Syllable-timing Interferes with Korean Learners' Speech of Stress-timed English

Ok-hwa Lee* · Jong-mi Kim*

ABSTRACT

We investigate Korean learners' speech-timing of English before and after instruction in comparison with native speech, in an attempt to resolve disagreements in the literature as to whether speech-timing is measurable (Lehiste, 1977; Roach, 1982; Dauer, 1983 vs. Low et al., 2000; Yun 2002; Jian, 2004). We measured the pair-wise variability between the adjacent stressed and unstressed syllables within a foot as well as that among adjacent feet in approximately 555 English sentences, which were read by 29 native speakers and 41 Korean learners in the intermediate proficiency level. The results show that in comparison with native American English, Korean learner speech is before instruction significantly ($p<.001$) smaller for the pair-wise variability between the adjacent stressed and unstressed syllables within a foot; and significantly($p=.01$) bigger for the variability among adjacent feet within the utterance. The learner speech after instruction showed significant ($p=.01$) improvement in the pair-wise variability of syllable sequence toward native speech values. The variability among adjacent feet was progressively smaller for learner speech before and after instruction and for native speech ($p=.03$). We thus conclude that the speech timing difference between Korean English and American English is measurable in terms of the duration of stressed and unstressed syllables and that the latter is stress-timed and the former is syllable-timing interfered.

Keywords: speech-timing, syllable-timing, stress-timing, Pair-wise Variability, Foot Variability, Korean English, English by Korean learners

1. Introduction[1]

Speech-timing has long been considered as a phonological entity (Pike, 1945; among others), and yet difficult to quantify (Lehiste, 1977; Roach, 1982; Dauer, 1983). In recent literature, however,

the distinction of stress-timing and syllable-timing manifests in acoustic measurements of non-native speech of English: Mochizuki-Sudo and Kiritani (1991) for Japanese English, Low *et al.* (2000) for Singapore English, Jian (2004) for Taiwanese English, and Lee and Jang (2004) for Korean English. Among these, we extend the investigation in Low *et al.* (2000) and Jian (2004) to Korean learners of American English as a foreign language. They investigated the acoustic correlates of speech rhythm in Singapore English in comparison with stress-timing in British English (Low *et al.*, 2000) and that in Mandarin English in comparison with stress-timing in American English (Jian, 2004). In their measurements of the Pair-wise Variability of the adjacent syllable pairs, both Mandarin and Singapore varieties of English displayed a smaller duration discrepancy between stressed and unstressed syllables as opposed to British and American varieties.

In this study, we investigate the acoustic features of speech-timing in the utterances of Korean learners of American English. Since Korean is syllable-timed and not stress-timed (Lee and Jang, 2004; Kang, 2004; Yun, 2002), we expect to find a smaller duration discrepancy between adjacent stressed and unstressed syllables in the English speech by Korean learners than those by native speakers.

To measure the duration difference between adjacent units, researchers (Low *et al.*, 2000; Jian, 2004) have used Pair-wise Variability Index. Although specifics of the formula differ from researcher to researcher, it essentially produces the duration variability among adjacent units after normalizing the duration difference within a foot by its average syllable duration for each pair. For this measurement, a greater variability among the adjacent syllables is expected for stress-timing. This leads to our first hypothesis.

(1) **Hypothesis 1**: The Pair-wise Variability for adjacent stressed and unstressed syllable sequences is significantly greater in American English by native speakers than that by Korean learners.

For acoustic measurement, the Pair-wise Variability for a pair of stressed and unstressed syllables is expected to be close to 0% in a perfect isochrony of syllable-timing, where there is no duration difference between stressed and unstressed syllables. On the other hand, the value would be significantly greater than 0% in stress-timing.

The Pair-wise Variability correlates to the relative duration or length of time, in that both deal with the relative value; but for the former in terms of differences, and for the latter in terms of proportions. The relative duration of an unstressed syllable to that of a stressed syllable is expected to be near 100% in a perfect isochrony of syllable-timing, but significantly smaller than 100% in stress-timing. A perfect stress-timing would then have the acoustic characteristics, in that the duration of each foot is expected to be the same, i.e., 0% in Pair-wise Variability or 100% in relative duration.

(2) **Hypothesis 2**: The native speakers of American English show a similar duration among
feet within an utterance, while Korean learners have varied durations.

In acoustic measurement, the relative duration of each foot to an average foot within an
utterance is close to 100% for perfect stress-timing, while the value varies for syllable-timing.
In particular, the inter-stress interval is expected to be constant regardless of the number of
intervening unstressed syllables in perfectly stress-timed rhythm. This contrasts to
syllable-timing where the inter-stress interval progressively increases as the number of
intervening syllable increases.

If the numerical difference corresponds to the characteristics of speech-timing, we may
expect that the learners will show a smaller discrepancy to the target value as their proficiency
of the target language improves, upon which Hypothesis 3 is proposed below.

(3) **Hypothesis 3**: The relative duration and Pair-wise Variability for adjacent stressed and
unstressed syllable sequences become progressively closer to the target
values as the learners' proficiency improves.

The developmental study for speech-timing as described in Hypothesis 3 is first launched in
this study to our knowledge. In order to investigate this, we study a large number of subjects
to further validate our findings. This differs from the previous investigation in speech-timing, in
that the number of subjects is sufficiently large (41 learners and 29 native speakers) to identify
the core acoustic cues of speech-timing.

