Check for updates

The effect of pronunciation teaching on the realization of English rhythm by Korean learners of English*

Wook Kyung Choe**

College of General Education, Dong-A University, Busan, Korea

Abstract

The current study was designed to explore whether taking English pronunciation classes could improve the realization of English rhythm by Korean learners of English. Specifically, this study used various rhythm metrics to examine the extent to which the learners' speech became rhythmically similar to the target language after taking classes that focused on English pronunciation. Sixteen learners who took a 15-week English pronunciation course at a university read an English passage twice (at the beginning and the end of the semester). The rhythm metrics such as Deltas, Varcos, and Pairwise Variability Indices were calculated for the learners' speech, as well as that of 8 native speakers of English. The results demonstrated that the learners' speech was slower, and they put more frequent within-sentence pauses than the native speakers even after the classes. The analyses also indicated that the speech recorded at the beginning of the semester was rhythmically much more different from the target language than at the end of the semester. After the classes, however, the learners' consonantal intervals became much more target-like, while the vocalic intervals were rhythmically even further from those in the target language. Overall, the findings suggested that the pronunciation classes helped the learners to produce English speech that was rhythmically similar to the native speakers.

Keywords: English rhythm, second language acquisition, rhythm metrics, Korean learners of English, teaching pronunciation

1. Introduction

According to rhythm typology, languages have been categorized into three types based on isochrony of speech units: 1) a "stress-timed" language (with regular occurrence of stressed syllables, such as English or Dutch), 2) a "syllable-timed" language (with regular occurrence of syllables, such as French or Spanish), and 3) a "mora-timed" language (with regular occurrence of morae—a mora consists of one consonant and one short vowel—such as Japanese) (Abercrombie, 1967; Ladefoged, 1975; Pike, 1945). Phonologically, syllable structures in syllable-timed languages are relatively simple with rare vowel reduction, whereas stress-timed languages often have

phonotactically complex syllable structures and the vowels in unstressed syllables are frequently reduced (e.g., Dauer, 1983).[1]

As researchers examined the rhythm classes of more languages, it was revealed that not all languages showed regular recurrences of speech units or that there was discordance between the isochrony and the syllable-internal structure of a certain language to be fallen into one rhythmic group. This led some acoustic studies to develop the rhythm metrics (e.g., Dellwo, 2006; Dellwo & Wagner, 2003; Grabe & Low, 2002; Ling et al., 2000; Ramus et al., 1999) and the follow-up acoustic studies have demonstrated that the classification of speech rhythm in different languages are not categorical but rather gradient (see White & Malisz, 2020 for a review).

1.1. Rhythm Metrics

Ramus et al. (1999) first proposed rhythm metrics such as %V (the proportion of vocalic intervals), and ΔC and ΔV (*SD* of consonantal and vocalic intervals, respectively). They argued that the correlation between %V and ΔC well differentiated rhythm classes of languages, in that %V was lower for stress-timed languages than for syllable-timed languages since unstressed vowels in the stress-timed languages were often reduced. Also, ΔC was higher for stress-timed languages because these often allowed consonant clusters.

Dellwo & Wagner (2003) questioned the extent to which speech rates could affect rhythm metrics such as %V and ΔC. Their study indicated that ΔC was more likely to vary with speech rates especially for stress-timed languages like English and German, and then later Dellwo proposed rate-normalized rhythm metrics, Varcos (Dellwo, 2006). Specifically, a Varco is calculated by dividing a delta value (ΔC or ΔV) into the average duration of the intervals.[2] Comparing %V, ΔC, VarcoV, and VarcoC, the study demonstrated that %V and VarcoC were best to differentiate stress-timed from syllable-timed languages.

Both Ramus et al. (1999)'s and Dellwo (2006)'s rhythm metrics focused on the proportions and/or variations of overall consonantal and vocalic intervals within a sentence. However, Grabe & Low (2002) argued that these rhythm metrics could not distinguish the situation where a longer interval (e.g., a full vowel) and a shorter interval (e.g., a reduced vowel) alternated from the situation where longer intervals successively occurred and then shorter ones successively occurred. To indicate the degree of variability in successive consonantal and vocalic intervals, Grabe & Low suggested an acoustic variability index—the raw Pairwise Variability Index (rPVI) and the speech rate normalized Pairwise Variability Index (nPVI). The formula for rPVI and nPVI are presented in (1) and (2) below: m indicating the number of measured intervals and d_k indicating the duration of the kth interval.

$$rPVI = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}| \right] / (m-1) \qquad (1)$$

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| \right] / (m-1) \qquad (2)$$

The analysis of the PVIs for 18 languages proved that rhythm classes were not categorical but gradient (or "weak categorical" as they mentioned): some languages should be considered as prototypical syllable-timed languages (e.g., French and Spanish) or prototypical stress-timed languages (e.g., English and German), but the others could not fall into either category (Grabe & Low, 2002). For PVIs, they argued that the correlation between nPVI-V and rPVI-C well represented the rhythm characteristics of languages. In detail, the nPVI-V for English was 57.2 which was higher than 43.5 for Spanish. The rPVI-C for English was also higher than Spanish (64.1 and 50.4, respectively).

In sum, previous research on rhythm metrics confirmed that English is a prototypical stress-timed language, in that it exhibited relatively lower %V, and higher ΔC, VarcoC, nPVI-V, and rPVI-C. Since unstressed vowels in English frequently reduced, it has lower %V and greater variability of vocalic intervals (i.e., higher nPVI-V). Also, due to complex syllable structures with consonant clusters, all the rhythm metrics related to consonantal intervals (ΔC, VarcoC, and rPVI-C) was higher than those for syllable-timed languages.

