Non-native speech rhythm:
A large-scale study of English pronunciation
by Korean learners*

Jong-mi Kim** · Suzanne Flynn · Mira Oh
(Kangwon National University · Massachusetts Institute of Technology
· Chonnam National University)

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. To what degree do the rhythmical properties of the native language emerge in the acquisition of a second language? Specifically, would non-native English—a stress-timed language—evidence effects of Korean—a so-called syllable-timed language? In this study, we analyze the speech data from 111 adult native speakers of Korean at an intermediate level learning English as a second language, in comparison with those from 29 adult native speakers of English. The study involves a pre- and post-test analysis of speech elicited in a reading task. We focus on the acoustic quantification of the vowel duration in relation to F0 trace and formant location in all elicited speech that involve 1) stress alternation (John was 'sick of Fred and Sandy), 2) stress assignment (a 'blackboard vs. a black 'board), and 3) stress reduction (add vs. addition). The results indicate that the speech rhythm contrast is manifested in both stress alternation and stress reduction in terms of the pair-wise variability of the vowel duration between the two groups of native and learner speech of English. These results are further supported by independent evidence from the developmental aspect in English pronunciation and perception. (Kangwon National University, Massachusetts Institute of Technology, and Chonnam National University)

Keywords: speech rhythm, non-native speech, syllable-timed rhythm, stress-timed rhythm, acoustic measurements of speech rhythm, vowel reduction

1. Introduction

A major distinction has often been between two types of speech rhythm; one has traditionally been labeled stress-timed and the other syllable-timed (Pike 1945; Abercrombie 1967; among others). Abercrombie (1967) argues that linguistic rhythm is based on the isochrony of either syllables, or stress intervals, for all human language. Stresses occur at regular intervals in stress-timed languages such as English, Russian, and Arabic; and syllables do so in syllable-timed languages such as French and Yoruba.

* This work was supported by the Korea Research Foundation Grant (KRF-2004-042-A00070). The earlier versions of this work have been presented at the MIT speech communication seminar on August 24, 2005, and the Boston University Conference on Language Development on November 3-5, 2006. The work has greatly been benefited by the valuable comments and help from Sun-Ah Jun, Kenneth Stevens, Michael Kenstowicz, Stephanie Seneff, Adam Albright, and Young-Seok Kim; although the remaining errors are our own.

** The first author.
The isochronic view of the rhythmic distinction cannot be measured in the phonetic signal (e.g., Bolinger 1965; Roach 1982; Wenk and Wiolland 1982). Thus, some weak forms of the rhythmic distinction account depart from a phonetic explanation of isochrony, but find the explanation in phonological characteristics of the syllable complexity and the vowel reduction (Dasher and Bolinger 1982; Dauer 1983). That is, complex syllables and reduced vowels are expected to be found only in the stress-timed languages, but not in the syllable-timed ones. Dasher and Bolinger (1982) argue that rhythm type as such results from the phonological structure of a given language.

The phonological account of vowel reduction and complex syllables as in Dasher and Bolinger (1982) and Dauer (1983) is supported in acoustic studies of these phenomena. Ramus et al. (1999) measured the duration correlation of vocalic and consonantal intervals. According to their study, syllable-timed languages in comparison with stress-timed ones demonstrate 1) a higher proportion of vocalic intervals, because a syllable-timed language prohibits vowel reduction; and 2) less variation in the duration of consonants, because a syllable-timed language does not permit complex syllables. Thus, there is apparent contradiction in the phonetic studies — whether to support the rhythmic distinction (Ramus et al. 1999) or not (Bolinger 1965; Roach 1982; Wenk and Wiolland 1982). Is there distinction between stress- and syllable-timed languages?

1.1 Previous phonetic studies of rhythm

To test the validity of rhythm distinction, previous phonetic studies used various acoustic cues. Ramus et al. (1999) adopt overall measurement of duration to find the acoustic correlation with the phonological attributes to speech rhythm. The overall measurement is then disputed by Grabe and Low (2002), who demonstrate a better correlation of speech rhythm with the Pairwise measurement of duration variability (Pairwise Variability, henceforth). Pairwise Variability refers to the durational variability between each pair of the comparing measurements, i.e., the durations of vowels, syllables, syllable sequence, or the intervals between vowels (excluding pauses), by taking the value of the difference and dividing it by the mean duration of the pair. The results show that stress- and syllable-timed languages differ in Pairwise Variability manifested in vowels. Low et al. (2000) finds the correlation of Pairwise Variability to the reduced vowel quality in terms of spectral patterns. The correlation is expected because the Pairwise Variability of duration reflects the binary nature of a rhythmic foot in stress-timed languages, where a stressed syllable is paired with an unstressed syllable. The unstressed syllable in a pair is expected to have the reduced vowel quality. Grabe and Low (2002) demonstrate a better correlation of rhythm with vocalic Pairwise Variability than with intervocalic
Pairwise Variability, an enhanced measurement comparable to the duration of complex syllables. Acoustic correlates as such are associated with phonological structure of language. The durational Pairwise Variability is associated with the binary phonological structure of stressed and unstressed syllables in a foot. Reduced vowel quality in spectral patterns is associated to unstressed syllables. Duration reduction of unstressed syllable sequence is associated to complex syllable. Among various acoustic correlates with phonological prominence, the most relevant phonological phenomena in English speech rhythm may be the stress alternation that pairs stressed and unstressed syllables, in addition to the rules of stress assignment and stress reduction.

We also find from the literature in phonology the phonetic attributes of phonological rules associated with rhythm. For stress alternation, the rhythm rule that pairs stressed and unstressed syllables in a speech string has been represented in terms of metrical grid, i.e., a weight unit (Hayes 1995), a phonological representation of prominence. The stress assignment rule is reflected in terms of pitch accents, H and L, i.e., tonal heights (Ladd 1996). The stress reduction rule has been represented in terms of vowel quality change (Burzio 1994). Our study attests these phonological associations to the acoustic cues as investigated by previous studies, but with a greater scale of new data.

A greater scale of data is needed in a new study, because previous studies have attempted to find the validity of speech rhythm based on limited amounts of data. For example, the whole language is represented only with 2 minutes of spontaneous speech by 1 speaker in Roach (1982), with 5 sentences by each of 4 speakers in Ramus et al. (1999), with a passage of approximately 140 syllables by 1 speaker in Grabe and Low (2002). Scarcity of data has also been a problem in the study of non-native speech rhythm to best accommodate the time-consuming nature of the task. For example, all of the 8 speakers in Jian (2004a,b) and 20 speakers in Low et al. (2000) read 5 sentences, each comprised of approximately 6 vowels of investigation. Thus, in this study the data we report are the result of a larger scale investigation from multiple subjects (29 native speakers of American English and 111 native speakers of Korean learning English as a second language).

1.2 Previous acquisition studies of second language rhythm

Several acquisition studies of adults that focus on acoustic measurements of speech rhythm have indicated the influence of the native language on the target language at early stages of acquisition. This effect emerges in terms of acoustic measurements of the factors that derive from differences between a syllable-timed native language and a stress-timed target language (Mochizuki-Sudo and Kiritani (1991) for Japanese English, Low

In this respect, speech rhythm of English as produced by Korean learners may offer an interesting result. English and Korean can mark a prototypical case of rhythmic distinction in terms of syllable complexity and vowel reduction according to the criteria in Dasher and Bolinger (1982) and Dauer (1983). That is, English has complex syllables and reduced vowels, while Korean has neither. Korean syllables allow only a single consonant regardless of onset and coda positions. The Standard Korean dialect (Korean, henceforth) has other stress-mute features of non-lexical tone, stress, or length (Jun 1995, 2000, 2005; Lim 2001; Yun 2002).