2. Method

To quantify the speech-timing of Korean English in terms of duration ratio between adjacent
stressed and unstressed syllables, we measured the syllable duration of stressed and unstressed
counterparts in approximately 555 English sentences as produced by 29 native speakers of
American English and 41 Korean learners of English in the intermediate proficiency level.

2.1 Participants

Research participants consisted of 41 Korean learners of English, 1 native speaker of
American English who provided the sample speech, and 28 other native speakers of American
English for data comparison. All participants had normal hearing, vision and speech articulation,
without a significant medical or psychiatric history. The model native speaker was an English
professor at a college, 45 years old, male, from Utah. The native speakers were mostly from
Northeast, Midwest, West, Northwest region of American English (92%), while the Korean

learners from Midwest region of Korea (95%).[2]

For the purposes of the present study, we attempted to hold the biological and educational variables constant for our Korean learner participants. The ages of the Korean participants were all in their twenties, although the ages of the Americans ranged from teenagers to fifties with the average of 29. All participants were college-educated, if not attending at a high school as a teenager (24%) for American participants. All Korean participants were enrolled in an English pronunciation class at a Korean college at the time of experiment. All students spoke only Korean at home.

The English proficiency level of Korean learners was intermediate in the light of listening and reading proficiency. Their average score of the standardized test, TOEIC in a 990 scale, was 586, with the standard deviation of 118.[3] None of the learners was in the range of low (10-200) or high (710-990) proficiency. We also evaluated the learners' proficiency level by an independent listening comprehension test that was administered before and after a 3-week instruction. The listening comprehension test was comprised of 9 dictation items to classify the intermediate learners in a 9-point scale. In the listening comprehension pre-test administered before a 3 week instruction, the learners' average score was 5.2 with the standard deviation 1.2. None of the learners belonged to the low (0, 1, 2) or high (9) level. We will return to the details of the listening comprehension tests in Table 1 and Table 4 in later sections in order to show their developmental correlation with the pronunciation production tests.

2.2 Speech materials

We discuss two types of speech materials: those speakers produced and perceived. The production speech materials were read by all participants for recording, and the perception speech materials were dictated by only Korean learners.

For production materials, we employed the sentences in Low et al. (2000) and Jian (2004) for cross-linguistic comparison. The sentences are listed in (4), where the stressed syllables are capitalized for expository purposes.

(4)   a. JOHN was SICK of FRED and SANdy
       b. DON was aCROSS at JONathan's
       c. PAULa PASSED her TRIAL of COURage
       d. JANE has FOUR to LAST the WINter
       e. GRACE was TIRED of MATthew FREEman.

2) The speakers were asked to mark the years spent in various parts of United States, and many had traveled around.

3) The linking of scores to proficiency levels is outlined in "TOEIC Can-Do Guide," provided by the Educational Testing Service. They rank the scores by 1) 10-200, 2) 210-450, 3) 460-700, 4) 710-850, and 5) 860-990. The Korean learners in this study belong to the rank 3).

The sentences in (4) demonstrate the stress alternation in adjacent syllables. The first syllable in each stimulus is a stressed syllable, while the second is an unstressed syllable. The third syllable may or may not be a stressed one, as in (4a) for a stressed one "*SICK*," and (4b) for an unstressed one "*a*." Learners did not have any capitalized cue for their stimulus list. We used the same set of production test materials before and after instruction.

The perception speech materials were to test the listening comprehension proficiency of the learners, and recorded by the model native speaker. There were two similar sets of perception test materials before and after instruction.

Table 1. Utterance list for the listening comprehension tests before and after instruction

	Test Set I before Instruction	Test Set II after Instruction
1	engineer	hostess
2	himself	parade
3	overthrow	runaway
4	An elf.	Not bad
5	In a row.	In the airport.
6	He was here.	She needed it.
7	I can understand.	I can't believe it.
8	He wants to leave.	I'll have a cup of coffee.
9	Please pass the pepper.	Leave the key at the desk.

For the two tests in Table 1, we tried to maintain the same task difficulty level to understand whether the listening comprehension level would be improved or not. Both stimulus lists were read by the same model speaker in the same recording session, but tested at a different time: Set I., before the instruction, and Set II., after the instruction.

2.3 Data acquisition procedure

The data acquisition procedure for the Korean learners' utterances consisted of three stages: 1) pre-listening and production tests after the introductory pronunciation instruction on the recording stimuli, 2) main pronunciation instruction for 3 weeks, and 3) post-listening and production tests. We tested the learners before and after the instruction to investigate their developmental patterns. On the other hand, the native speakers were involved in only one stage of recording.

At the first stage of the data acquisition procedure, the Korean learners were first given a 20-minute pronunciation instruction, during which all 5 sentences in the stimulus list and some other 5 sentences were presented in a handout. The teacher (the first author) explained to the students the meaning of all the words in the list. The learners listened to and repeated after a model speech for the list for one time. Then, they were asked to go to the two quiet rooms at a scheduled time during the day, and were allowed to do more practices on their own if time had permitted. They were then asked to read the stimulus materials for recording. The

recording was digitally done in 16 KHz, 16 bit by trained teaching assistants and the devices used were Computerized Speech Lab by Kay Elemetrics, and/or other PC equipments. The distance of the microphone from the participants' mouths was adjusted to the amplitude of their voice displayed by the waveform signal of test speech. Utterances with errors or hesitations were re-recorded at the end of each session, when requested by the learner. All learner recordings were completed in one day. A brief listening comprehension test on the speech materials in Set I. in Table 1 was administered three days later to compare with the pronunciation development on speech timing.[4] The net time spent for the listening test lasted only for 5 minutes, in order not to take away the class time that should concentrate on the production instruction and assessment.

