ISCA Archive http://www.isca-speech.org/archive



INTERSPEECH 2004 - ICSLP 8th International Conference on Spoken Language Processing ICC Jeju, Jeju Island, Korea October 4-8, 2004

What Makes a Non-native Accent?: a Study of Korean English

Jong-mi Kim* and Suzanne Flynn**

*Department of English Language and Literature Kangwon National University, Republic of Korea

kimjm@kangwon.ac.kr

**Department of Linguistics and Philosophy Massachusetts Institute of Technology

sflynn@mit.edu

Abstract

We report a set of results in the second language (L2) acquisition of English phonology by first language (L1) speakers of Korean. Specifically, we focus on significant differences isolated between L2 speakers' production of isolated words in English and their production of the same words in sentence and phrasal contexts. Results indicate significantly more accurate production of words in isolation than in the production of the same words in phrasal contexts. The particular phonological phenomena focused on concern both stress reduction and placement. We also consider several other aspects of segmental phonology. We argue that the discrepancy in results observed between tasks may account for many of the seemingly disparate results indicated in other studies of L2 phonology. We discuss several possible explanations for these data in terms of which production task most closely provides a measurement of developing linguistic competence and which might reflect the role of either general learning strategies (overgeneralization) or reversion back to the L1 grammar under conditions of stress or when the L2 grammar is not fully developed.

1. Introduction

What makes a non-native accent? Several factors may contribute to this. An accent may involve segmental insertion, deletion, and/or substitution. On the other hand, an accent may be due to differences between the L1 and the L2 prosodic patterns, which may in turn reflect differences having to do with duration, amplitude and/or pitch. Among these different possibilities, most research has focused on issues concerning segmental factors. Few studies have investigated the L2 acquisition of stress and prosody, however see Broselow; Archibald and Carson [1, 2] and fewer involve any acoustic analyses. As a consequence, we are left with a very incomplete understanding of what accounts for a non-native accent.

In this paper, we contribute to the study of accent through quantitative acoustic analyses of a large database consisting of recorded speech samples from L1 speakers of Korean learning English as an L2. Our analyses focus on prosodic aspects of the L2 speech samples in a systematic and controlled way. We believe that the conclusions suggested by these data contribute to our understanding of the development of L2 phonology.

At other levels, the results of this current study can uniquely inform speech engineering applications as well as other such endeavors. The data reported here represent an expansion of a research project that is currently being implemented in a non-native speech recognition system [3].

2. The L2 study

The data was collected from 50 native speakers of Korean learning English as an L2 as well as six native speakers of English. The subjects were recorded while reading aloud phonologically controlled English texts. All the L2 learners spoke the standard dialect of Korea. The L2 level of competence in English varied, although most of the subjects were at the intermediate level as determined by their academic standing in a pronunciation class. The following describes the methodology chosen for the major part of this study that involves 48 subjects. A small subset of the data was collected by various other methods for a balanced study.

The methodology of this study consisted of four parts. First, a native speaker of English was chosen as the model speaker. This native speaker was 45 years old, male and a native of the state of Utah, U.S.A. His dialect was standard American mid-western.

The second part of this task involved the recording of all the stimuli by the native speaker (ns). To do this, the ns read the experimental stimulus items in a quite room. The readings were recorded and digitized at a 16 kHz sampling rate using *Computerized Speech Lab* by Kay Elemetrics, Inc.

The third part of the task protocol involved distributing the model speech files to all the participants in the study in both CD form and as downloadable files from an internet website. The participants were told to listen to the materials and to practice reading them for an average of a week. The learners were asked to imitate the model speech as closely as possible. In addition, prior to recording, the learners were given explicit phonetic instruction that focused on potential pronunciation difficulties with the stimuli.

In part four, the 48 native speakers of Korean were recorded. The learners read each stimulus item in a quiet room. The recordings were again digitized in 16 kHz sampling rate by *Computerized Speech Lab* by Kay Elemetrics. Examples of the stimulus items are in Table 1.

Table 1: Examples of Stimuli. (Note: Cons. = Consonants, Into. = Intonation)

Property	Example
Vowel	b[ea]t, b[i]t, b[ai]t, b[e]t, b[a]t, b[o]tt, b[ou]ght
Cons.	[p]eep, [t]eat, [k]ey, [b]eep, [d]eed, [g]ee
Syllable	[r]ice, [l]ice; ba[th], ba[the], [sp]ain, [spr]ain
Stress	[a]dd, [a]ddition; WH[I]TE h[ou]se, wh[i]te H[OU]SE
Rhythm	[Deliver] [books] [Friday]. [Deliver] [books by]
	[Friday]. [Deliver the] [books by] [Friday].
Into.	Teaching languages is harder than learning them.