Even in Korean linguistics, however, there have been a number of controversies over the type of speech rhythm of the Standard Korean dialect. Lee (1973) and Lee (1990) claim that Korean is stress-timed. They propose that stress in Korean is determined by the syllable duration, although its fundamental frequency ($F_0$) plays a marginal role in stress assignment. In contrast, Jun (1995) and de Jong (1994) convincingly argue that Korean is syllable-timed, since the prosodically strongest syllable in the word changes depending on its position within Accentual Phrase (AP) — stress is not a property of a word but rather that of AP. In their studies, AP has its own pitch pattern, LHLH, but every word does not exhibit duration or pitch contour. In a nutshell, duration plays a prominent role in the stress-timed approach to prominence in Korean, while fundamental frequency does so in the syllable-timed approach. In this respect, the duration study by Yun (2002) negates Korean as a stress-timed language. Lee and Kim (2005) along with other studies of duration (Kim and Lee 2002; Lee and Jang 2004) further demonstrate that syllable-timedness is manifested in Korean learner speech of stress-timed English.

Given the syllable-timing influence on stress-timed English speech of Korean learners, we then question, "What phonetic correlates should we investigate to test the rhythm contrast?" Although duration has been most widely studied in relation to speech rhythm, literature survey entails the relevance to vowel formants (Low et al. 2000) and fundamental frequency (de Jong 1994; Jun 1995; Lee 1973; Lee 1990). The three acoustic parameters of duration, formants, and fundamental frequency have long been proposed to render perceptual prominence either in Korean, or in English, or both. There are debates, however, on which acoustic cue mainly functions to indicate prosodic prominence in English. A change in fundamental frequency is reported as the most reliable cue for prosodic prominence in English (Bolinger 1958). On the other hand, Kochanski et al. (2005) view that English speakers mark prominence with patterns of duration and loudness, but not with fundamental frequency. In contrast

---

1 The Standard Korean dialect, which our research participants speak, is spoken in Seoul and in the mid west Korea. It is the most receptive and prestigious dialect used in educational and broadcasting media.
with the disagreement in acoustic cues for prominence of English, Korean has been agreeably reported to have a change in fundamental frequency as the most reliable cue for prosodic prominence (Koo 1986; Jun 2000). Jun (2005) further claims that Koreans perceive rhythm by Accentual Phrases which is defined by $F_0$. Do these studies on acoustic correlates of prominence suggest that Korean and English speakers may employ different acoustic cues in perceiving and producing linguistic rhythm? We are then forced to test all these acoustic cues relevant to the rhythm in both languages: duration, vowel formants, and fundamental frequency.

To test the validity of speech rhythm distinction in native and non-native speech, we have already completed three background experiments that utilizes the acoustic cues for phonological prominence: 1) duration differences ($\text{msec}$) for stress alternation phenomenon, 2) fundamental frequency changes ($\text{Hz}$) for stress assignment environment, and 3) vowel formant frequencies ($\text{Hz}$) for stress reduction environment. These experiments have already been published and reviewed in detail in Section 2 as background for our new experiment in Section 3. The new experiment in Section 3 investigates duration differences for stress reduction environment. The conclusion is drawn in Section 4 from all the four experiments in Sections 2 and 3.

2. Background experiments

This section explores three background experiments of 1) stress alternation by Lee and Kim (2005), 2) stress assignment by J.-M. Kim (2005), and 3) stress reduction by Kim and Lee (2005) that outline English speech rhythm and its contrastive realization by Korean learners. For the sake of a coherent account of speech rhythm, the same research methodology is used for background experiments in Section 2 and the new experiment in Section 3, as outlined below.

For research participants, each experiment in Sections 2 & 3 involves one model speaker of American English who provides the sample speech, 23-28 other native speakers of American English who read the same data for learners, one instructor who is the first author, and 24-46 Korean learners of English. The native speakers spoke the Mainstream American English, while the learners spoke the so-called Standard Korean as a native language. All Korean participants were enrolled in an English pronunciation class, as the first author’s students at a Korean college at the time of experiment. The learners’ English proficiency level was mid-

---

2 Although $F_0$ is more important than duration, focus is also realized in duration, so that Koreans would still be sensitive to the postlexical duration difference.

3 The same 25 native speakers were in all of the four experiments in Sections 2 and 3 with some addition of more speakers in Section 2.1. The learner participants were the same for the experiments in Sections 2.1 and 2.3, but different elsewhere in order to avoid overloading. Although the native speakers were from various parts of the United States, many had changed their residences and none had noticeable regional accent.
intermediate in the scale of five ranks according to Test Of English for International Communication (TOEIC). For speech materials, each experiment in Sections 2 & 3 uses two types: production and perception. All participants are recorded for speech production, and only the Koreans are dictated for perception materials. The production materials are the recording list relevant to stress, such as stress alternation, stress assignment and vowel reduction. Both the native and learner participants read the same production materials for the comparison of native speech with learner speech. The perception materials are the two sets of native speech for learners to take the listening comprehension proficiency test before and after instruction. We keep the task difficulty level as same as possible for the two sets to recognize the improvement of listening comprehension. Thus, the pair of words and sentences is held constant for low-level vocabulary, the similar number of syllables in the sentence, and simple sentence structures. Both lists are read for recording by the same model speaker in a single recording session, but displayed for testing at different times: before and after the instruction. The listening test was a co-variance reference to determine if the rhythm development in production has a correlation to the development of listening comprehension.

The data acquisition procedure from the Korean learners consists of three stages: 1) pre-listening and production tests after a brief pronunciation instruction about the recording stimuli, 2) main pronunciation instruction in class, and 3) post-listening and production tests. We test the learners before and after instruction to investigate their developmental patterns of speech production with relation to listening comprehension. On the other hand, the native speakers underwent only one stage of recording for production stimuli. Each recording stage takes less than a week.

At the first stage of the data acquisition procedure, the learners take listening and production tests on the speech materials on the very first day of instruction. For the listening comprehension test, the learners write down the recorded listening materials of native speech for pre-instruction. The average time spent is 5 minutes, in order not to take away the class-time during which the learners must concentrate on production instruction and test. For the production test, the learners are first given a 20-minute pronunciation instruction, during which the production list is presented in a handout. The teacher (the first author) explains to the learners about the meanings of the words in the list. The learners listen to and repeat after the model native speech for one time. Then, they are asked to go to one of the recording rooms immediately after the class, and allowed to do more practices on their own if time permits. They are then asked to record their

---

4 TOEIC is a standardized test of English proficiency provided by the Educational Testing Service. According to TOEIC Can-Do Guide, different ranks are assigned to the listening scores of 1) 10-100, 2) 105-225, 3) 230-350, 4) 355-425, and 5) 430-495. The Korean learners in this study belong to the third rank.
speech reading the production list. The word list is read in a fixed sentence frame, “Say X again,” to elicit the data in the invariable phonetic environment. The recording was digitally done in 16 KHz, 16 bit by trained teaching assistants. The devices used are Computerized Speech Lab by Kay Elemetrics and other PC equipments. The distance to the microphone from the mouth is adjusted to the amplitude of the voice displayed by the waveform signal of test speech. Utterances with errors or hesitations are re-recorded at the end of each session, when requested by the learner. All recordings by learners have been completed within a week.