At Stage 2, the learners were taught the regular class materials including the expected shortening of unstressed syllables in the embedded words in English.[5] During the instruction, they listened to and repeated again after the model native speech in the stimulus list. Further, the model native speech material was distributed in CD, and by web-downloads. The pronunciation lessons lasted for 3 weeks.

At Stage 3, the learners were briefly tested again on the listening comprehension with the speech material Set II in Table 1. Then, they were asked to read and record the stimulus materials again in two quiet rooms, but only once in the post-test. All recordings were completed in a day.

In contrast, the native speakers in this study did not have any lesson or training session in English. The native speakers were immediately asked to read the recording materials twice that were mixed among other approximately 60 utterances. And yet, all native speakers answered that they understood the meanings of all utterances in the recording list. The second-time recording was taken for the data, although the first-time recording was occasionally taken to replace a weak, noisy or unnatural speech signal. The recorded corpus of native and learner data contained 1,554 stressed and unstressed pairs of syllables (= 14 pairs x (41 learners x 2 recordings + 29 native speakers x 1 recording), excluding the sentence final feet.

2.4 Analysis

To analyze the data, we measured the duration of the stressed and unstressed syllable constituents of a foot. Our method of measurement is illustrated in Figure 1, which represents the spectral pattern, F0 trace, and intensity level with periodic cycles.

---

4) During the listening test, the learners were asked to write down as hearing the recorded speech materials in Table 1.

5) The teacher taught some segmental and rhythm sections in the textbook by Grant (2000).
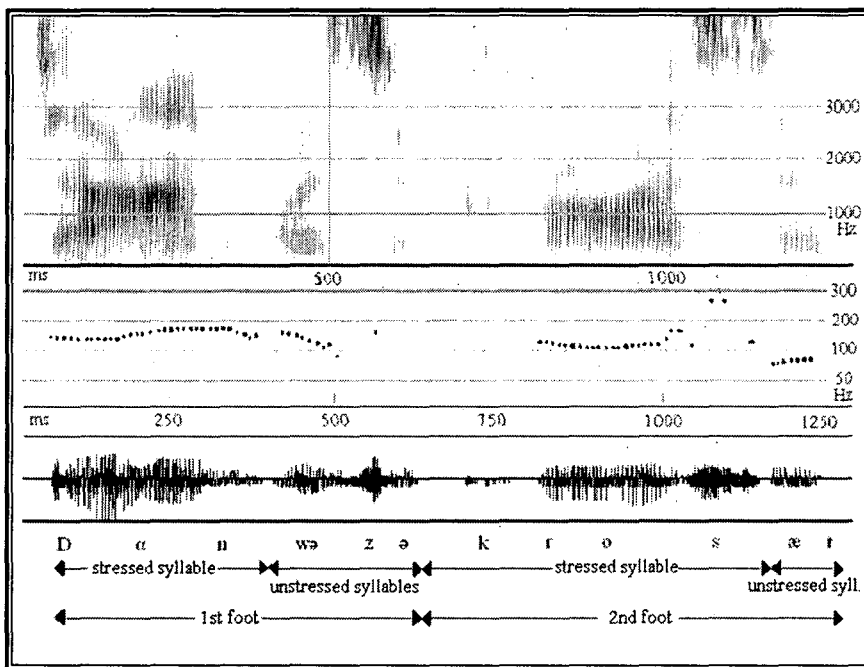
Figure 1. Duration measurement of the syllables and feet for the sentence
"Don was across at Jonathan's."

As seen in Figure 1, we measured the duration of the stressed syllable "Don" and the subsequent unstressed syllable sequence "was a" of the first foot, as well as the duration of the stressed syllable "cross" and the subsequent unstressed syllable "at" of the second foot for the sentence, "Don was across at Jonathan's." We did not measure the phrase final foot, "Jonathan's," because it is subject to the potential phrase-final lengthening effect. Neither we measured pauses and repairs of the learners, caused by lack of proficiency. We primarily used the software Speech Station Ⅱ (commercially available by Sensimetrics), although we also compared the unclear spectral views or abnormal F0 traces with different software such as CSL by Kay Elemetrics, or PitchWorks by Scicon R&D.

The syllable boundary location between stressed and unstressed syllable sequences as in Figure 1 is subject to theoretical disagreements, as there are various theoretical positions on the syllable boundary in literature. The word sequence "across at," for instance, may be differently syllabified with or without ambi-syllable as in "a.cross.at," "a.cro.ssat," and "a.cros.sat." We located the syllable boundaries at word boundaries, if available, as in "a.cross.at." This is because the word boundary is often blocked from re-syllabification in learner English, by the insertion of a dummy vowel or a glottal stop. In other cases, we followed the principle of Maximal Onset and Minimal Coda by Hammond (1998:246), where he syllabifies the words "can.a.da, ba.nan.a, rac.coon, ba.lloon." We consider our data, "PAUL.a" in (4c) as the analogous

example. For the data acquisition purposes, we excluded all the sentential final feet to be free from the potential phrase-final lengthening effect. All the measured data were, then, employed as "the raw duration values."