From a phonological view, Korean has been regarded as a syllable-timed language, considering 1) that the vowels in Korean are rarely reduced, 2) that Korean has simple syllable structures as (C)V(C), and 3) that no consonant clusters are permitted in Korean (Song, 2006). In terms of rhythm metrics, despite a few conflicting results (e.g., Cho, 2004), most research has suggested that Korean could be categorized as a syllable-timed language, which was not prototypical, though (e.g., Arvaniti, 2009; Arvaniti 2012; Jang, 2009a; Lee et al., 1994; Mok & Lee, 2008). For example, Arvaniti (2012) measured aforementioned rhythm metrics of 6 different languages. The Table 1 showed the mean metric values for English, Korean, and Spanish from Table 9 in Arvaniti (2012).

Table 1. The mean metric values for 3 languages from Arvaniti (2012)

Language	ΔC	%V	rPVI-C	nPVI-V	VarcoC	VarcoV
English	60.0	45.7	68.9	59.9	55.0	54.8
Korean	50.5	49.2	56.7	54.3	54.8	58.3
Spanish	46.6	49.5	53.7	49.1	50.2	53.3

ΔC, *SD* of consonantal intervals; %V, proportion of vocalic intervals; rPVI, raw Pairwise Variability Index; nPVI, normalized Pairwise Variability Index.

As shown in Table 1, ΔC, %V, rPVI-C, and nPVI-V for Korean were quite similar to those for Spanish (a syllable-timed language), but different from those for English (a stress-timed language). Specifically, the %V for Korean was higher, and the nPVI-V was lower than those for English due to rare vowel reductions. In addition, because of the disallowance of consonant clusters in Korean, both ΔC and rPVI-C for Korean were lower than English. These patterns were mostly consistent in Jang (2009a) and Mok & Lee (2008), which also explored Korean speech rhythm using these

---

1 Korean is widely considered as a "syllable-timed" language. Since the current study deals with Korean learners' realization of English rhythm, the characteristics of mora-timed languages will not be discussed in the rest of the study.

2 For better readability, Dellwo (2006) suggested to multiply the obtained value by 100.

metrics. There was one noticeable difference for nPVI-V; namely, the nPVI-V for Korean were more similar to that for English [61 in Jang (2009a) and about 60 in Mok & Lee (2008)] than that for Spanish.

1.2. Rhythm in the Second or Foreign Language Acquisition

As many languages in the world can be phonologically and/or acoustically divided into 2 or 3 rhythm classes, researchers have examined the second or foreign language acquisition of speech rhythm. One area of the research is the L1 (positive or negative) effect on the realization of target language rhythm (e.g., Galaczi et al., 2017; White & Mattys, 2007). It is assumed that learners whose L1 falls into the same rhythm class as the L2 would easily learn or realize the rhythmic characteristics of the L2, whereas learners whose L1 is rhythmically different from the L2 would have difficulties in doing that. For example, White & Mattys investigated the rhythm metrics of L1 and L2 for English, Dutch (stress-timed languages), and Spanish (a syllable-timed language). They showed that the VarcoV values from the English speech by Dutch learners and the Dutch speech by English learners were quite similar to those from the English speech and the Dutch speech by their native speakers. However, the VarcoVs from the English speech by Spanish learners and the Spanish speech by English were in between those from the English speech and the Spanish speech by their native speakers. These suggested that the L1 rhythm patterns were both positively and negatively transferred to the L2 production.

On the other hand, some other studies regarding the L1 effect on the realization of L2 rhythm showed the opposite results. These studies suggested that the L2 acquisition of rhythm for stress-timed languages tended to have the direction from syllable-timing to stress-timing patterns no matter whether the L1 was syllable-timed or not (Li & Post, 2014; Ordin & Polyanskaya, 2015; Zhang & Lee, 2019). For example, Ordin & Polyanskaya examined speech rate and rhythm metrics of English speech by the German learners—whose L1 is stress-timed—and the French learners—whose L1 is syllable-timed—at different English proficiency levels. The results indicated that the advanced learners (either the German or the French) could produce faster speech than the beginners, showing that the average number of syllables per second for the advanced learners was not statistically different from that for the native speakers. In terms of speech rhythm, the Varcos and the nPVIs for both vocalic and consonantal intervals proposed that the beginners produced more syllable-timed pattern no matter whether their L1 rhythm classes were similar to or different from English. However, the effect of L1 rhythm classes was revealed in the advanced learners' speech. That is, only the nPVI-C for the advanced German learners was significantly different from that for the native speakers, while all the values except VarcoC and nPVI-C for the advanced French learners were significantly different from those for the native speakers. This proposed that the advanced learners with rhythmically similar L1 could reach more native-like rhythmic patterns of the target language, but the learners with rhythmically different L1 were less likely to realize the target rhythmic patterns despite their higher level of proficiency.