At the next stage of the data acquisition procedure, the learners are taught the class materials including the expected reduction of unstressed vowels and syllables in English. The class materials consist of art works from poems, chants, and pop-songs; conversational speech from skits and movie extracts, and reading speech from TOEIC and TOEFL. During the instruction, the teacher (the first author) explains rhythm differences in English and Korean, and the learners listen to and repeat after the model speech in the stimulus list, once more. The teacher gives a one-time feedback on the learner speech for each speaker. Further, the model speech is distributed in CD, and by web-downloads for optional practice at home. The classroom instruction on pronunciation lasts for 3-5 weeks.

At the last stage of the data acquisition procedure, the participants are tested again on the listening comprehension by writing down the recorded listening materials of native speech for post-instruction. Then, they are again asked to read and record the materials in one of the recording rooms, but only once in the post-test. All recordings are completed in a day.

In contrast with the data acquisition procedure for learner speech, the native speakers in this study do not have any lesson or training session in English. They are immediately asked to read the recording list twice that is mixed with other approximately 60 utterances. All of the native speakers have acknowledged that they understood the meanings of all sentences in the recording list. The second-time recording is taken for the data, although the first-time recording is occasionally taken to replace a weak, noisy or unnatural speech signal.

For data analysis of each experiment in Sections 2 & 3, we first measure the acoustic values that are considered relevant to prominence: 1) duration of vowels, syllables and feet, 2) $F0$ height, and 3) formant frequencies of $F1$ and $F2$. Then, we normalize the raw measurement values in terms of speech rate and pitch range. After that, we identify the correlation between phonetic and phonological elements with speech rhythm. Thus, our initial goal of investigation is to identify what the core phonetic and phonological correlates of speech rhythm in Korean English are.

---

5 TOEFL is the Test of English as a Foreign Language, respectively, administered by the Educational Testing Service, NJ, USA.
2.1 Experiment on stress alternation and speech timing

To quantify English stress alternation in native and learner speech, Lee and Kim (2005) measure the Pairwise Variability (%) of duration (\text{msec}) between the adjacent stressed and unstressed syllables within a foot as well as that among adjacent feet in 555 English sentences, which were read by 29 natives and 41 learners before and after instruction. The speech materials consist of five sentences containing stressed and unstressed syllables in a sequence as in 'John was sick of Fred and Sandy'. The recorded corpus contains 1,554 stressed and unstressed pairs of syllables (= 14 pairs \times (41 learners \times 2 recordings + 29 native speakers \times 1 recording), excluding the sentence final feet). For analysis, the authors first measure the duration value (\text{msec}) of the stressed syllable and the unstressed syllable sequence within a foot, and then normalize the raw duration values in terms of the mean value of the pair. The authors then compare the Pairwise Variability and Foot Variability of duration. The results show that in comparison with native speech, learner speech before instruction had a significantly lower Pairwise Variability between adjacent stressed and unstressed syllables within a foot. Figure 1 demonstrates the differences of Pairwise Variability in native and learner speech.

![Figure 1. Pairwise Variability for the duration between stressed and unstressed syllables increases in the order of learner speech before instruction, after instruction, and native speech (Pairwise Variability (%) = 46.0 < 50.7 < 81.3; n = 1,554; p < .001) (This figure is reproduced from Lee and Kim 2005:105).](image-url)
In Figure 1, the Pairwise Variabilities (%) in native speech (dotted bars) are significantly large in positive values, which indicate that the duration (msec) of the stressed syllable is significantly longer than that of unstressed syllable sequence within a foot (eg., John: was). The average Pairwise Variability is 81.3 % in native speech. On the other hand, the Pairwise Variabilities (%) in learner speech before instruction (crisscrossed bars) are significantly smaller than those in native speech, which indicate that the duration (msec) difference of the stressed syllable and unstressed syllable sequence in learner speech is significantly shorter than that in native speech. The average Pairwise Variability is 46.0 % in learner speech before instruction. The Pairwise Variability (%) on average increases in the order of learner speech before instruction, after instruction, and native speech (46.0 < 50.7 < 81.3; p = .001). From these results of stress alternation in tandem with an experiment on Foot Variability,6 Lee and Kim (2005) conclude that stress alternation is acoustically quantifiable by means of the Pairwise Variability for duration, as in (1).

(1) Acoustic quantification of stress alternation by Pairwise Variability of duration.
   a. Native speech shows a greater range of duration alternation between stressed and unstressed parts within a foot.
   b. Learner speech shows a smaller range of duration alternation between stressed and unstressed parts within a foot.
   c. Korean learners show a progress in acquiring the stress alternation in English after three weeks of classroom instruction on pronunciation.

For native speech in (1a), the pairs of stressed and unstressed syllables demonstrate the largest Pairwise Variability of duration in Figure 1. This follows the expectation that English, being a stress-timed language, has a greater variety of syllable lengths depending on stress alternation, and a less variety of feet, indicating less sensitivity to the number of embedded unstressed syllables. For learner speech in (1b) as opposed to native speech, the syllables in Figure 1 demonstrate the lowest Pairwise Variability. This follows the expectation that Korean, being a syllable-timed language, has a lower variety of syllable durations, and a greater variety of foot durations depending on the number of embedded syllables. For learner development in (1c), the Pairwise Variability in learner speech moves toward the pattern of native speech as the proficiency increases. Further, Lee and Kim (2005) demonstrate that the better command of stress alternation after instruction accompanies a better performance of listening comprehension task. This follows the expectation that the acquisition in production is part of overall proficiency development that involves perception as well. From these

6 Foot Variability is the duration variation among adjacent foot that is normalized in terms of the average foot duration within the utterance. See Lee and Kim(2005: 102) for the algorithm.
findings, Lee and Kim (2005) conclude that the stress alternation phenomenon in English is acoustically cued by Pairwise Variability of duration. To further investigate speech timing, we conjecture that stress alternation may interplay with the stress assignment and stress reduction as the factors. In relation to this, Low et al. (2000) and Jian (2004a,b) discuss the interplay with reduced vowel quality, but not stress assignment. The following sections discuss the interplay of stress related factors in speech timing in terms of stress assignment and stress reduction. Can we quantify by acoustic means the phonological rules of stress assignment and stress reduction? If it is supported, what are the acoustic cues that distinguish native speech from learner speech in terms of the two speech rhythms?

2.2 Experiment of stress assignment and F0 slope

To quantify English stress assignment in native and learner speech, J-M. Kim (2005) measures the changes in fundamental frequency (Hz) between adjacent stressed and unstressed syllables in the compound nouns in comparison with English noun phrases which were read by 25 native speakers and 46 learners. The speech materials consisted of 12 contrasting sets of compound nouns and noun phrases, 7 sentences, and 1 paragraph that include contrasting compound nouns and noun phrases. The examples were a 'blackboard' for a compound stress rule vs. a black 'board' for a nuclear stress rule. The recorded corpus contains 1,523 samples of compound nouns and noun phrases (= 3,046 constituent words): 777 samples from learners and 746 samples from native speakers. For analysis, the author first measures the F0 value (Hz) at the mid-point of the given vowels, and then normalizes the raw F0 values in terms of the average pitch of each speaker. The author then compares the degree of F0 drop and rise in the contrasting noun compounds and noun phrases. The results show that learner speech whose native language has no lexical stress manifest the F0 slope (Hz), or the difference of F0 values between constituent words, in the same pattern as native speech. Figure 2 demonstrates the similar pattern of F0 slope in native and learner speech.
Non-native speech rhythm