The raw duration values (in milliseconds) were normalized by the mean duration between stressed and unstressed counterparts of the foot. This normalization method has been referred to as Pair-wise Variability: A measure of the difference between the two comparing values within a pair, as normalized by their mean value. The measure has widely been used in studies of non-native speech rhythm to compare with the variability between two related pairs of speech samples (Low *et al.*, 2000; Jian, 2004, among others). Among various modifications of the algorithm, the method in the present paper is the simplest form: the value difference of the pair divided by their mean. The value is represented in percentage by multiplying it by 100.

(5) *Pair-wise Variability* (%) = 2 $(d_s - d_w)$ / $(d_s + d_w)$ ✳ 100

    (where: Pair-wise Variability (%) = normalized Pair-wise Variability between a stressed syllable and one or more unstressed syllables within a foot, d_s = the duration of the stressed syllables in a given foot, d_w = duration of unstressed or weak syllables in a given foot)

In (5), a pair of stressed and unstressed counterparts was normalized by the mean values. The measurement purports to find whether the deviation from the mean value for stressed vowels had been sufficiently larger than that of their unstressed counterparts. The resulting values based on the Pair-wise Variability correlate with the relative duration values in percentage between the comparing pairs.

In other proposals (Jian, 2004), the normalization was termed in the absolute value, in order to compare the variability of any successive syllables. For our own purpose, however, we did not use the absolute value, since the directionality in our data had to be fixed by the stressed or unstressed nature of the target syllables. In other words, we wanted to detect the negative value of the cases when the learners had pronounced the stressed and unstressed counterparts in the reverse way.

We further computed the duration variation among adjacent feet, when the number of intervening unstressed syllables was different. The measurement of foot variability was also normalized in terms of the average foot duration within the utterance.

(6) *Foot Variability* (%) = n $\sum\limits_{i=1}^{n}$ ABS$(F_{i-1} - F_i)$ / $(F_{i-1} + F_i)$ ✳ 100

    (where: Foot Variability (%) = normalized foot variability, ABS = absolute value, F_{i-1} = duration of the preceding foot, and F_{i-1} = duration of the given foot)

Foot Variability in (6) averages out the duration difference of the adjacent feet within an utterance. We expect that a perfect stress-timed utterance will have the value of 0%. Thus, for our own research purpose, the native speech of American English would have a smaller Foot Variability than the learner speech. The difference would be tested by the varied number of syllables within a foot (i.e., varied number of intervening unstressed syllables). This is because a foot, by definition, contains only one stressed syllable and one or more unstressed syllable(s). We made a scale of the value in percentage by multiplying it by 100 to facilitate reading.

3. Results

The results indicate that learner English is more syllable-timed than that of native English in terms of the duration of stressed and unstressed syllables. The following Table 2 presents Pair-wise Variability for the duration of stressed and unstressed counterparts within a foot as normalized by the average syllable duration of the pair. The values are compared to the learner speech before and after instruction and native speech. The data present all 1,554 pairs of stressed and unstressed syllable sequences acquired in this experiment.

Table 2. Average Pair-wise Variability of unstressed syllables over stressed syllables in learner speech as compared to that in native speech. The unstressed syllables are progressively shorter in learner speech before and after instruction and native speech (n=1,554).

	John: was	sick: of	Fred: and	Don: was a	cross: at	Paul: a	passed: her	trial: of	Jane: has	four: to	last: the	Grace: was	tired: of	Mat: thew	Aver- age
Learner _before	-1***	24**	73**	-49***	62***	83*	69*	88	6	48***	104***	42***	78***	16***	46.0***
Learner _after	16[a]	39	60	-39	66	67	64	93	8	61	116	43	81	36[aa]	50.7[a]
Native	49···	78···	91···	38···	117···	101···	79···	96···	9···	107···	138···	74···	109···	52···	81.3···

*<.05, **<.01, ***≤.001 of p values from t-tests of the learner speech before instruction to native speech
[a]<.05, [aa]<.01, [aaa]≤.001 of p values from t-tests of the learner speech before instruction to after instruction
···≤.001 of p values from t-tests of the duration differences between the stressed and unstressed portions of a foot

In Table 2, the average Pair-wise Variability was as large as 81.3 % in the native speech of American English, while only 46.0 % in learner speech before classroom instruction (p<.001). In other words, the duration difference between the stressed and unstressed parts of a foot was significantly (p<.001) greater in native speech than in the learner speech before

classroom instruction. The t-test results marked by asterisks (*) show a significant contrast ($p<.05$) between the native speech and the learner speech before instruction for the majority (86%) of the data. The only less significantly contrasting pairs were: "trial/of" and "Jane/has." In fact, all the pairs of words showed greater Pair-wise Variability in native speech than those in the learner speech before instruction. The t-test result marked by superscript crosses (⁺⁺⁺) shows a very significant contrast ($p\leq.001$) between the duration of the stressed syllable part and the unstressed syllable part of the given foot in native speech. In fact, every individual foot in native speech, but not in learner speech shows the statistically significant level of $p\leq.001$.

An important observation in Table 2 is that the Pair-wise Variability is -49% in the learner speech for the stressed and unstressed syllable sequence of the foot "Don was a." The negative value indicates that the learners read the unstressed syllable sequence much longer than its stressed counterpart for the foot. The average duration of the unstressed part "was a" was 384 milliseconds, while the stressed counterpart "Don" was 230 milliseconds.[6] On the other hand, the native speech shows the positive Pair-wise Variability of 38%, which means that the native speakers read the unstressed syllable sequence much shorter than the stressed counterpart for the same foot. The average duration of the unstressed part "was a" was 180 milliseconds, while the stressed counterpart "Don" was 267 milliseconds.