Several studies also researched the realization of English rhythm by Korean learners using rhythm metrics (e.g., Choe, 2019; Jang, 2009b; Kim, 2008; Kim, 2021; Kim & Chung, 2016; Lee & Kim, 2005; Sa, 2015). For example, Jang (2009b) obtained the rhythm metrics from Korean learners' read speech, and demonstrated that

the Korean learners' %V, VarcoV, nPVI-V, and rPVI-C were higher than the native speakers' values. As far as the consonantal intervals, Choe (2019)'s results showed similar patterns, in that the ΔC and the rPVI-C for the learners were higher than those for the native speakers. However, the variabilities of the vocalic intervals in two studies exhibited inconsistency. Specifically, the nPVI-V for the learners in Choe was significantly lower than that for the native speakers. This result proved that the Korean learners' production of English vowels was relatively less variable, which suggested that the rhythmic characteristics of the learners' L1 negatively transferred to the realization of English rhythm.

These less variable vocalic intervals for the Korean learners of English were also found in Kim (2021)'s study. She investigated the effect of the Korean learners' proficiency level on the realization of English rhythm. In this study, she measured the rhythm metrics relevant only to vocalic intervals from the English speech by three groups—the native speakers, the learners with weak accent, and those with strong accent. The results revealed that the Korean learners with strong accent had more syllable-timing patterns—the highest %V and the lowest VarcoV and nPVI-V. On the contrary, the English speech by the learners with weak accent moved towards more stress-timing patterns which are still significantly different, but similar to the speech by the native speakers.

In sum, previous research on the second language acquisition of speech rhythm has revealed that when learners' L1 was rhythmically different from the target language, the learners' L1 rhythmic characteristics could be negatively influence the realization of the rhythm in the target language. Also, the more advanced the learners' level of proficiency was, the more target-like rhythm (s)he could produce. Furthermore, it seems that rhythm metrics related to vocalic intervals rather than consonantal intervals better represented the extent to which the learners' speech was rhythmically similar to or different from the native speakers' (e.g., Choe, 2019; Kim, 2021; White & Mattys, 2007). Lastly, these general patterns of the second language acquisition of speech rhythm have also been observed in the research on the rhythmic patterns by Korean learners of English.

1.3. Current Study

The current study was designed to explore the realization of English rhythm by Korean learners of English. Especially, the study focused on whether taking classes for English pronunciation could improve the learners' ability to produce more target-language-like rhythmic patterns in their L2 speech. To examine the effect of teaching pronunciation on learning and realizing English rhythm, the current study analyzed rhythm metrics for the learners' speech recorded before and after the classes. The study also compared these with the native speakers' speech to determine the extent to which the learners' English rhythm was similar to or different from the native speakers' rhythm. By doing this, this study could investigate how helpful taking pronunciation classes was for Korean learners to produce English speech with stress-timing patterns.

2. Methods

2.1. Participants

Sixteen Korean learners participated in the current study. All of them were in their 20s [aged from 20 to 28, mean M=21.19], and the undergraduate students with various majors at a university in Busan.

In order to eliminate the influence of L1 dialectal differences, this study recruited the participants who were raised and educated in South Kyungsang province of Korea. The participants' self-reported TOEIC scores was ranged from 550 to 830 M=678.13, SD=99.97. This suggested that their level of English proficiency was intermediate. They had no experience of living in English-speaking countries more than 1 month.

All the Korean participants took a one-semester course about English pronunciation. The class meets twice a week for 15 weeks and each class lasts for 75 minutes. This is an elective course designed to teach undergraduate students how to speak English with more native-like pronunciation and to better understand spoken English so that their listening skills could be improved. The instructor—the author—provided some basic knowledge about English phonetics such as segmental and suprasegmental characteristics of English. The beginning two-thirds of the classes focused more on the segmental perspectives (e.g., how to pronounce English consonants and vowels correctly), while the last one-third of the classes focused more on the suprasegmental perspectives (e.g., syllable, stress, phrasing, and intonation). With the basic information about segmental and suprasegmental aspects of English and useful tips to have better English pronunciation, the students were asked to repeat either the pre-recorded sound files or the instructor's demonstration to apply the acquired knowledge to their actual English pronunciation.

As a control group, 8 native speakers of English also participated in the study. They were all graduate students at a university in the northwestern area of the USA. They were aged in between 26 to 45 (M=34.88). The native speakers were not fluent in any language other than English, and they were raised and educated only in the US. This suggested that they all are the native speakers of American English.

All 24 participants voluntarily participated in this study. The gender of the participants was controlled, and all the participants reported not to have speech and hearing problems.

2.2. Stimuli and Procedure

The reading material was an extract from a TIME for Kids article about polar bears. As the Appendix shows, the extract has three paragraphs, each of which consists of 5 sentences. The lengths of 15 sentences vary from 7 syllables to 30 syllables.[3] The total number of syllables for the extract was 256, and the average number of syllables per sentence was 16.53 (SD=7.00). This material was chosen because articles for kids usually consist of easy vocabulary and expressions. This could lead the Korean participants to easily understand the context, and so to read the texts relatively naturally.

The Korean learners were asked to participate in two recording sessions to get the pre- and post-education speech data. The pre-education session was in the second or third week of the semester, and the post-education was in the fourteenth and the fifteenth week.[4] To minimize the extent to which each participant practice the reading material between the two sessions, the reading material was not used during class meetings but was given to the participants during the experiment only. The settings and the procedures were exactly same for the two sessions.