In Figure 2, F0 slope values (Hz) in native speech are stronger negative values for the noun compounds (vertical and horizontal lines on the right) and weak negative or positive values for the noun phrases (diagonal lines on the right), which indicate that the stress placement is on the left element for noun compounds and on the right element in noun phrases. The average F0 slope values (Hz) in noun compounds are \(-37.9\) in isolation, \(-24.1\) in sentences, and \(-26.1\) in a paragraph; while those in noun phrases are \(2.3\) in isolation, \(-1.8\) in sentences, and \(-7.1\) in a paragraph. Likewise, F0 slope values (Hz) in learner speech (before instruction) are stronger negative values for the noun compounds (vertical and horizontal lines on the left) and positive values for the noun phrases (diagonal lines on the left), which indicate again that the stress placement is on the left element for noun compounds and on the right element in noun phrases. The average F0 slope values (Hz) in noun compounds are \(-34.2\) in isolation, \(-27.7\) in sentence, and \(-12.2\) in paragraph; while those in noun phrases are \(22.7\) in isolation, \(14.4\) in sentence, and \(7.5\) in paragraph. The learner speech is elicited only once before the main pronunciation instruction in class; because the data elicitation after the instruction became unnecessary due to the satisfactory performance even without it. From these results of stress

\[^7\] When stress is on the right, then the slope values are weak negative or positive values, given that the F0 trace declines in human utterances.
assignment, J-M. Kim (2005) concludes that stress assignment is acoustically quantifiable by means of $F_0$ height (Hz), as in (2).

(2) Acoustic quantification of stress assignment by $F_0$ slope.
   a. Native speech shows a sharp contrast between $F_0$ drop for compound stress and $F_0$ rise for nuclear stress.
   b. Learner speech also shows a sharp contrast between $F_0$ drop for compound stress and $F_0$ rise for nuclear stress.
   c. Korean learners readily acquire the stress assignment rules in English by less than an hour of classroom instruction on pronunciation.

For native speech in (2a), compound stress in noun compounds shows stronger negative $F_0$ slope values (Hz) in Figure 2; while nuclear stress in noun phrases shows weak negative or positive $F_0$ slope values (Hz). This follows the expectation that compound stress in English has a descending $F_0$ trace (negative $F_0$ slope values) with the stress assignment on the left, while nuclear stress a flat or slightly ascending $F_0$ trace (weak negative or positive $F_0$ slope values) with the stress assignment on the right. For learner speech in (2b) as similar to native speech, these two contrasting stress assignment rules are acoustically cued by the contrastive $F_0$ slope values between the pairs of compound nouns and noun phrases in Figure 2. The striking similarity between native and learner speech suggests for learner development in (2c) that the Korean learners readily learn compound and nuclear stress assignment in English noun compounds and phrases, which is realized as pitch. From these findings, J-M. Kim (2005) concludes that the stress assignment rules of English are acoustically cued by $F_0$ height (Hz).

While stress assignment is easy, would the stress reduction in English be difficult for Korean learners to acquire? The question may boil down to two possible issues: 1) The acoustic cue $F_0$ for pitch lends little for prominence (Kochanski et al. 2005), and 2) the difficulty lies in stress reduction, not in stress assignment. The stress assignment in English may be easy for Korean learners by assigning a high pitch to a stress location, because the accentual pattern in Korean is realized by pitch height (Koo 1986; Jun 2000; among others). In contrast, stress reduction may be difficult for learners, because Korean phonology does not allow unstressed vowels. Among various features of stress reduction, vowel reduction in relation to rhythm has been extensively studied for Singapore English (Low et al. 2000) and Taiwanese English and Mandarin English (Jian 2004b). These studies suggest that stress-timed languages allow vowel reduction in contrast with syllable-timed languages. For this reason, a

---

8 The good pitch command by Koreans may be related to the fact that Korean intonation patterns are characterized in terms of pitch height, or fundamental frequency traces, in Accentual Phrase and Intonational Phrase (Koo 1986; Jun 1995, 2005; among others).
different experiment in the following section explores the difficulties of vowel reduction in unstressed syllables for learners.

2.3 Experiment on vowel reduction and acoustic formants

To quantify English stress reduction in terms of reduced vowel quality in native and learner speech, Kim and Lee (2005) measure the location of the first and second formants ($F_1$ and $F_2$) of stressed and reduced vowels in 10 morphologically-related word pairs of English for stressed and reduced vowel contrasts. The examples are add vs. addition; deprecate vs. deprecate; society vs. social, chosen from the greatest variety of vowel pairs from open sources. The words were read by 25 native speakers and by 41 learners before and after instruction. The recorded corpus contains 2,140 stressed and reduced vowels (= 10 pairs x 2 words x (41 learners x 2 recordings + 25 native speakers x 1 recording)). For analysis, the authors first measure the locations of $F_1$ and $F_2$ (in Hz) of the stressed and reduced vowel pairs, and normalize the raw formant frequency values with respect to gender, according to the method in Yang (1996). The authors then compare the average distance of formant location from the vowel centroid in stressed and reduced vowels. The results show that in comparison with native speech, learner speech before instruction has a significantly dispersed formant values for reduced vowels. Figure 3 demonstrates the differences of reduced vowel quality in native and learner speech.

![Figure 3. The dispersion difference between reduced vowels and their stressed counterparts increases in the order of learner speech before instruction, after instruction, and native speech (n=2,140 for stressed and reduced vowels; dispersion of reduced vowels from the centroid (Hz) = 354 > 310 > 219 ; n=1,320; p=.001) (This figure is reproduced from Kim and Lee 2005:84-7).](image)

9 The full list is the same as in Table 1 of Section 3, in an attempt to take as a consistent research methodology as possible throughout the experiments in Sections 2 and 3.
In Figure 3, the dispersion in native speech (two figures on the right) demonstrates a sharp contrast between stressed and reduced vowels, where the formant frequencies (Hz) are far more centralized in reduced vowels (bottom row on the right). The average distance of the formant location from the vowel centroid, or the dispersion value is 219 Hz in native speech. On the other hand, the dispersion in learner speech before instruction (two figures on the left) demonstrates little contrast between stressed and reduced vowels, where the formant frequencies are not much centralized in reduced vowels (bottom row on the left). The average distance of the formant location from the vowel centroid, or the dispersion value is 354 Hz in learner speech before instruction. The average dispersion value of reduced vowels decreases in the order of learner speech before instruction, after instruction, and native speech (354 > 310 > 219; p = .001). From these results of stress reduction in terms of reduced vowel quality, Kim and Lee (2005) conclude that stress reduction is acoustically quantifiable by means of formant frequencies (Hz), as in (3).

(3) Acoustic quantification of vowel reduction by first and second formant frequencies.
   a. Native speech shows a greater centralization of vowel formants when unstressed.
   b. Learner speech shows a greater dispersion of vowel formants when unstressed.
   c. Korean learners show a progress in acquiring the reduced vowel quality in English after three weeks of classroom instruction on pronunciation.

For native speech in (3a), the dispersion of stressed vowels is significantly greater than that of unstressed vowels in Figure 3. This follows the expectation that a stress-timed language places a reduced vowel in the unstressed position (Dasher and Bolinger 1982; Dauer 1983). For learner speech in (3b) as opposed to native speech, the learner speech data before instruction contrastingly demonstrate a large dispersion for reduced vowels as that for stressed vowels in Figure 3. This follows the expectation that a syllable-timed language has no vowel reduction in the unstressed position. For learner development in (3c), the dispersion value of learner speech moves toward the pattern of native speech as the proficiency increases. Further, J-M. Kim (2005) demonstrates that the better command of reduced vowel quality after instruction accompanies a better performance of listening comprehension task. This follows the expectation that the acquisition in production is part of overall proficiency development that involves perception as well. From these findings, J-M. Kim (2005) concludes that the stress reduction rules in English are acoustically cued by formant frequency values (Hz).
To further investigate stress reduction, we consider the study of native speech by Lindblom (1963) and Flemming (2005) that vowel quality reduction is the result of undershooting by a shorter duration of reduced vowels. In second language studies, both Low et al. (2000) and Jian (2004a) discuss the reduced vowel quality in connection with the duration alternation for speech rhythm. Can we quantify stress reduction in terms of the reduced duration of vowels to pair with the reduced vowel quality? If the quantification is achievable, then do the acoustic cues distinguish native speech from learner speech whose native language has different speech rhythm?