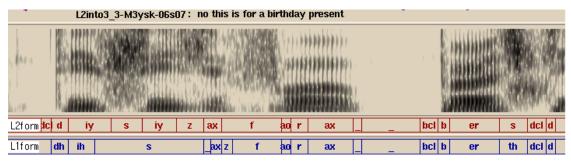


Figure 1: Illustration of L2 phones by a Korean learner of spoken English where L2 form shows the inserted vowel [ax], substituted phonemes [d], [iy], and [s] in the utterance "No, this is for a birthday present". (Note: L2form = Phonetic surface forms of Korean English, L1form = Corresponding forms of American English)

The data consisted of 9681 speech points recorded by the L2 Korean participants in the study. The speech data collected were processed in the speech recognition engine of MIT SUMMIT [4]. This engine was specifically modified for Korean English [3]. The output of the engine was manually corrected depending on the acoustic analysis of a given speech signal and the perceptual judgments of a native speaker of American English. We present the linguistic aspects of the results in the following text; we begin with a discussion of the segments and then proceed to the prosody results.

3. L2 accents on segmental level

Some of the segmental phenomena isolated in this study in the L2 forms can be understood in terms of critical differences between the L1 and the L2.

3.1. Distribution of L2 phones

Figure 1 illustrates several distributional aspects of Korean L2 phones that are perceived as different phonemes by a native speaker of American English in the learner's utterance, "No, this is for a birthday present". The figure represents a forced aligned output of SUMMIT recognizer, which is acoustically accurate.

The dental fricatives $[\eth]$ and $[\theta]$ in the words 'this' and 'birthday' are not present in the phoneme inventory for Korean. This difference between English and Korean has apparently triggered the substitution of an asymmetrical phonological form, the stop [d] and the fricative [s].

However, the L2 pronunciation of [z] as in the word 'is' is also absent in the Korean phoneme inventory; yet [z] is correctly produced by the non-native speakers in English. The phone insertion of [w] at the end of the word 'is' -- [w] in IPA is marked by [ax] in Figure 1 -- is derived from the different syllabic constraints that hold for Korean and English. [w] is a default epenthetic vowel in Korean. Korean does not allow fricatives in a syllable final position; this in turn triggers the insertion of a vowel to make [z] syllable initial thus creating a segment whose phonetic quality is faithful to the L1 phoneme [z] in English. If in Korean native phonology, any syllable final [z] is realized as a stop [t].

Another example demonstrates the incorrectly stressed vowel [i] in the word 'is', -- [i] in IPA is marked by [iy] in Figure 1-- where the target form is an unstressed vowel [ə]. This is due to the vowel inventory in Korean, where there is no unstressed vowel. Stress is not a lexical property of Korean. In our data, the unstressed vowel [ə] is realized in many different L2 phones: r-colored unstressed schwa [ə-], a stressed mid back

vowel [Λ], an orthographic dictation of the foreign sounds [a, u, e, i] as in 'Korea. calculate, given, dolphin', respectively.

We can understand these modifications as deriving from the phonological differences between English and Korean in terms of phonemic distribution, syllabic structure, and/or stress placement. In this case, learners seem to make accommodations for the distributional discrepancies between the L1 and L2 phones. What the Korean speakers produce reflects their attempts to articulate new target sounds in L2 English. As is well known, such L2 productions can result in the perception of a non-native accent by native speakers of a particular target language.

3.2. Phonological rules of L2

In addition to this type of accommodation, learners also resolve differences between the conflicting phonological rules in the L1 and L2. The utterance in Figure 2 is the word "splendid", articulated by another L2 learner. Critical to note in this case as well is that the articulation of the phonemes is perceived differently by the native English speakers than that intended by the L2 speaker in his articulation.

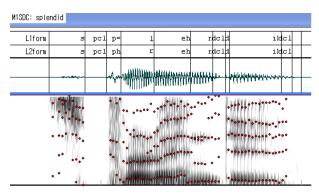


Figure 2: Illustration of L2 phones by a Korean learner of spoken English where L2 form shows the allophonic alternation of native phonology [r], but not that of target phonology [p=], in the utterance "splendid". (Note: L2form = Phonetic surface forms of Korean English, L1form = Corresponding forms of American English)

In Figure 2, the [r] in the L2 'sp[r]endid' derives from the liquid alternation rule in Korean phonology while the heavy aspiration of [ph] in 's[ph]lendid' disregards the English rule of de-aspiration in such an environment. The native speaker of American English in this study perceived and transcribed

this sequence with a vowel inserted between [ph] and [r], as in [sphərendəd]. The L2 articulation here suggests that at least at the level of production, the rules of English phonology have not yet been completely internalized by this speaker.