The experiment for the learner group was conducted in a quiet office. The author asked each participant first to read the material (printed on paper) in mind so that (s)he got familiar with the vocabulary and understand the context. If (s)he could not either know the meaning or the pronunciation of any word, the author informed the participant. When the participant felt ready to read the material, (s)he was asked to read it aloud wearing a head-worn microphone (Shure SM35-XLR). To obtain analyzable and more natural speech data, when the participant put more than two mistake/error/disfluency-driven pauses within a sentence, the experimenter interrupted and asked the participant to re-read the exact sentence. Other than this, the participants could manipulate their own prosodic structures for the read speech (e.g., where to put prosodic boundaries or the types of intonation to use), and read the material at their normal speech rate. The learners' speech was digitally recorded via a Marantz PMD 661 MKIII.

The procedures for the native speakers were exactly same as that for the learners, except a few settings. The native speakers' speech was recorded to a Marantz PMD 660 in a quiet laboratory room in the US, using Shure ULXS4 wireless receiver and lavaliere microphone.

2.3. Acoustic Measurement

A total of 600 sentences (480 sentences for the learners and 120 sentences for the native speakers) were recorded and then analyzed. To calculate different the values for aforementioned rhythm metrics and the speech rate, the digitally recorded speech data was analyzed using Praat software (Boersma & Weenink, 2014). The boundaries for syllables, consonantal and vocalic intervals were identified and labelled by the author with the help of visual (i.e., spectrograms and waveforms) and auditory information. The criteria in the previous research were adopted to determine the boundaries and to identify pauses (Grabe & Low, 2002; Krivokapić, 2007; Lee & Kim, 2005; White & Mattys, 2007). A mistake, speech error, self-correction, hesitation, or filler was labeled as a disfluency. These disfluencies and pauses driven by these disfluencies were then excluded from further analysis. Using Praat script, the numbers of intervals, syllables, and pauses as well as the durations of consonantal and vocalic intervals were measured. Figure 1 shows the sample segmentations and labels.

---

3 The numbers of syllables were counted based on the online dictionary (https://dict.naver.com/). Depending on the way of saying the year 2050 as 'twenty fifty' or 'two thousand and fifty,' the longest sentence—the 4th of paragraph 2—can have 28 syllables. Another difference was on the way of reading −30 as 'minus thirty' or 'negative thirty'. Since many of the participants read them as 'twenty fifty' and 'minus thirty,' the overall number of syllables was calculated in this way. However, the actual measures of the number of syllables followed by the way of each participant's pronunciation.

4 Since the first week of the semester was for the class introduction, and the eighth and the fifteenth weeks were for the class exams, the actual time difference between pre- and post-education sessions was 12 weeks. However, for easy reference, the current study will refer to this as a one-semester or 15-week pronunciation course.
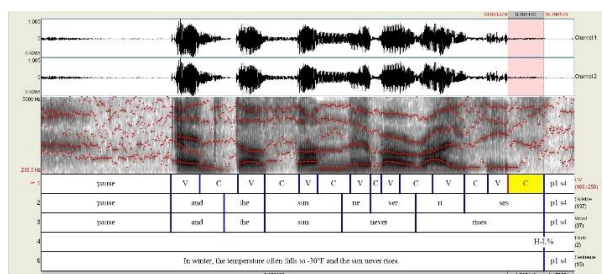
Figure 1. A sample Praat window with segmentations and labels

A total of 8 rhythm metrics were calculated with the durations of consonantal and vocalic intervals; %V, ΔC and ΔV in Ramus et al. (1999); 2 Varcos in Dellwo (2006); and 4 PVIs in Grabe & Low (2002). Additionally, since it is reported that speech rate and the number of pauses were good indicators of the learners' proficiency level (e.g., Huang & Gráf, 2020), the speech rate (hereafter Rate)—dividing the number of syllables by the articulation duration, and the number of pauses within a sentence (hereafter NumP) were calculated.

3. Results

The first set of analyses focused on comparing the learners' pre-education speech with the native speakers' speech. Before investigating the rhythm metrics, the effect of Group (learners vs. native speakers) on Rate and NumP was tested. The one-way analysis of variance (ANOVA) test on Rate revealed that the native speakers produced significantly faster speech (M=5.07, SD=0.75) than the learners (M=3.74, SD=0.68) [$F_{(1, 358)}$=285.21, p<.001]. Since NumP violated homogeneity,[5] a Mann-Whitney U test was conducted. The result revealed that the native speakers put significantly fewer within-sentence pauses (M=0.34, SD=0.59) than the learners (M=1.31, SD=1.38) [U=7,774, p<.001]. These results suggested that the pre-education speech data in the current study showed typical characteristics of learners' speech—slower speech with frequent pausing.