3. Experiment on stress reduction

This section quantifies stress reduction in terms of the duration of vowels. The duration study in this section pairs with the vowel quality study of stress reduction in Section 2.3. The combined results of stress reduction will later be comprehensively studied with respect to stress alternation in Section 2.1 and stress assignment in Section 2.2 for our conclusion on speech rhythm.

To quantify English stress reduction in terms of reduced duration of vowels and syllables in native and learner speech, we measure the Pairwise Variability (%) of stressed and reduced vowels and syllables in morphologically-related word pairs in word level as well as sentence level. To our knowledge, the quantification has not been done other than our own preliminary works to this study, not only for learner speech, but for native speech as well.\footnote{Preliminary versions of this section have been presented at the conferences of Pan-Pacific Association of Applied Linguistics (PAAL) Conference, Okayama, Japan on August 2003 (a poster only), International Conference on Spoken Language Processing (ICSLP/INTERSPEECH), Jeju, Korea on October 2004, and the 16th Conference on Speech Sciences, Daegu, Korea on May 2005; although all presentations included less than half of the present discussion. The learner data were also newly acquired.} Although there have been some acoustic studies of reduced vowels in comparison to adjacent stressed syllables on non-native English (Low et al. 2000; Yang 2002; Jian 2004b), they have not dealt with the well-known phonological rules of stress reduction in morphologically-related examples (eg., add vs. addition) that serve the basis of shared underlying representation (Chomsky and Halle 1968: 110-126). Our speech materials use the same morphemes with contrasting stress, while other studies (e.g., Low et al. 2000; Jian 2004a) use all unrelated words embedded in the sentence unit. It is due to the different focus of the study that our interest is vowel reduction in a general sense (Chomsky and Halle 1968:110-26; Ladefoged 2006:94), while they are interested in stress alternation in the successive syllables. Vowel reduction in a general sense refers to the reduction of unstressed syllables from a stressed syllable, as in the underlined vowels in telegraphy [təlɪɡrɑːfi] that are reduced from the underlined stressed vowels in telegraph [tɛlɪɡrɑːf]. In this sense, the
speech materials are expected to share the same baseform, or the same underlying representation in Generative Phonology. To share the same baseform, the segments in comparison must belong to a morphologically-derived pair of words.

In contrast, Kim and Lee (2005), as summarized in Section 2.3, deals with vowel reduction in terms of $F1$ and $F2$ formant frequencies in words with morphologically-shared underlying representation. To pair with this study of vowel quality, we launch a duration reduction measurement to demonstrate that the duration reduces when unstressed. To do so, we first demonstrate that the phonological stress reduction is acoustically cued by duration measurement.

(4) **HYPOTHESIS 1:** Stress reduction is acoustically quantifiable in terms of temporal duration.

Hypothesis 1 in (4) may hold only for native English, but not for learner English of Korean speakers, since Korean phonology does not distinguish the stressed or unstressed opposition of vowel phonemes. We thus hypothesize that learners will have difficulty reducing the English vowels in the unstressed syllables.

(5) **HYPOTHESIS 2:** Korean learners have difficulty in reducing the duration of vowels in unstressed syllables.

If stress reduction is difficult for the learners as hypothesized, will the learners do better by classroom instruction? Previous literature has different views on this aspect of adult speech development: hardly any learning (Scovel 1988) vs. significant improvement (Kim and Lee 2005; Lee and Kim 2005). Along the line of our background studies in Section 2, we hypothesize that the learners will improve in stress reduction by learning.

(6) **HYPOTHESIS 3:** Duration reduction of unstressed vowels and syllables by Korean learners assimilates to the native values by classroom instruction on pronunciation.

The remainder of this section tests the three hypotheses aforementioned. To investigate duration reduction in unstressed vowels we compared the vowel and syllable durations in the word and sentence levels in learner speech with those in native speech.

### 3.1 Participants

Research participants consist of 24 Korean learners of English, 1 native speaker of American English as the model speech, and 24 other native
speakers of American English. The backgrounds of these speakers are the same as those described in Section 2.\textsuperscript{11}

3.2 Speech materials

We discuss two types of speech materials: production and perception. The production materials contain 10 morphologically-related word pairs for the stressed and reduced vowel contrasts. Of these pairs, we make 10 contrasting sentences that include the stressed and unstressed counterparts. The list is chosen for a greater variety of vowels from the open sources (Ladefoged 2006 and TIMIT 1990) to facilitate potential cross-checking of the data by other researchers. The word list is given in Table 1, where the contrasting vowels are underlined. We use the same set of production materials, before and after instruction.

Table 1. Production materials for stress reduction in word pairs.

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Stressed</th>
<th>Reduced</th>
<th>Vowels</th>
<th>Stressed</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>depreciate</td>
<td>deprecate</td>
<td>u</td>
<td>compute</td>
<td>Computation</td>
</tr>
<tr>
<td>i</td>
<td>implicit</td>
<td>implication</td>
<td>ñ</td>
<td>confirm</td>
<td>configuration</td>
</tr>
<tr>
<td>e</td>
<td>explain</td>
<td>explanation</td>
<td>ñ</td>
<td>confront</td>
<td>configuration</td>
</tr>
<tr>
<td>e</td>
<td>allege</td>
<td>allegation</td>
<td>ñ</td>
<td>majority</td>
<td>major</td>
</tr>
<tr>
<td>æ</td>
<td>add</td>
<td>addition</td>
<td>ø</td>
<td>society</td>
<td>social</td>
</tr>
</tbody>
</table>

In Table 1, the ten stressed vowels as in explain are on the left column and the reduced counterparts as in explanation are on the right column. We chose the word pairs in Table 1 for the sake of the maximal variation of different vowels and familiarity for the participants. These vowels do not exhaust all English vowels; but represent the potential alternation between stressed and reduced vowels that are morphologically-related.\textsuperscript{12} These pairs of contrasting words are embedded in the sentences as in Table 2.

Table 2. Production materials for stress reduction in sentence pairs.

<table>
<thead>
<tr>
<th>Vowels</th>
<th>Stressed</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>æ</td>
<td>What explains this uni-directional paralysis?</td>
<td>But that explanation is partly true.</td>
</tr>
<tr>
<td>æ</td>
<td>Add remaining ingredients and bring to a boil</td>
<td>Addition and subtraction are learned skills.</td>
</tr>
<tr>
<td>ñ</td>
<td>Will you please confirm the policy?</td>
<td>There was no confirmation about the policy.</td>
</tr>
<tr>
<td>ñ</td>
<td>This is nevertheless the majority.</td>
<td>He is a man of major talent.</td>
</tr>
<tr>
<td>ø</td>
<td>Does society really exist as an entity?</td>
<td>Differences were related to social backgrounds.</td>
</tr>
</tbody>
</table>

\textsuperscript{11} Readers are referred to the beginning of Section 2 for further details on the learners’ proficiency in English and the data collection procedure as the detail apply to this Section 3 as well with no change.