3.3. Non-nativeness of L2 phones

Incidentally, many of these L2 rules are also found in L1 native speech. We tested how general L2 rules are by implementing in the speech recognition system all the L2 segmental rules including the types of the rules discussed in this section. A total of 72 rules of L2 speech were implemented in SUMMIT [4]. Then we forced the machine align the phonetic transcription for both the native speech and the learner speech. Table 2 summarizes the L2 speech forms with the corresponding L1 phonetic strings on a randomly chosen data of 1182 words (542 Non-native speech, and 540 Native speech). Each of the machine aligned segmental string was manually checked for the acoustic and perceptual accuracy, by one researcher and an American listener.

Table 2: L2 segmental phenomena, as appeared both in native speech and learner speech. (Note: L2 words=Words that contain L2 segmental phenomena)

	L2 words	Tested words	Ratio (%)
Learner speech	183	540	33.9 (%)
Native speech	63	542	11.7(%)

Table 2 suggests that what has traditionally known as L2 speech is not confined to the non-native accent (33.9%), but can also be found in native speech as well (11.7%). For example, some productions of sounds [æ] can be acoustically and perceptually marginal to the phone categories of [ε] or [æ] for both native speech and L2 speech of American English. Besides, many of non-native speech share identical or similar phonetic surface forms to native speech (the remaining 66.1% of learner speech), although literature in non-native accents focus only on the different forms (33.9%).

4. L2 accents on prosodic level

Much of what is perceived as a non-native accent reflects subtle deviations from native prosodic patterns. In the following sections, we outline the L2 development of stress, rhythm, and intonation. Stress assignment is explored in the greatest detail.

4.1. Stress reduction by affixation

In our database, the duration of vowels was measured for contrastive words containing stressed, and unstressed vowels, as in '[a]dd' and '[a]ddition'. The Korean learners were hypothesized to have difficulty in stress reduction because Korean is a syllable timed language that places an even duration on each syllable regardless of whether or not it is stressed. We [5] also measured the duration of the vowels of the same words when they were embedded in sentence contexts. A pair of example sentences: 'Add the remaining ingredients and bring to a boil; Addition and subtraction are learned skills'. The measurement of the durations, although not exemplified in Korean phonology suggest that learners better command the stress reduction rule for English at the word level rather than at the sentence level as indicated in Figure 3. The chart in Figure 3 represents the percentage of stress reduction made by the Korean learners in isolated in comparison to the ns of American English.

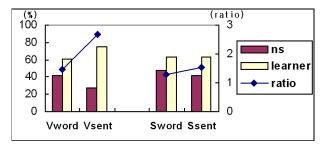


Figure 3: Percentage of stress reduction in learner speech as compared to that in native speech. Learner ratio shows greater discrepancy from that of native speech in sentence level. (Note: Vword=vowel duration in words, Vsent=vowel duration in sentences, Sword=syllable duration in words, Ssent=syllable duration in sentences)

As shown in Figure 3, Korean learners have reduced the unstressed vowel in the stimulus words in a manner that closely resembles the native speaker at least in terms of the measurements of this experiment. These results, quite paradoxically contrast with those for vowel reduction in the same words by the same speakers in a sentence domain. Learner speech shows greater distance from the native speech, i.e., less command of the rule at the sentence level.

A number of explanations are possible. One may be simply that the additional demands involved in the articulation of the sentence seriously compromise the L2 learner's ability to "attend" to all other phonological aspects such as vowel reduction at the word level. Articulation of the vowels in non-reduced forms does not interfere with comprehensibility. It may result in slightly "accented" speech depending upon the context for the utterance.

A second possibility may be that the word-level results indicate simply lexical learning for these words as a part of some list. Under such a scenario, when the learner is forced to articulate the word in a sentence -- which demands much more in terms of attention to other linguistic details -- the learner "forgets" what s/he has lexically learned.

The first scenario suggests that the learner has internalized the phonology of English and that reading a longer utterance demands much more in terms of performance. Thus, what is observed is that the learners have developed a grammar (competence) for the L2 and the evidence elicited suggests performance but not competence problems. The second scenario suggests that the learners have not yet developed a competence for the L2 phonology. While we do not have the data at hand to tease apart these two scenarios, we find the results both intriguing and important in terms of subsequent investigations along these lines.

4.2. Stress assignment by compounding

The differences in percentages of target performance between isolated words and sentences are also observed when we focus on stress placement in compound words and phrases. Examples are the contrast between noun compounds and noun phrases such as 'the White house' and 'a white house' in isolated phrases, sentences, and paragraphs.

A priori, the Korean learners were expected to have difficulty in their command of the noun phrases which involve a rise in pitch. This is because the raised pitch must be derived from the stress on the following element; this does not exist in Korean. In contrast, the pitch drop in compound nouns was not

expected to be difficult for the learners to acquire because pitch declination in an utterance is hypothesized to be universal. Figure 4 summarizes the average difference of pitch values in Hz for the two different types of noun phrases in isolation and in sentential and contexts.