Next, the rhythm metrics for the two groups were analyzed. The effect of Group on %V, Varcos, and nPVI-C was assessed using the one-way ANOVAs. The results indicated that the group effect was significant on %V [$F_{(1, 358)}$=10.62, p=.001], VarcoV [$F_{(1, 358)}$=17.61, p<.001], and nPVI-C [$F_{(1, 358)}$=5.18, p=.023]. The Mann-Whitney U tests for deltas, rPVIs, and nPVI-V yielded significant group differences for all the metrics: ΔC [U=8,111, p<.001], ΔV [U=9,858, p<.001], rPVI-C [U=9,180, p<.001], rPVI-V [U=10,368, p<.001], and nPVI-V [U=21,292, p<.001]. The significant differences are presented in Figure 2.
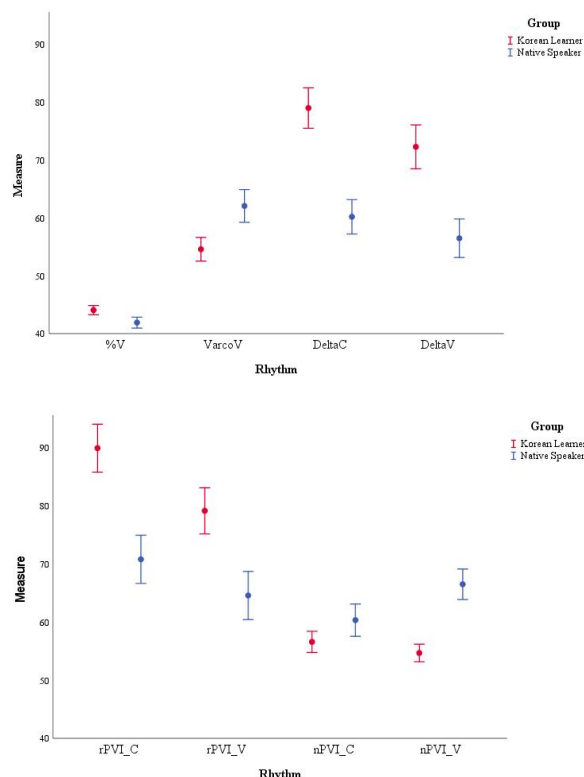

Figure 2. Mean values of 8 rhythm metrics as a function of Group (learners vs. native speakers) with 95% confidence interval error bars. %V, proportion of vocalic intervals; rPVI, raw Pairwise Variability Index; nPVI, normalized Pairwise Variability Index.

Figure 2 demonstrates three findings about the learners' pre-education speech. First, all measured rhythm metrics except VarcoC for the learners were significantly different from those for the native speakers. This proposed that the learners' English speech was rhythmically different from the native speakers' one. The second finding is the higher %V for the learners, which suggested that the proportion of vocalic intervals in the Korean learners' English speech was bigger than that in the native speakers' one.

The last interesting finding is the effect of speech-rate normalization on the rhythm of the learners' speech. As presented in Figure 2, all the non-rate-normalized rhythm metrics (i.e., deltas and rPVIs) represented the patterns against our expectation. Specifically, the higher values for the learners indicated that the learners produced English speech with greater variabilities in both vocalic and consonantal intervals. This could imply that the learners' speech was more "stress-timing" than the native speakers' one. However, all the rate-normalized rhythm metrics showed the opposite directions—lower values for the learners. Especially because the analysis for speech rate yielded a significant difference between the two groups (learners vs. native speakers), it is legitimate to consider the rate-normalized metrics as the evidence to understand the learners' realization of English rhythm. Altogether, the results of VarcoV and nPVIs suggested that the learners' speech showed

---

5 For further analyses, a Mann-Whitney U test instead of a one-way ANOVA was conducted whenever any rhythm metric violated homogeneity with the independent variable, Group (learners vs. native speakers).

syllable-timing patterns, and so they could not produce native-like rhythmic patterns before they took the English pronunciation classes.

The next analyses were conducted to examine the effect of Group for the post-education data. Like the analyses for the pre-education data, the Group effect on Rate and NumP was tested using Mann-Whitney U tests. The results showed the significant differences between two groups; that is, even after taking the classes, the learners' speech was slower (M=3.77, SD=0.56) [U=9,628, p<.001], and had more within-sentence pauses (M=1.32, SD=1.51) [U=8,509.5, p<.001] than the native speakers' one.

The Group effect on rhythm metrics was assessed using a one-way ANOVA (%V, Varcos, and nPVIs) or a Mann-Whitney U test (deltas and rPVIs). The ANOVA results demonstrated the significant Group differences on VarcoV [$F(1, 358)$=18.00, p<.001] and nPVI-V [$F(1, 358)$=87.16, p<.001]. The Mann-Whitney U tests showed that there were significant differences between two groups for ΔC [U=6,468, p<.001], ΔV [U=12,155, p=.016], and rPVI-C [U=7871, p<.001]. These significant differences are shown in Figure 3.

The first noteworthy finding for the learners' post-education data was about the number of rhythm metrics with significant differences. That is, the results revealed that fewer rhythm metrics (i.e., VarcoV, ΔC, ΔV, rPVI-C, and nPVI-V) showed significant Group differences compared with the pre-education data (i.e., all except VarcoC). Based on this, it is possible to argue that the learners' speech became rhythmically more target-like after they took English pronunciation classes for 15 weeks.
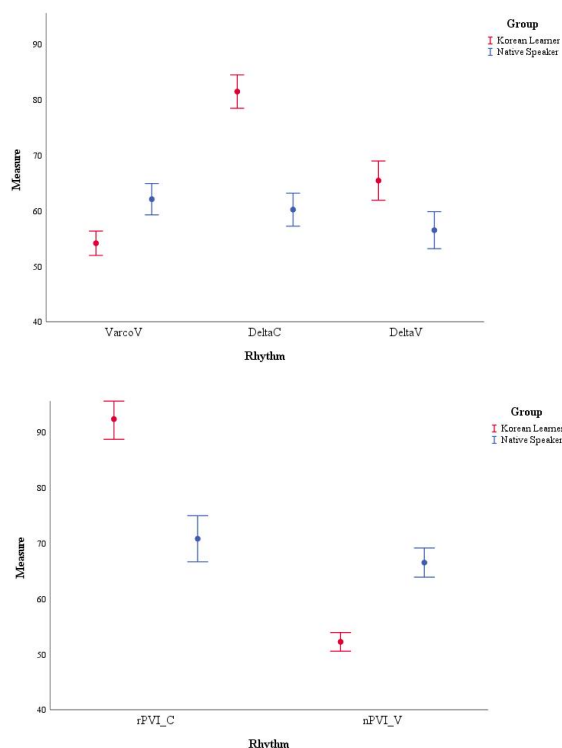
Figure 3. Mean values of 5 rhythm metrics as a function of Group (learners vs. native speakers) with 95% confidence interval error bars. rPVI, raw Pairwise Variability Index; nPVI, normalized Pairwise Variability Index.