The correlation between speech-timing and acoustic duration in Pair-wise Variability is further supported by the developmental aspect of the learners whose values in overall significantly ($p=.01$) improved from the average 46% to 50% toward the native value. The improvement per individual syllable sequences however, did not show significant improvement in many cases. As marked by the superscript 'a' symbol, only 15 % of data showed a significant improvement.

The following Figure 2 graphically presents Pair-wise Variability in Table 2. Figure 2 presents the values of Pair-wise Variability in (6) that normalizes the duration differences of stressed and unstressed counterparts in terms of the average syllable duration.

---

6) The difference of the two comparing components is 154 msec, which is half(50%) of their average duration 307 msec. The reported value 49% in Table 2 is the average of the individual Pair-wise Variability of every speakers, which is more accurate than the calculation from average duration.
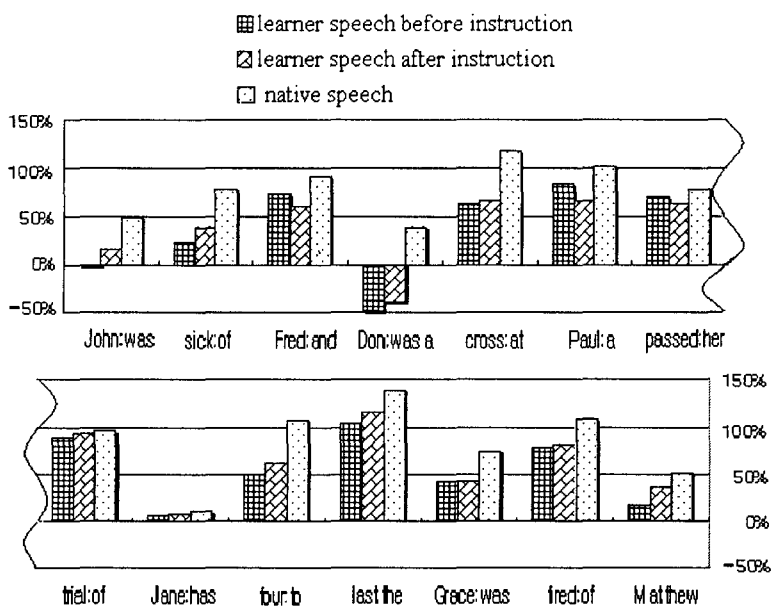
Figure 2. Pair-wise Variability for the duration between stressed and unstressed syllables increases in the order of learner speech before instruction, after instruction, and native speech (Pair-wise Variability (%) = 46.0 < 50.7 < 81.3 ; n=1,554 ; p<.001).

In Figure 2, Pair-wise Variability for each foot progressively increases for learner speech before instruction, after instruction, and native speech. Native speech show significantly bigger Pair-wise Variability than learner speech before instruction. The value near 0% means that there was no significant difference between the durations of stressed and unstressed portions of the foot. In the instances of the stressed syllable "John" and the following unstressed sequence of syllable "was," the learner speech before instruction on the average showed -1% Pair-wise Variability. This indicates that the learners produced both the stressed and the unstressed syllables with almost half duration each within the foot. This is expected for the syllable-timed rhythm as a perfect example of isochrony. In stress-timed rhythm as in native speech, the stressed syllable "John" is longer than the following unstressed syllable "was" within the foot, as the duration difference between them is about 49% of the average syllable duration of the foot. Figure 2 demonstrates that Hypotheses 1 and 3 are correct.

(8) **Hypothesis 1 Attested**: The Pair-wise Variability for adjacent stressed and unstressed syllable sequences is significantly greater in American English by native speakers than that by the Korean learners

(9) **Hypothesis 3 Attested**: The relative duration and Pair-wise Variability for adjacent stressed and unstressed syllable sequences become progressively closer to the target values as the proficiency improves.

Hypothesis 3 is only partially attested, because it also needs to be supported in terms of the relative duration. The following Figure 3 presents the relative duration (%) between the stressed and unstressed syllables in native speech, and learner speech before instruction and after instruction.
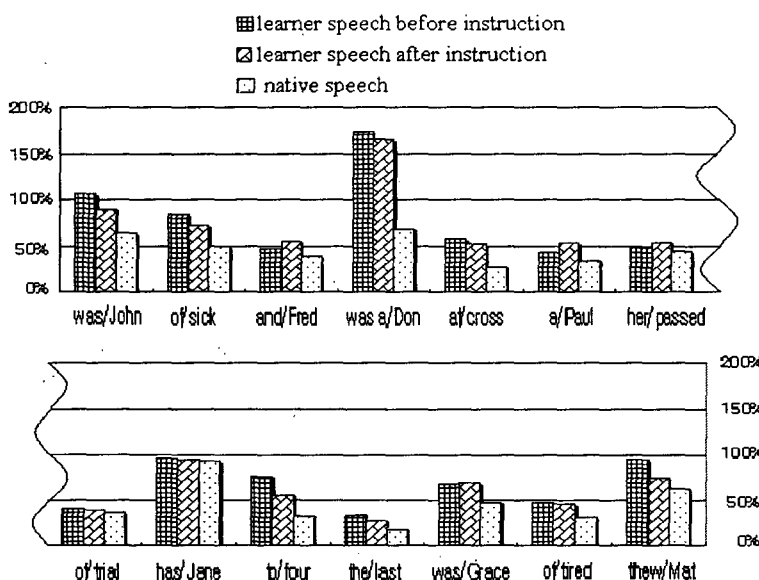

Figure 3. Relative duration of unstressed syllables to stressed syllables decreases in the order of Korean learner speech before instruction, after instruction, and native speech (Relative duration (%) = 72.4 < 67.5 < 45.8 ; n=1,554 ; p<.001).

