rating results of learner and native speech on pronunciation accuracy in (a) and fluency in (b) show that the raters assigned higher ranks on the post-test than the pre-test (*n=87,158; *p<0.05*). The improvement does not reach beyond Ranks 5 and 6, the borderline between non-native and native ranks (*Notes:* NS= Native Speaker; L2 High = Learners with high proficiency; L2 Low = Learners with low proficiency; Novice= Learners with no background in the Korean language.)

In Figure 4, learner speech fluency ratings are significantly better on the post-test (bars with light color) than on the pre-test for both groups of learners (L2 Low and L2 High). The inter-rater reliability was good, as the average standard deviation among the raters for the groups L2 High, L2 Low and NS was 0.73 (*n*=86,158). The rating of novice learners on a different test set was done by one trained phonetician (the author), and reported here for reference (*n*=100, *p*=0.04 by t-test for pre- and post-tests). The native speaker speech did not differ much, which indicates that the raters accurately detected native Korean speech. The results positively answer our research question in (1) that learner speech after instruction had a significantly higher score in accuracy and fluency than learner speech before instruction.

It is interesting to note that learner improvement did not cross over the ceiling between Ranks 5 (equal to non-native with high proficiency), and 6 (equal to perhaps native speech). To further investigate course efficacy, the researcher examined whether the evaluation results showed a correlation between the listening and pronunciation tests.

4.4 Relationship between pronunciation rank and listening test

Figure 5 shows whether the listening test results show some correlation with the pronunciation test results.

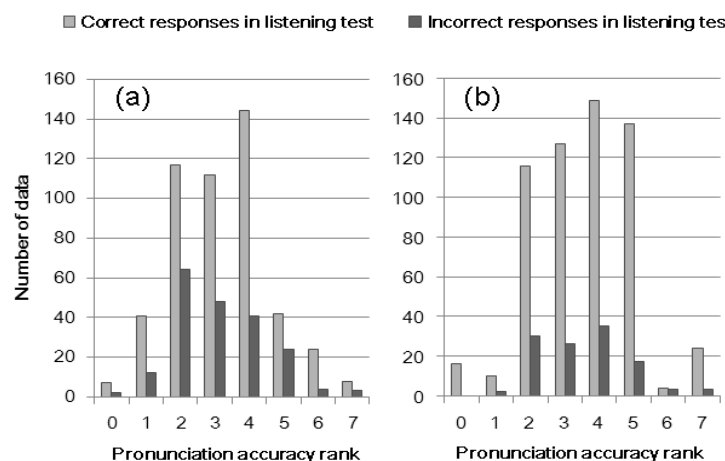■ Correct responses in listening test   ■ Incorrect responses in listening test

Figure 5. Results of the listening pre-test in (a) and post-test in (b) show that the learners' listening comprehension on the post-test is higher than on the pre-test for each pronunciation rank of accuracy. (83%>71%, *n* = 1,392, *p*=0.04).

Figure 5 shows that learners of the same pronunciation rank performed significantly better on the post-test (b) than on the pre-test (a) as they correctly answered larger proportion of listening test items (bars with light color). For computational consistency, the listening and speaking tests of novice learners were excluded, as they were unable to take the listening pre-test. Therefore, the data in Figure 5 represent the accuracy ratings for the speaker groups of L2 Low and L2 High, excluding novices and native speakers. Among the possible 1,400 hundred data (28 learners * 25 items * 2 pre-post readings), we invalidated 8 data, either because of inconsistent listening or pronunciation responses by native speakers or inconsistent ratings by native speaker raters.

The results positively answer our research question in (1) that learners' listening comprehension after instruction had a significantly higher score than before instruction. It is important to note that listening scores on the post-test are significantly higher than those on pre-test. In other words, improvement in listening scores leads to improvement in pronunciation scores.

5. Discussion

The results in Section 4 not only indicate the intended efficacy of pronunciation training, but some other interesting findings in learner phonology.

Figure 1 shows that the raters gave a higher rating for learner's accuracy and fluency on the post-test than the pre-test. Although improvement

occurred in accurate production of segmental vowel and consonant features, it was not as high as in fluency. Figure 2 shows that learners after instruction received a significantly higher score on the listening comprehension test than before instruction. Figure 3 shows that raters detected non-native accents in both pre- and post-test recordings though some improvement occurred. Figure 4 shows that learner improvement did not cross over the ceiling between Rank 5 (equal to 'non-native with high proficiency'), and 6 (equal to 'perhaps native speech').

The results indicate various points in non-native phonology development. First, phonological properties make non-native accents, while fluency has little effect. This is because the raters gave higher ratings on the post-test than the pre-test pronunciation tests in fluency, but not in vowel or consonant segments, as demonstrated in Figure 1. The results indicate that segmental features of consonants and vowels are what make a non-native accent, although fluency may improve a great deal.

Secondly, non-native accents are noticeable. This is because the raters detected non-native accents with great accuracy, as demonstrated in Figures 2, 3, and 4. The results support the ceiling effects demonstrated by Lenneberg (1967), Scovel (1969, 1988), Selinker (1972), Bley-Vroman (1989), and Flege (1992, 1995), but do not demonstrate the native-like achievement of non-native speakers in Ioup et al. (1994), Bongaerts (1999), and Bongaerts, Mennen, Van der Slik (2000).

Lastly, improvement in listening scores leads to improvement in pronunciation scores. This is shown by Figure 5 because listening scores on the post-test are higher than on the pretest for each pronunciation rank. This result aligns with Flege (1988:75), but not with Major (1995, 2001: 55).