In addition, the effect of speech-rate normalization—completely opposite directions for normalized and non-normalized rhythm

metrics—was captured in the current analyses as in the analyses for pre-education data. For the rhythm metrics without normalizing speech rate (i.e., deltas and rPVI-C), the learners' speech showed more vocalic and consonantal variabilities, which are often considered as stress-timing characteristics, than the native speakers' one. On the contrary, the rate-normalized rhythm metrics (i.e., VarcoV and nPVI-V) proposed that the vocalic intervals in the learners' speech were less variable, so the speech had more syllable-timing characteristics compared to the native speakers' speech.

Although some findings could be obtained by comparing the pre-education with the post-education results, this provided us with overall ideas of how much the learners' speech became rhythmically similar to the L2 across all the learners. Since each learner might use different strategies to realize more native-like rhythmic patterns, it is necessary to directly compare the rhythm metrics in pre-education with the post-education speech data for each learner. In this way, we could firmly determine whether taking the English pronunciation class helped the individual learners to successfully produce English speech with more "stress-timing" rhythms.

To explore the extent to which each learners produced rhythmically different English speech after the 15-week course on English pronunciation, the current study performed paired-samples t-tests. Namely, the paired-samples t-tests were conducted to compare the rhythm metrics (including Rate and NumP) in pre-education speech and in post-education speech for each speaker. The results indicated significant differences for %V, ΔV, rPVI-V, nPVI-V, VarcoC, and nPVI-C. The means and standard deviations for each metric and session (pre- vs. post-education) are presented in Table 2.

Table 2. The mean and the standard deviation of 6 rhythm metrics with significant differences between two sessions

Rhythm metrics	Pre-education	Post-education
%V[**]	44.07 (6.26)	41.38 (5.02)
ΔV[**]	72.27 (29.80)	65.41 (27.77)
rPVI-V[**]	79.11 (31.27)	69.63 (29.30)
nPVI-V[*]	54.68 (11.82)	52.22 (13.21)
VarcoC[*]	50.03 (11.46)	51.93 (11.81)
nPVI-C[*]	56.58 (14.32)	59.05 (15.13)

[*]p<.05, [**]p<.001.
%V, proportion of vocalic intervals; ΔV, SD of vocalic intervals; rPVI, raw Pairwise Variability Index; nPVI, normalized Pairwise Variability Index.

With the results of the paired-samples t-tests, several findings regarding the changes in rhythm patterns by the individual learners were noted. First, no significant differences on Rate and NumP were observed between the pre- and post-education speech. This proposed that the individual learners could read English texts neither at a significantly faster rate nor with significantly fewer within-sentence pauses even after taking the classes (actually, the mean NumP for post-education slightly increased: M=1.31 for pre-education vs. M=1.32 for post-education).

As shown in Table 2, the rhythm metrics related to vocalic intervals showed opposed directions to those related to consonantal intervals. That is, with respect to C values, the post-education speech had significantly greater variabilities for consonantal intervals than the pre-education one. This suggested that taking the pronunciation classes enabled the individual learners to produce

more native-like rhythmic patterns for the duration of consonants.

However, the individual differences related to vocalic intervals demonstrated quite surprising results. First, the %V values for pre- and post-education suggested that the learners seemed to have rhythmically more native-like vowel productions, in that the proportions of vocalic intervals in their post-education speech were lower than those in the pre-education speech. On the other hand, the other rhythm metrics for vocalic intervals such as ΔV and both PVI-Vs conflicted with our expectation. Specifically, even after taking the English pronunciation classes, the individual learner produced English speech with less variable vocalic intervals, which is not L2-like (i.e., stress-timing) but rather L1-like (i.e., syllable-timing) rhythmic patterns.

These differences between vocalic and consonantal intervals in the individual learners' pre- and post-education speech were captured in Figure 4. Since nPVIs were the only rhythm metrics with significant differences for both consonantal and vocalic intervals, the distribution of three speech data are presented over the nPVI-V and nPVI-C plane as in Figure 4.
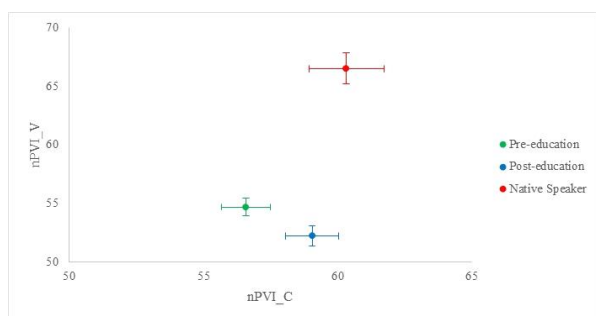
Figure 4. Distribution of the learners' pre-education and post-education speech as well as the native speakers' speech over the nPVI-V and nPVI-C plane. Each bar represents one *SE* around the mean. rPVI, raw Pairwise Variability Index; nPVI, normalized Pairwise Variability Index.