In Figure 3, the relative duration of unstressed syllables over the preceding stressed syllables progressively decreases for learner speech before instruction, after instruction, and native speech. The value near 100% means that there was no reduction. All native speech show a large reduction as the duration of unstressed syllable sequence drops below 70% of the stressed syllable within a given foot, except the foot "Jane has." Nevertheless, even this pair of syllables showed a significant (p<.001) reduction of 91±11 (%) from the average duration value 270 of "Jane" to the value 247 ms for the word "has."

In contrast, the Korean learners did not significantly reduce the unstressed syllables over their stressed counterpart within the foot. In the instances of the stressed syllable "Don" and the following unstressed sequence of syllables "was a," the learners produced the unstressed sequence of syllables much longer than its comparing stressed syllable within the given foot. This is expected for the syllable-timed rhythm because the unstressed sequence "was a" has

two syllables, while its stressed counterpart "Don" has one syllable. In stress-timed rhythm, the unstressed sequence "was a" is shorter than its stressed counterpart "Don" within the foot as in native speech in Figure 1. The facts further support that our Hypotheses 1 and 3 are correct.

(10) **Hypothesis 1 Attested**: The Pair-wise Variability for adjacent stressed and unstressed syllable sequences is significantly greater in American English by native speakers than that by the Korean learners.

(11) **Hypothesis 3 Attested**: The relative duration and Pair-wise Variability for adjacent stressed and unstressed syllable sequences become progressively closer to the target values as the proficiency improves.

Hypotheses 1 and 3 are now attested for both duration ratio and Pair-wise Variability. To recapitulate our developmental results, thus far, on stress alternation, we summarize the results in Figure 1 and 2, and present them all in Figure 4 for a holistic view of the difference between native and non-native speech rhythm at a glance. The data present all 1,554 pairs of words.
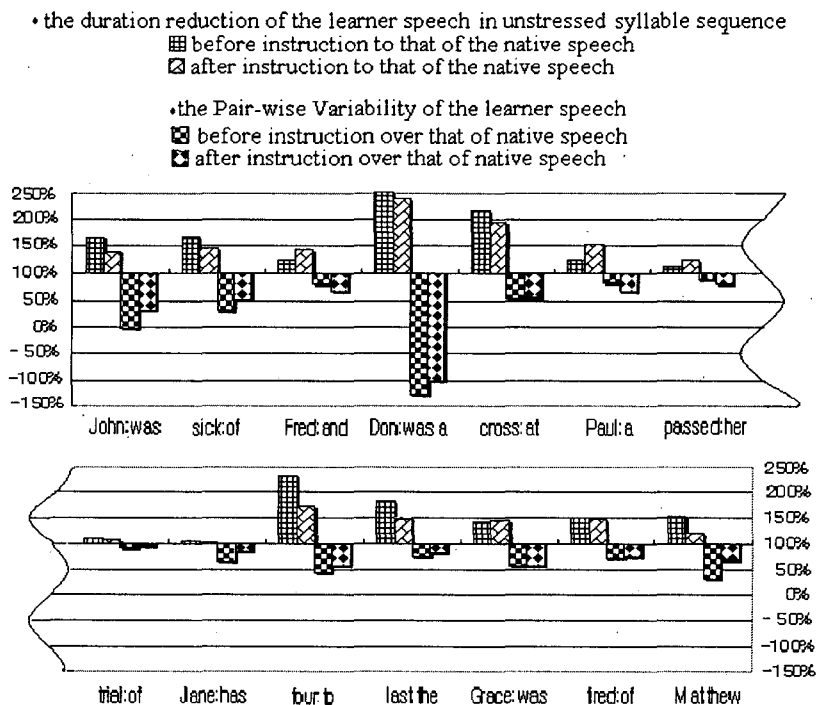

Figure 4. Duration reduction in learner speech decreases after instruction toward the native speech value (100%), while the Pair-wise Variability increases (n = 1,554 ; Pair-wise Variability (%) = 45.8 < 55.4, p = .01 ; Relative duration (%) = 160.6 < 148.8, p = .01).

In Figure 4, the most ideal case would be that the bars distribute near 100%, in which the learners speak like the natives. In other words, the English pronunciation proficiency of the Korean learners is the same as the English natives' one. What we find in Figure 4 is the duration ratio above the standard line (100%) and the Pair-wise Variability below the standard line. This indicates that the Korean learners do not make a durational distinction between stressed and unstressed syllables within a foot as the native speakers do.

All these results in our previous discussion support the fact that stressed syllables are longer than the unstressed syllable sequence within a foot. Then, would stressed syllables come in a more regular interval for stress-timed rhythm than those for syllable-timed rhythm? To test this, we need to compare the feet whose inter-stress interval contains a different number of syllables. In the stimulus list in (4), the foot sequence in the phrase-final feet of the sentence in (4b), "DON was aCROSS at Jonathan's" is a candidate. We compute the duration difference of the two feet, [DON was a] and [CROSS at], that are normalized by the average foot duration of the utterance. We exclude the final foot, because its length may be affected by phrase-final lengthening. The following Table 3 presents the Foot Variability, as computed according to the formula in (6). The data represents a total of 222 feet (2 feet x (41 learners x 2 recordings + 29 native speakers x 1 recording)).