REFERENCES

BLEY-VROMAN, ROBERT. 1989. What is the logical problem of foreign language learning? In Susan M. Gass and Jacquelyn Schachter (eds.). *Linguistic perspectives on second language acquisition,* 41-68. Cambridge: Cambridge University Press.

BLOMMAERT, JAM. 2009. A market of accents. *Language Policy* 8.3, 243-259.

BONGAERTS, THEO. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In David Birdsong (ed.). *Second language acquisition and the critical period hypothesis*, 133-159.Mahwah, NJ: Erlbaum.

BONGAERTS, THEO, SUSAN MENNEN, and FRANS VAN DER SLIK. 2000. Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica* 54.2, 298–308.

CHOO, MIHO and WILLIAM D. O'GRADY. 2003. The Sounds of Korean: A Pronunciation Guide, Honolulu: University of Hawaii Press.

DUCATE, LARA and LARA LOMICKA. 2009. Podcasting: An effective tool for honing language students' pronunciation? *Language Learning & Technology* 13.3, 66-86.

FLEGE, JAMES E. 1988. Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84, 70-79.

FLEGE, JAMES E. 1992. Speech learning in a second language. In C. A. Ferguson, L. Menn and C. Stoel-Gammon (eds.). *Phonological Development: Models, Research, and Implications*, 565-604. Timonium, MD: York Press.

FLEGE, JAMES E. 1995. Second language speech learning: Theory, findings, and problems. In W. Strange (ed.). *Speech perception and linguistic experience: Issues in cross-language research*, 233–277. Timonium, MD: York Press.

IOUP, GEORGETTE, ELIZABETH BOUSTAGUI, MANAL EL TIGI and MARTHA MOSELLE. 1994. Reexamining the Critical Period Hypothesis. *Studies in Second Language Acquisition* 16, 73-98.

JUN, SUN-AH. 2000. K-ToBI (Korean ToBI) labeling conventions. *Speech Sciences* 7.1, 143-170.

KIM, JONG-MI. 2001. Computerized sound dictionary of Korean and English. *Speech Sciences* 8.1, 271-296.

KIM, JONG-MI. 2010. Issues inhumanities therapy in Korea: An applied linguistics perspective. *Studies in Humanities* 24, 49-74. Kangwon National University Humanities Institute.

KIM, JONG-MI. 2012. Segments predict non-native accents: Prosody lends little. Presented at the CUNY Conference on the Segment in Phonology, City University of New York.
Downloadable from http://www.cunyphonologyforum.net/segconf.php

KIM, JONG-MI, SUZANNE FLYNN, and MIRA OH. 2007. Non-native speech rhythm: A large-scale study of English pronunciation by Korean learners. *Studies in Phonetics, Phonology and Morphology* 13.2, 219-250.

KOHNERT, KATHRYN, MARY R. T. KENNEDY, LESLIE GLAZE, PUI FONG KAN, and EDWARD CARNEY. 2003. Breadth and depth of diversity in Minnesota: Challenges to clinical competency. *American Journal of Speech-Language Pathology* 12, 259-272.

LENNEBERG, ERIC H. 1967. *Biological Foundations of Language*. New York: Wiley.

MAJOR, ROY C. 1995. Native and nonnative phonological representations. *IRAL-International Review of Applied Linguistics in Language Teaching* 33.2, 109–128.

MAJOR, ROY C. 2001. *Foreign Accent: The ontogeny and phylogeny of second language phonology.* New Jersey and London: Erlbaum.

MORALES PECH, DIANA and JESÚS IZQUIERDO. 2011. L2 phonology learning among young-adult learners of English: Effects of regular classroom-based instruction and L2 proficiency. *Revista Electronica de Investigacion Educativa* 13.1. Retrieved May 14, 2011 from: http://redie.uabc.mx/vol12no2/contents-moralesizquierdo.html

MOYER, ALENE. 2004. *Age, Accent, and Experience in Second Language Acquisition: An Integrated Approach to Critical Period Inquiry*. Clevedon, UK: Cromwell.

SCHMIDT, ANNA MARIE and JAMES EMIL FLEGE. 1995. Effects of speaking rate changes on native and nonnative speech production. *Phonetica* 52, 41-54

SCHMIDT, ANNA MARIE. 1997. Working with adult foreign accent: Strategies for intervention. *Contemporary Issues in Communication Science and Disorders,* 24. Retrieved May 15, 2011 from: http://www.nsslha.org/uploadedFiles/NSSLHA/publications/cicsd/1997 Working With Adult Foreign Accent Strategies.pdf

SCOVEL, THOMAS. 1969. Foreign accent: Language acquisition and cerebral dominance. *Language Learning* 19, 245-254.

SCOVEL, THOMAS. 1988. *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. New York: Newbury House.

SELINKER, LARRY. 1972. Interlanguage. *IRAL* 10, 209-231.

SHIN, SUNJEONG. 2008. *Read and Speak Korean for Beginners with Audio CD*. 1st edition. McGraw-Hill.

STOCKMAN, IDA J., JOHANNA BOULT, and GREGORY C. ROBINSON. 2008. Multicultural/multilingual instruction in educational programs: A survey of perceived faculty practices and outcomes. *American Journal of Speech-Language Pathology* 17, 241-264.

WOLFE, VIRGINIA, DAVID MARTIN, THOMAS BORTON, and HEATHER CONNER YOUNGBLOOD. 2003. The effect of clinical experience on cue trading for the /r-w/ contrast. *American Journal of Speech-Language Pathology* 12, 221-228.

Jong-mi Kim
Department of English Language and Literature
Kangwon National University
Kangwondaehak-gil, Chuncheon-si,
Gangwon-do, 200-701 Republic of Korea
e-mail: kimjm@kangwon.ac.kr