Figure 4 illustrated that when speech rate was normalized, the learners' pre-education speech was significantly different from the native speakers' speech in the rhythmic perspective. After taking the classes about English pronunciation, each learner could realize the English rhythm in significantly different way. However, the direction for vocalic and consonantal intervals were dissimilar to each other. To be specific, after the 15-week course, the individual learners' consonantal intervals were not different from the native speakers' ones, insisting that the learners' consonantal production rhythmically became more stress-timing. On the contrary, the learners' vocalic intervals moved towards the opposite direction after taking the English pronunciation classes.

4. Discussion and Conclusion

The purpose of the current study was to explore the effect of English pronunciation classes on the realization of English rhythm by Korean learners of English. To investigate how much the learners' speech became rhythmically similar to the target language, the study examined the rhythm metrics in the learners' pre-education and post-education English read speech, and then compared them with the native speakers' rhythm metrics. The following four findings were noted in the current study: 1) with

regard to the speech rate and the number of pauses, the learners' speech did not show significant differences even after taking the 15-week pronunciation course; 2) to understand the substance of learning L2 rhythm, the speech-rate-normalized rhythm metrics should be considered; 3) as for the interval-based rhythm metrics, the pre-education speech was rhythmically much more different from the L2 compared to the post-education speech; and 4) the learners seemed to differently manipulate the durations of consonantal intervals from those of vocalic ones after taking the pronunciation classes to read passages with more native-like English pronunciation. These main findings will be discussed regarding the effect of pronunciation classes on the L2 acquisition of speech rhythm. Also, this section will speculate the possible strategies which the learners used to produce more native-like English speech.

Overall, the current study demonstrated that in the rhythmic perspective, the Korean learners seemed to learn how to improve their English pronunciation through the one-semester course on English pronunciation. To better understand in what way the learners' speech became rhythmically more native-like, it is necessary to check detailed class operation. As mentioned earlier in the Methods Section, the beginning two-thirds of the classes (i.e., up to tenth week) focused on how to correctly pronounce English consonants and vowels and the rest 4 weeks were devoted to the suprasegmental concepts such as syllables, word and sentence stresses, phrasing, and intonation. The characteristics related to rhythm were instructed during the classes about syllables (e.g., consonant clusters) and about stress (e.g., reduced vowels). For example, the instructor taught the differences in syllable-internal structures for English and Korean, and then asked the students to practice the accurate way of pronouncing consonant clusters without inserting epenthetic vowels. Also, the instructor provided the students with the information about lexical stresses and the differences between content words and function words. They were also taught that English vowels could be reduced ("have weaker pronunciation") for unstressed syllables. For this class, the students should practice various words and sentences to make difference between stressed and unstressed syllables in their own speech.

Going back to our findings, the current study indicated that there were no significant differences in the speech rate and the number of within-sentence pauses between the pre-education and the post-education speech, but both were significantly different from the native speakers. One possible explanation lies in how the learners were instructed during the classes. More specifically, the instructor advised the students not to speak English at a faster rate with incorrect or inaccurate pronunciation, but to slow down a bit to have more native-like pronunciation (advice based on the previous research on the speech rate and accentedness as in Anderson-Hsieh & Koehler, 1988). Alternatively, since the higher-level prosodic features such as speech rate and prosodic phrasing are rarely taught or since learning these higher-level prosodic features is thought to be more difficult than learning segmental pronunciation (e.g., Trouvain & Gut, 2007), the learners' English proficiency might not get advanced enough to manipulate the higher-level prosodic features such as speech rate and phrasing. In any case, the current study revealed that the 15-week pronunciation course was not sufficient for the learners to read English sentences at a faster speech rate without stopping in the middle of sentences. Also, it proposed that even after taking pronunciation classes, the Korean learners' English speech kept the typical characteristics of learners'

speech such as slower speech rate and frequent phrasing as in Choe (2019), Jang (2009b), and Ordin & Polyanskaya (2015).

Unlike the speech rate and pausing, the Korean learners were able to successfully improve their production of English rhythm by learning and practicing English pronunciation through the classes. The number of the rhythm metrics with significant difference in the pre-education data was 8, which was more than the number (i.e., 5) in the post-education data. This direct comparison tells us that the learners' English speech got rhythmically changed, and then became more similar to the target language after taking the classes. Before reaching the conclusion of the effect of pronunciation classes, however, we need to speculate more details of the extent and the direction to which the learners' speech got changed.

One interesting finding from the current study was that not raw rhythm metrics but the rate-normalized metrics well represented the actual learning patterns and directions of English rhythm by Korean learners. That is, as for the raw rhythm metrics (i.e., deltas and rPVIs), both pre- and post-education speech showed greater variabilities. Considering the stress-timing L2, these values were supposed to be lower for the learners' speech than the target language. In contrast, the rate-normalized rhythm metrics such as Varcos and nPVIs in the current study were lower for the learners than the native speakers, proving that the learners' speech had L1 rhythmic characteristics (smaller variabilities in consonantal and vocalic intervals) in their L2 speech. The current findings related to rate-normalized rhythm metrics were consistent in the previous research on the acquisition of L2 rhythm (e.g., Ordin & Polyanskaya, 2015; White & Mattys, 2007), in that Varcos and nPVIs should be considered for L2 rhythm acquisition since most L2 speech was slower than native speaker's speech.