\textsuperscript{12} Among the list from Ladefoged (2006), we omitted the contrasts of another mid back vowel [œ] in invoke, invocation, two unreduced [ɔɪ] in exploit, exploitation, and morphologically unrelated [uː] in deviant, outsider.
Sentences in Table 2 embed 10 contrastive words with five vowels listed in Table 2. To take an example, the first sentence pair includes the words *explains* and *explanation*. The former is the stressed form and the latter is the reduced counterpart.

The perception speech materials are shown in Table 3, where the task difficulty level is kept as same as possible for the two sets in terms of vocabulary, number of syllables in the sentence, and sentence structures.

Table 3. Listening comprehension materials before and after instruction.

<table>
<thead>
<tr>
<th></th>
<th>Before instruction</th>
<th>After instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>become</td>
<td>Hello</td>
</tr>
<tr>
<td>2</td>
<td>discover</td>
<td>Understand</td>
</tr>
<tr>
<td>3</td>
<td>in the car</td>
<td>in a class</td>
</tr>
<tr>
<td>4</td>
<td>at a gym</td>
<td>on the phone</td>
</tr>
<tr>
<td>5</td>
<td>I won.</td>
<td>She cried.</td>
</tr>
<tr>
<td>6</td>
<td>They see you.</td>
<td>We hear him.</td>
</tr>
<tr>
<td>7</td>
<td>His answer was right.</td>
<td>It’s only a game.</td>
</tr>
<tr>
<td>8</td>
<td>He had a bad dream.</td>
<td>He gave us some paper.</td>
</tr>
<tr>
<td>9</td>
<td>I can’t remember her address.</td>
<td>The woman in line was angry.</td>
</tr>
<tr>
<td>10</td>
<td>A lot of people can sit at this table.</td>
<td>He listened and asked some questions.</td>
</tr>
</tbody>
</table>

3.3 Data acquisition procedure

The data acquisition procedure for learner speech consists of three stages: 1) pre-tests, 2) main pronunciation instruction for 5 weeks, and 3) post-tests as outlined in the beginning of Section 2. The instruction time is set the longest among all four experiments in Sections 2 and 3, because duration reduction of unstressed vowels and syllables may not occur, even after 5 weeks of main instruction. The recorded corpus of native and learner speech contains 2,160 utterances from 1,080 stressed and unstressed pairs of syllables (= 15 pairs × (24 learners × 2 recordings + 24 native speakers × 1 recording)). A total of 2,124 syllables have been acquired from the recorded corpus of 2,160 utterances, among which we have discarded 36 for the reason of poor signals. None has been discarded for other reasons, regardless of phonetic or phonological considerations.

3.4 Analysis

To analyze the stress reduction in native and learner speech, we first measure the duration of stressed and reduced vowels and syllables in the corresponding pairs of words and sentences before and after instruction. The duration measurement is illustrated in Figure 4, representing the spectral pattern and intensity level with periodic cycles.
As seen in Figure 4, we measure the duration of the reduced vowel [ə] and its embedding syllable [pʰə] for explanation. The syllable boundary between stressed and unstressed syllable sequences is subject to theoretical disagreements, as there are various theoretical positions on the syllable boundary in literature. The word may be differentially syllabified for the first [n] depending on ambi-syllabiccity, which is disallowed in this study for the sake of consistency.13

Another pitfall for duration measurement concerns the phrase-final lengthening effect. We measure the phrase final vowels as in explain and major, although it is subject to the phrase-final lengthening effect. The word major needs attention because the duration in native speech may be opposite to our expectation, and end up with a long reduced vowel in major, and short stressed vowel in majority, due to both the phrase-final lengthening effect and the multi-syllabic shortening effect. We examine such odd examples as well, since our important concern is the difference between native speech of a stress-timed language and learner speech of a syllable-timed first language. Our sentence data, on the other hand, are free from the phrase-final positioning of contrasting stress.

The measured duration values (in milliseconds), or the raw duration values are normalized by the mean duration between stressed and unstressed counterparts of the foot, i.e., Pairwise Variability. Among

---

13 Thus, the words in Table 1 are syllabified as the following: From the left column to the right, de.pre.ciate im.pli.ca tion ex.plain a.lege add; de.pre.ciate im.pli.ca tion ex.pli.ca tion a.lege a.lege a.lege; com.put.e con.fig.ure con.fig.ure ma.jor ity con.fig.ure ma.jor ity; com.put.e con.fig.ure con.fig.ure con.fig.ure con.fig.ure con.fig.ure ma.jor ity ma.jor ity ma.jor ity ma.jor ity.
various modifications of the mathematical formula, the method in the present paper is the simplest form: the value difference of the pair divided by the mean. The value is represented in percentage by multiplying it by 100 as in the following formula (7).

(7) Pairwise Variability (%) = \(100 \times \frac{2(d_s - d_w)}{(d_s + d_w)}\)

(\(d_s\) = the duration of the stressed vowel or syllable, \(d_w\) = the duration of reduced vowel or syllable)

In (7), a pair of stressed and unstressed counterparts is normalized by the mean values.\(^{14}\) The measurement purports to find whether the average value for stressed vowels is sufficiently larger than that of the reduced counterparts. The resulting values based on the Pairwise Variability correlate with the relative duration values in percentage between the comparing pairs.

3.5. Results and discussion

The results indicate that the duration of stressed vowels and syllables are significantly longer than their unstressed counterparts of native speech, but not so much for learner speech. Table 4 demonstrates the differences of Pairwise Variability.

<table>
<thead>
<tr>
<th>Word level reduction</th>
<th>Sentence level reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowel</td>
</tr>
<tr>
<td>Native speech</td>
<td>81 (30)***</td>
</tr>
<tr>
<td>Learner before</td>
<td>37 (41)+++</td>
</tr>
<tr>
<td>Learner_after</td>
<td>52 (42)aaa</td>
</tr>
</tbody>
</table>

*** \(p \leq .001\) of \(t\) values from \(t\)-tests for the duration differences between the stressed and reduced vowels in native speech.

+++ \(p \leq .001\) of \(t\) values from \(t\)-tests for learner speech before instruction to native speech.

aaa \(p \leq .001\) of \(p\) values from ANOVA tests for learner speech before instruction, after instruction, and native speech.

b insignificant of \(p\) values.

\(^{14}\) In another proposal (Jian 2004b), normalization is termed in the absolute value, in order to compare the variability of any successive syllables. For our own purpose, however, we do not use the absolute value, since the directionality in our data has to be fixed by the stressed or unstressed nature of the target syllables. In other words, we want to detect the negative value of the cases when the participants pronounce the stressed and unstressed counterparts in the reverse way.
Table 4 shows that Hypotheses 1-3 are supported. The average Pairwise Variability for native speech attests Hypothesis 1 that stress reduction is acoustically measurable in terms of temporal duration. The native speakers significantly ($p=.001$) shorten the duration of reduced vowels into (a) 81% of the average vowel duration in word pairs, (b) 75% of the average syllable duration in word pairs, (c) 76% of the average vowel duration in sentence pairs, and (d) 60% of the average syllable duration in sentence pairs.

The average Pairwise Variability for learners before instruction supports Hypothesis 2 that Korean learners have difficulty in duration reduction for unstressed vowels and syllables. The Pairwise Variability (%) of stressed vs. reduced units are significantly ($p=.001$) smaller for learner speech before instruction than native speech in terms of (a) vowel duration in word pairs (37<81), (b) syllable duration in word pairs (47<75), and (c) vowel duration in sentence pairs (48<76). The tendency is also observed in (d) for syllable duration in sentence pairs (53<60).