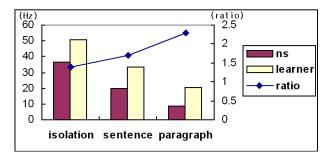


Figure 4: Difference of F0 values (Hz) in noun compounds from that of noun phrases in learner speech as compared to that in native speech. Learner F0 values show progressively greater discrepancy from that of native speech in isolation, sentence and paragraph level.

In Figure 4, the results indicate that the learners have imitated the ns model significantly better in word level than in sentence or paragraph level.

It is important to note that our result conform the previous study[6] in which the data collection method differed from our data reported in this paper. In Jeong's study[6], the learners did not have a ns model for a practice period prior to testing. The speakers were asked to read controlled texts at the testing session. The learners were assumed to have learned the stress placement rule in their English classes in high school. Four learners and three American speakers were asked to read, seven times, the noun compounds and noun phrases such as 'the White house' and 'a white house' in isolated phrases, sentences, and paragraphs.

The result in Figure 4 might be construed as a form of overgeneralization or it might indicate difficulties with other contextual factors present in a sentence context and not relevant at the isolated word level.

4.3. Non-nativeness of L2 prosody

The prosodic discrepancy between learner speech and the native speech is indeed perceived as "non-native" by a native speaker. We report below some results of perceptual experiments that are based on temporal sequencing of prosody[3].

English rhythm involves the placement of a timing beat on stressed syllables, i.e., foot, while Korean rhythm does not. These typological differences in rhythm between Korean and English led us to hypothesize that the Korean learners will have problems with temporal sequencing in English. Table 3 summarizes the perceived non-nativeness related by the duration of the sentences and the embedded non-final function words in 127 test sentences that comprise 10 native speech and 117 learner speech data.

Table 3: Perceived duration range of non-final function words and sentences

	Data	Percentage	Perception
Funtion words	65/101	64.4 (%)	Staccato Rhythm
Sentences	36/117	30.8 (%)	Not fluent

In Table 3 the native speaker perceived "staccato rhythm" or "not-fluent" when the duration of the learner data was deviated 20% from that of the native speech. The perceptual degree is gradual, and yet there was some consistent judgment beyond the threshold of 20% deviation. The mean duration ratio was 1.23 for function words, and 1.12 for sentences. There were 16 sentences excluded for the function word measurement because they did not have non-final function words.

The overall duration of the learners' speech patterns was found to be longer than the native speakers. In addition the articulations also indicated more pitch peaks and more breaks than the native English speaker. These peaks and breaks are due in large part to hesitations, repairs, and pronouncing unstressed syllables pronounced with longer durations than that of English speaker. As expected, the utterances of the L2 learners tended to exemplify longer durations than the native speaker.

5. Conclusion

The data we have discussed here in this paper are meant to be representative of several phenomena involved in the development of the L2 phonology in English by native speakers of Korean. The data reported here are from Intermediate learners of English. We focused on the acquisition of certain aspects of prosody. Results suggest as indicated throughout the L2 literature that the language or languages already known "play" a role in the acquisition of subsequent languages especially with respect to the phonology. What is especially interesting in the data discussed above is that contrary to traditional accounts of L2 phonological learning, not all contrasts between the L1 and the L2 are problematic. And, in different contexts, phrasal versus isolated words, results in different types of articulations by the same speaker. This variation is much wider than that observed in the steady state of a native speaker but clearly indicated during prior developmental stages. Understanding the factors that constitute a non-native accent is vast and infinitely more complex than traditionally envisaged. Future research is needed to determine which of several explanations discussed in this paper can be empirically supported.

6. References

- [1] Broselow, E., "Prosodic phonology and the acquisition of a second language", *Linguistic Theory in Second Language Acquisition*, eds. by S. Flynn and W. O'Neil, Academic Publishers, Boston, pp. 295, 1988.
- [2] Archibald, J. and Carson, J., "Acquisition of Quebec-French Stress", Ms., University of Calgary, 2003.
- [3] Kim, J-M., Wang, C., Peabody, M., and Seneff, S., "An interactive English pronunciation dictionary for Korean EFL learners", Submitted at ICSLP, 2004.
- [4] Glass, J., "A probablistic framework for segment-based speech recognition", Computer Speech and Language, 17, 2003.
- [5] Park, K., Lee, O.-H., and Kim, J.-M., "English Rhythm in Foreign Language Education for Korean Learners", [Abstract only], Proc., PAAL, Okayama, Japan, 2003.
- [6] Jeong, K., Accentuation of English Noun Compounds and Phrases by Korean Learners. M.A. thesis, Kangwon National Univ., Chuncheon, Korea, 2003.