Table 3. Foot Variability of the sentence "Don was across at Jonathan's." The variability (%) is decreases in the order of learner speech before instruction, in after instruction, and native speech. (n=222, p=.03)

	Foot Duration (ms)		Foot Variability (%)	
	Don was a	cross at	Average	Standard Deviation
Learner before instruction	600···	500	30.9**	18.9
Learner after instruction	550···	450	27.8***	20.5
Native speaker	450°	400	18.9	11.3

<.01, *≤.001 of p values from t-tests of the learner speech to native speech
+++≤.001 of p values from t-tests of the duration differences between the comparing adjacent feet
° insignificant p values from t-tests of the duration differences between the comparing adjacent feet

In Table 3, Foot Variability in native speech (=18.9%) is significantly (p<.001) smaller than those in learner speech (=30.9%, 27.8%) regardless of the timing of instruction. The t-test results marked by superscript crosses (···) show a significant contrast (p<.001) between the comparing adjacent feet in the learner speech before and after instruction, but not for native speech. As the t-test result indicate, the foot duration with more syllable numbers was significantly greater than its counterpart in learner speech, but not in native speech. This indicates that native speech follows more stress-timed rhythm, while learner speech obeys more duration additions by syllable numbers.

The developmental aspect of the Foot Variability was not shown to be significant (p=.46), although the average value improved from 30.9% before instruction to 27.8% after instruction toward the native value. The numerical results in Table 3 are graphically presented in Figure 5.
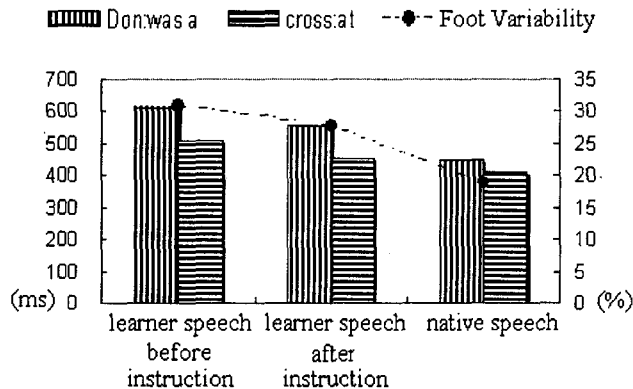

Figure 5. Foot Variability of the sentence "Don was across at Jonathan's." The variability decreases in the order of learner speech before instruction, after instruction, and native speech of American English respectively (Foot Variability (%) = 30.9 < 27.8 < 18.9 ; n=222 ; p=.03)

In Figure 5, Foot Variability in native speech is significantly less than that in learner speech before instruction (18.9 < 30.9 ; p = .01). The result entails that our hypothesis about Foot Variability is attested.

(12) **Hypothesis 2 Attested**: The native speakers of American English show a similar duration among feet within an utterance, while the Korean learners have a varied duration.

In (12), we note that the result in Figure 5 confirms Hypothesis 2 on the cross-linguistic difference of Foot Variability. Notice, in addition, that the results in Table 5 also lean toward our developmental expectation that learner speech has produced less Foot Variability toward the native speech value, supporting our Hypothesis 3 as well.

Having demonstrated the developmental aspect of stress reduction in Figure 2-5 and Tables 2-3, we would like to consider whether the learners' speech proficiency has indeed been improved. We would like to confirm the validity of our acoustic measurements by the independent co-variant criteria of the listening comprehension task. The listening comprehension tests were conducted twice before and after instruction by the methods described in Table 1 of Section II. The results are summarized in Table 4.

Table 4. Listening comprehension improvement of the Korean learners after instruction (n=82)

Results	Number of utterances comprehended, given 9 words and sentences	
	before instruction	after instruction
No. ± SD	5.2 ± 1.3	7.5 ± 1.3
t-TEST	p<.001	

In Table 4, the learners' listening proficiency has been improved after instruction. From this fact, we infer that the production assimilation in Figure 2-5 and Tables 2-3 accompanies competence improvement in listening comprehension in Table 4. In other words, the production assimilation in Figure 2-5 and Tables 2-3 is considered as part of the overall linguistic development rather than a mechanical manipulation of articulation for a specifically given task in a narrow scope.

4. Conclusion

Based on the results in the preceding sections, we conclude that speech-timing is acoustically measurable in terms of the differences in the Pair-wise Variability and Foot Variability. The speech-timing difference was explicitly quantified in acoustic values between stress-timed American English and Korean learner speech of English whose native language is known to be syllable-timed.

In comparison with native American English, the duration difference between the stressed and unstressed syllable sequences of a foot was significantly (p<.001) smaller in the learner speech before classroom instruction. In other words, the timing in native speech is more obedient to stress alternation between adjacent syllables than in learner speech, and thereby it is construed as being more stress-timed. The correlation between speech timing and acoustic duration was further supported by the learners' developmental aspect, whose Pair-wise Variability significantly (p=.01) improved to be stress-timed after the classroom instruction.

In addition, Foot Variability between adjacent syllables was progressively smaller for learner speech before and after instruction and for native speech (p=.03). This follows our expectation of the rhythmic difference that a stress-timed language manifests less varied foot duration. American English shows contrastively smaller foot duration than Korean learner speech whose native language is syllable-timed.