The average Pairwise Variability for learners after instruction supports Hypothesis 3 that the duration reduction of unstressed vowels and syllables in learner speech assimilates to the native values after instruction. The Pairwise Variability between reduced vowels and their stressed counterparts slightly increases in the order of learner speech before instruction, learner speech after instruction, and native speech for (a) vowel durations in word pairs (37<52<81; $p=.001$) and (b) syllable durations in word pairs (47<57<75; $p=.001$). The tendency is also observed for (c) vowel durations in sentence pairs (48<49<76; $p=\text{insignificant}$) and (d) syllable durations in sentence pairs (53<56<60; $p=\text{insignificant}$).

These facts demonstrated in terms of average values in Table 4 are also steadily observed for individual pairs of words and sentences, as shown in Figures 5, 6 and 7. Figure 5 demonstrates duration contrasts between stressed and reduced pairs of vowels and syllables in both word and sentence level utterances in native speech. The measured duration unit is indicated by a bracket in the abscissa. The duration (msec) is represented by the vertical bars, and their Pairwise Variability(%) by the horizontal lines. The $p$ values from the t-tests are $\leq .05^*, \leq .01^{**}, \leq .001^{***}$. 

Figure 5. Compared to the reduced counterparts, the duration in native speech is significantly ($p=.001, n=1,440$) longer when stressed for (a) vowels in words, (b) syllables in words, (c) vowels in sentences, and (d) syllables in sentences. (duration (msec) = (a) 165>68, (b) 325>148, (c) 147>67, (d) 284>171; $n=1,440; p=.001$)

In Figure 5 for native speech, the duration (msec) of stressed vowels and syllables (white bars) is significantly longer than their reduced counterpart (shaded bars) in each pair of morphologically-related words (eg., *add*: *addition*) at word level (a & b) and sentence level (c & d). The average duration (msec) for comparing vowels are (a) 165>68 for vowels in words, (b) 325>148 for syllables in words, (c) 147>67 for vowels in sentences, and (d) 284>171 for syllables in sentences. The only exception is the pair *major* and *majority* in the word level that involves an unbalanced comparison of two segments /or/ merged in the reduced position to one segment /o/ in the stressed position, in addition to the earlier discussed phrase-final lengthening effect and the multi-syllabic shortening effect.\(^{15}\) The corresponding Pairwise Variability values (line chart in %) other than this exceptional case are all positive, reflecting a relatively greater duration in the stressed units than in their unstressed counterparts. We thus consider

\(^{15}\) This example pair, as in others, *society* vs. *social* is also subject to the phrasal final lengthening and the multi-syllabic shortening effect, as we discussed earlier in Section 3.4. The pair, however, follows the generalization of a long stressed and a short reduced unit in sentence level ($p=.001$), in which the phrasal final lengthening and the multisyllabic shortening effects cancel each other in our stressed example, *This is nevertheless the majority.*
that stress reduction in native speech is acoustically measurable in terms of duration; i.e., the Pairwise Variability of vowels and syllables. Let us then conclude that Hypothesis 1 is attested, as in (8).

(8) **HYPOTHESIS 1 ATTESTED**: Stress reduction is acoustically quantifiable in terms of temporal duration.

Once stress reduction is successfully quantified for native speech, we are now able to test our second hypothesis that the same phenomenon is realized differently for learner speech whose native language is syllable-timed. That is, Korean learners would have difficulty in reducing unstressed vowels. Figure 6 demonstrates that stress reduction in native speech is significantly greater than learner speech for both word and sentence level utterances. This is numerically quantified in terms of Pairwise Variability between stressed and reduced pairs of vowels and syllables.

Figure 6. The values of Pairwise Variability (%) for stressed vs. reduced units are significantly ($p=.001$, $n=1,404$) greater for native speech than learner speech for (a) vowel durations in word pairs, (b) syllable durations in word pairs, (c) vowel durations in sentence pairs, and (d) syllable durations in sentence pairs. (Pairwise Variability (%) = (a) 81>37, (b) 75>47, (c) 76>48, (d) 60>53; $n = 1,404; p = .001$)

In Figure 6, Pairwise Variability (%) of stressed vs. reduced units are significantly greater for native speech (triangle markers) than learner speech (star markers). It is interesting to note that the tendency is more discernable for the vowels than syllables, and more for words than
sentences. From these results of stress reduction in terms of duration reduction, we conclude that Korean learners have difficulty in reducing the vowels in unstressed syllable, as in (9).

(9) **HYPOTHESIS 2 ATTESTED:** Korean learners have difficulty in reducing the duration of vowels in unstressed syllables.

The learner speech seems to be improved in the course of classroom instruction of pronunciation, and the improvement is significant in the word level. Figure 7 presents the Pairwise Variability of vowels and syllables in words and sentences of learner speech before and after instruction compared to native speech. It shows that learner speech after instruction moves toward the value of native speech.

![Figure 7](image)

In Figure 7, the Pairwise Variability (%) on average increases in the order of learner speech before instruction (star markers), after instruction (square markers), and native speech (triangle markers). The significance levels are indicated as *** (p ≤ 0.001), ** (p ≤ 0.01), and * (p ≤ 0.05).

In Figure 7, the Pairwise Variability (%) on average increases in the order of learner speech before instruction (star markers), after instruction (square markers), and native speech (triangle markers). The significance levels are indicated as *** (p ≤ 0.001), ** (p ≤ 0.01), and * (p ≤ 0.05).

16 A possible factor that affects consistency of computing the syllable duration can be the insertion of vowels in learner speech. For instance, the duration measurement of a stressed syllable in *add* may well be seriously distorted in learner speech [ædɪd] if the vowel [ɪ] is inserted.

On the other hand, a possible factor that the learners command the stress reduction rule better in word level than in sentence level may be due to a more complex linguistic property contained in the longer utterances (Kim and Flynn 2004). For instance, a paragraph-level utterance must deal with topics and focuses, as well as intonation that a word-level utterance does not need.
markers), and native speech (triangle markers). The improvement is more discernable in the word, but not in the sentence level. From these results of stress reduction in terms of duration reduction, we conclude that Korean learners show a small progress in acquiring the duration reduction of unstressed vowels and syllables in English, as in (10).

(10) HYPOTHESIS 3 ATTESTED: Duration reduction of unstressed vowels and syllables by Korean learners assimilates to the native values by classroom instruction of pronunciation.

For (10), it is important to note that the improvement is significant only in the word level, and the improved value does not reach close to the native level in 5 weeks of classroom instruction. In Figure 7, native speech shows the greatest Pairwise Variability between the stressed units and the reduced units than learner speech regardless of classroom instruction. For instance, the values of Pairwise Variability for the vowels in words are 37% and 52% for learner speech before and after instruction. They are significantly \( p = .001 \) in both cases) smaller as compared to 81% for native speech. Such a large difference of duration only in native speech, but not in learner speech may characterize that English undergoes the phenomenon of stress reduction in terms of vowel duration, and in a larger scale, a stress-timedness.

Next, we would like to consider if the learners’ proficiency in spoken English has been improved by confirming the validity of our independent acoustic measurements with a co-variant criterion of the listening comprehension task. The listening comprehension tests have been conducted twice before and after instruction with the methods described in the beginning of Section 2. The results are summarized in Table 5.

Table 5. Average scores of listening comprehension tests and their standard deviations (in parentheses) before and after instruction.

<table>
<thead>
<tr>
<th>Results</th>
<th>Number of utterances comprehended, out of 10 words and sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before instruction</td>
</tr>
<tr>
<td>Score (SD)</td>
<td>3.5 (1.7)</td>
</tr>
<tr>
<td>t-test</td>
<td>( p = .001 )</td>
</tr>
</tbody>
</table>

According to Table 5, listening comprehension has been improved after instruction. From this fact, we infer that the production improvement in Figure 7 accompanies an improvement of listening comprehension. From these results of stress reduction in terms of shortened vowel duration, we arrive at the conclusion in (11).
(11) Acoustic quantification of stress reduction by Pairwise Variability for duration.
   a. Native speech shows a greater reduction of duration for unstressed vowels and syllables.
   b. Learner speech shows a smaller reduction of duration for unstressed vowels and syllables.
   c. Korean learners show a progress in acquiring the duration reduction of unstressed vowels and syllables in English after 5 weeks of classroom instruction on pronunciation.

For native speech in (11a), the duration of vowels and syllables are longer when stressed, but shorter when unstressed in Figure 5. This follows the expectation that a stress-timed language has prominence only in the stressed syllable, where the prominence can, in part, be characterized by a longer duration. For learner speech in (11b) as opposed to native speech, the Pairwise Variability between stressed and reduced vowels and syllables is much smaller than native speech in Figure 6. This follows the expectation that a syllable-timed language does not reduce the temporal duration in unstressed vowels and syllables. For learner development in (11c), the duration values in learner speech move toward the pattern of native speech as the proficiency increases, as shown in Figure 7. Further, the improved command of duration reduction in unstressed vowels after instruction accompanies a better performance of listening comprehension task. This follows the expectation that the acquisition in production is part of overall proficiency development that involves perception as well.

From these findings in Section 3, we conclude that the stress reduction rules in English are acoustically cued by temporal duration. It is interesting to note that among all acoustic cues for stress alternation, assignment, and reduction, the duration reduction of unstressed vowels is hardest to acquire even after the longest instruction of 5 weeks. The findings from all the four experiments of non-native speech rhythm in Sections 2 and 3 are summed up in next section in terms of 1) native speech rhythm of English, 2) non-native speech rhythm of English by Korean speakers and 3) development of speech rhythm in adult language acquisition.

4. Conclusion

To conclude, the speech rhythm difference between stress-timedness and syllable-timedness is explicitly quantified in acoustic values between stress-timed native English and non-stress-timed learner English by Korean speakers whose native language is syllable-timed. On native speech rhythm of English, we found that English phonology of prominence, i.e., stress alternation, stress assignment, and stress reduction, is acoustically quantifiable in terms of duration, fundamental frequency, and vowel formants. On non-native speech rhythm, we found that the alleged timing
The contrast between stress-timed English and syllable-timed Korean is reflected only in stress alternation and reduction, but not in stress assignment. Given that stress timing is a by-product of the rules of stress assignment and stress reduction and that stress assignment rules have been reported to be successfully commanded by Korean learners in terms of pitch (J-M. Kim 2005), the stress-timed rhythm in English may then boil down to successful command of stress reduction. As for stress reduction, learner English as opposed to native English is significantly smaller in Pairwise Variability of duration for stressed and unstressed vowels. For another aspect of stress reduction, the $F_1$ and $F_2$ formant frequencies of unstressed vowels are centralized to the extent of schwa quality in native English, but not in learner English by native Korean speakers. The tendency is consistently observed in learner speech after instruction, whose production and perception values assimilate to native English. All these findings confirm the traditional distinction of speech-timing between stress-timedness and syllable-timedness.

4.1 Acoustic correlates of phonological prominence

What do these findings in speech rhythm offer to our understanding of acoustic correlates of phonological prominence? For native speech rhythm of English, the findings offer new insights and resolutions for the understanding of stress phonology in English in that the relevant phenomena of alternation, assignment, and reduction are, in fact, acoustically real, and can be quantified by specific acoustic measurements. Most important in this acoustic correlates is that we successfully quantified speech rhythm by means of Pairwise Variability of duration, but unlike any isochrony of stress in other previous studies that had failed (Roach 1982; among others).

For non-native rhythm of English as produced by Korean speakers, the acoustic correlates of speech rhythm are more prevalent in duration reduction of unstressed syllables. Fundamental frequency, another acoustic correlate of stress assignment rules, has turned out to be insignificant. These production results in our study are consistent to the perceptual study of English prominence by Kochanski et al. (2005: 1038), claiming that “loudness predicts prominence, while fundamental frequency lends little.” Considering the fact that our measurement excludes the duration of phrase final syllables, we can loosely correlate the duration in production in our study with perceptual loudness in their study.

These disparate production results of pitch from duration in learner speech are consistent with the perceptual study in Lim (2001), which claims that Korean speakers are more sensitive to relative pitch movement rather than to durational change. Lim attributes this to lacking vowel length distinction in Korean (also, refer to Ingram and Park 1997). S-J. Kim (2005) also supports our duration account of stress alternation. She claims
that native speakers of English correlate the unstressed weak vowel /ə/ with both shorter duration and lower pitch, while Korean learners correlate the vowel with lower pitch only. These studies all together are now consistent to our findings that the learning of segmental duration is independent from the learning of pitch.

Our duration reduction account can explain many acoustic findings in our study. First, it correctly predicts the non-native difference in Pairwise Variability for stress alternation, because learners would not have as much duration reduction of unstressed syllables as native speakers do. Second, it correctly isolates in the learner speech a successful command of stress assignment in terms of pitch from unsuccessful productions of stress reduction in terms of duration. By separating pitch from duration, the poor stress alternation is the result of a poor command of stress reduction, but not stress assignment in that both phenomena — stress alternation and stress reduction — depend on duration reduction of unstressed vowels. Third, it correctly predicts dispersed vowel formants of unstressed vowels in learner speech, because the Korean learners do not have as much duration reduction of the vowels as native speakers do. Duration reduction causes the undershooting of the full vowel quality in native English (Lindblom 1963; Flemming 2005), but not in learner English. In English, all reduced oppositions can be categorized into a single reduced vowel, schwa (Flemming 2005), while Korean has no reduced vowel at all in the phoneme inventory.

4.2 Unstressed vowels in phoneme inventory

The findings may be related to and derived from the absence of unstressed vowels in Korean phonology. An important implication is that non-native syllable-timed rhythm of a stress-timed language may depend on the absence of stress distinction in the vowel inventory of the native language, as indicated in Dasher and Bolinger (1982), among others. In this respect, we reconfirm that speech rhythm does not associate with isochrony, but with stress alternation that may depend crucially on the phonological distribution of unstressed vowels. In other words, what matters to speech rhythm is whether the phonological distribution or process meets stress alternation, a property that may accompany the presence of an unstressed vowel in phoneme inventory. In order to have stress-timed rhythm, English phonology requires stress alternation and, thereby, reduces an unstressed vowel. To contrast, Korean phonology with syllable-timing does not have unstressed vowel that may result in stress alternation.17 From our data,

17 There are less typical, mixed cases, as Nespor (1990) cites Polish, which has been classified as stress-timed but does not exhibit vowel reduction. Grabe and Low (2002) find an acoustic correlation with such mixed cases that Polish would be low on the vocalic axis and high on the intervocalic axis.
speech rhythm correlates with stress alternation in the phonological distribution and unstressed vowels in the phoneme inventory.

REFERENCES


FLEMMING, EDWARD. 2005. A Phonetically-Based Model of Phonological Vowel Reduction. Ms. MIT.


Non-native speech rhythm


Jong-mi Kim
Department of English Language and Literature
Kangwon National University
192-1 Hyoja-dong, Chuncheon-si, Gangwon-do
Korea 200-701
e-mail: kimjm@kangwon.ac.kr

Suzanne Flynn
Department of Linguistics and Philosophy
Massachusetts Institute of Technology
77 Massachusetts Avenue, Bldg. 32-D808
Cambridge, MA02139, U. S. A.
e-mail: sflynn@mit.edu