Let us then recapitulate the issue "What makes the difference in speech-timing?" Phonologically, we consider stress alternation between the successive syllables characterizes stress-timing. That is, a stressed syllable makes a pair with one or more unstressed syllables to form a foot. Without the general pattern of stress alternation, speakers would not feel the

rhythmic beat in terms of stress location. In our study, the phonological rule of stress alternation for successive sound sequence has been adequately represented by large acoustic Pair-wise Variability between the duration of stressed syllable and the unstressed counterpart of a foot. In contrast, syllable-timing would not have the duration distinction between stressed and unstressed syllables, so that the speakers feel the rhythmic beat by syllable unit. This has also been acoustically represented as smaller Pair-wise variability. Although our data is limited to Korean English, they consistently conform to the data in Singapore English (Low *et al.*, 2000) and Taiwanese English (Jian, 2004) to support the traditional distinction of stressed and syllable timing as the two-contrastive axes of rhythmic types.

What follows is another implication of speech-timing that the stressed syllable comes at regular intervals in stress-timed languages. The expectation falls out in the sense that the duration of unstressed syllables contracts, but not the stressed ones. We have seen the contraction in the light of Foot Variability, where we had much less variability in native speech than in learner speech when the given foot contains more than one unstressed syllables. Foot Variability, a new measure of rhythm proposed in this study, has effectively calculated the contraction differences of multi-syllabic unstressed sequences in native speech from learner speech that has different types of speech rhythm. Although we have shown a limited amount of data, due to the fact that the data is taken from other studies for cross-linguistic comparison, our investigation conforms to Lee and Jang (2004) who concluded that inter-stress intervals are more sensitive to syllable numbers in Korean learners' speech than in native speech of English.

In addition, our results in developmental study demonstrated consistent transition from one type of rhythm to another with respect to both Pair-wise Variability and Foot Variability. This indicates that learners assimilate their rhythm to the target language values as their proficiency increases. In particular, Pair-wise Variability of all items in a group improved significantly (p=.01), although the improvement was not statistically so significant for many individual items in isolation for t-tests. Such less conclusive improvement in stress alternation in terms of duration drastically contrasts to the instantaneous learning of stress assignment in terms of pitch (Kim, 2005) for Korean learners of English. We may conjecture, then, for the present purpose, that stress alternation in duration may be a slower learned linguistic property than stress assignment in pitch when Koreans learn English.

To conclude, the native and learner differences in Pair-wise Variability and Foot Variability in this study conform to the traditional distinction of speech-timing in that English is stress-timed, whereas L2 Korean-English is interfered by L1 syllable-timing.

References

Dauer, R. M. 1983. "Stress-timing and syllable-timing reanalyzed." *Journal of Phonetics*, 11: 51-62.

Grant, L. 2000. *Well Said*. 2nd ed. MA: Heinle & Heinle.

Hammond, M. 1999. *The English Phonology of English*. Oxford University Press.

Jian, H-L. 2004. "An improved pair-wise variability index for comparing the timing characteristics of speech." *ICSLP Proceedings*, 1261-1264.

Kang, Y. 2004. *Understanding English Stress*. [In Korean]. Seoul: Kyungjin.

Kim, J-M. 2005. "Stress assignment rules in Korean English." *Studies in Phonetics, Phonology and Morphology*, 11(2): 71-82. The Phonology-Morphology Circle of Korea.

Lee, J-P. & Jang, T-Y. 2004. "A comparative study on the production of inter-stress intervals of English speech by English native speakers and Korean speakers." *ICSLP Proceedings*.

Lee, O-H. & Kim, J-M. 2005. "Syllable-timing of Korean English." *Proceedings of the 16th Conference on Speech Sciences*, 295-305. The Korean Association of Speech Sciences.

Lehiste, I. 1977. "Isochrony reconsidered." *Journal of Phonetics*, 5: 253-263.

Low, E. Grabe, L., E. & Nolan, F. 2000. "Quantitative characterization of speech rhythm: syllable-timing in Singapore English." *Language and Speech*, 43(4): 377-401.

Mochizuki-Sudo, M. & Kiritani, S. 1991. "Production and perception of stress-related durational patterns in Japanese learners of English." *Journal of Phonetics*, 19: 231-248.

Pike, K. L. 1945. *The Intonation of American English*, Ann Arbor, MI: Univ. of Michigan Press.

Roach, P. 1982. "On the distinction between 'stress-timed' and 'syllable-timed' languages." In David Crystal (Ed.), *Linguistic Controversies*. London: Edward Arnold. 73-79.

Yun, I. 2002. "Compression effects of number of syllables on Korean vowel." *Speech Sciences*, 9(1): 173-184. The Korean Association of Speech Sciences.

▲ Ok-hwa Lee
   Department of English Language and Literature, Graduate School
   Kangwon National University
   1 Kangwon Daehak-gil, Hyoja 2-dong
   Chuncheon, Gangwon-do 200-701
   Republic of Korea
   Tel: +82-33-250-8105   Fax: +82-33-250-8259
   E-mail: oghwa@msn.com

▲ Jong-mi Kim, Corresponding author
   Department of English Language and Literature
   Kangwon National University
   1 Kangwon Daehak-gil, Hyoja 2-dong
   Chuncheon, Gangwon-do 200-701
   Tel: +82-33-250-8150   Fax: +82-33-250-8259
   E-mail: kimjm@kangwon.ac.kr