Although the rate-normalized rhythm metrics in the current study confirmed the effect of pronunciation classes on the learners' ability to manipulate the durations of consonants and vowels, the directions of the consonantal and vocalic metrics were different from each other. That is, as seen in Figure 4 and Table 2 above, the rhythm metrics of consonantal intervals became more target-like (i.e., stress-timing); whereas the vocalic intervals were rhythmically even further apart from the target language. These results could demonstrate that after taking the classes, the learners were able to vary the durations of English consonants, but the durations of English vowels became more similar to each other in their English speech. Revisiting the rhythm characteristic of English and Korean might lead us to possible explanation of these findings.

English is a prototypical stress-timed language which allows the consonant clusters in onset and coda position. In English, not only the durations of the vowels in unstressed syllables were often shortened, but also in most cases, these unstressed vowels changed their qualities into the most relaxed vowel, a schwa. On the contrary, Korean is one of the syllable-timed languages with a simple syllable-internal structure of CVC. Since it is widely believed that Korean does not have lexical stress (see Song, 2006 for a review), the vowels in Korean are produced with relatively equal durations and never changed into different vowels. With these differences between L1 (Korean) and L2 (English), previous research on the prosodic acquisition by Korean learners of English has reported that they had difficulties in successfully reducing unstressed vowels (cf. Jung & Rhee, 2018; e.g., Kwon, 2007), and that they often inserted vowels to avoid consonant clusters (e.g., Hong et al., 2010). The higher %V in pre-education speech, which then became similar to

the native speakers' %V, suggested that the learners' L1 was negatively transferred into their L2 speech (i.e., no vowel reductions or vowel epenthesis) before taking the classes.

However, as learning the different rhythmic characteristics and practicing how to realize English rhythmic patterns, one (i.e., consonantal intervals) was quite successfully acquired, but the other (i.e., vocalic intervals) was not yet. Specifically, the lower %V and the higher VarcoC and nPVI-C—none of which were significantly different from the native speakers' values—in the post-education data implied that the learners were able to vary the durations of consonantal intervals without inserting epenthetic vowels in between consonant clusters. In contrast, the lower nPVI-V value for the post-education data proposed that the learners tried to differentiate the durations of vocalic intervals (i.e., significant paired-samples t-test value for nPVI-V), but their manipulation of vocalic intervals were not sufficient enough to be realized as native-like variabilities in producing vocalic intervals.

Then, what could make the difference in realizing English rhythms related to consonantal versus vocalic intervals? One possibility could lie in the intrinsic differences of consonantal and vocalic intervals. The variabilities in the durations of consonantal intervals can be somewhat structural, then so categorical (i.e., how many consonants are in either onset or coda position); while the vocalic variabilities depend more on gradient (i.e., how long or short a vowel is). That is, the duration of the consonantal interval in CCCV (as in 'spray') should be the longer than that in CCV (as in 'pray'), which is longer than that in CV (as in 'ray'). However, as there is limitation on lengthening or shortening the duration of a vowel within a syllable, varying the vowel durations is more fine-tuning procedure. Also, the duration of the vocalic intervals can be influenced by other prosodic characteristics such as stress or sentence focus more easily than that of the consonantal intervals. For example, the second [ɪ] in 'religion' should be longer than the first [ɪ] since the word has a penultimate stress. As the learners with the syllable-timed L1, learning and realizing more structural and categorical manipulation such as variabilities in consonantal intervals can be easier than fine-tuning the durations of vocalic intervals. Of course, this speculation must be assessed with more controlled experiments in the future studies.

In conclusion, the current study aimed to examine whether teaching English pronunciation could help the Korean learners to realize English rhythm in a native-like manner. Analyzing the rhythm metrics for the native speakers' speech and the learners' speech at the beginning and the end of the semester proposed that the learners were able to apply their acquired knowledge about English rhythm to their own production, and so to read English passages with more native-like rhythms.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, Scotland: Edinburgh University Press.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*(4), 561-613.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66*(1-2), 46-63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351-373.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer (version 5.3.66) [Computer program]. Retrieved from http://www.praat.org/

Cho, M. H. (2004). Rhythm typology of Korean speech. *Cognitive Processing*, *5*(4), 249-253.

Choe, W. K. (2019). The realization of English rhythm by Busan Korean speakers. *Phonetics and Speech Sciences*, *11*(4), 81-87.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51-62.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for delta C. In P. Karnowski, & I. Szigeti (Eds.), *Language and language processing: Proceedings of the 38th Linguistic Colloquium, Piliscsaba 2003* (pp. 231-242). Frankfurt, Germany: Peter Lang.

Dellwo, V., & Wagner, P. (2003, August). Relations between language rhythm and speech rate. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 471-474). Barcelona, Spain.

Galaczi, E., Post, B., Li, A., Barker, F., & Schmidt, E. (2017). Assessing second language pronunciation: Distinguishing features of rhythm in learner speech at different proficiency levels. In T. Isaacs, & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 157-182). Bristol, UK: Multilingual Matters.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*:515-546.

Hong, H., Kim, J., & Chung, M. (2010, September). Effects of Korean learners' consonant cluster reduction strategies on English speech recognition performance. *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (pp. 1858-1861). Makuhari, Japan.

Huang, L. F., & Gráf, T. (2020). Speech rate and pausing in English: Comparing learners at different levels of proficiency with native speakers. *Taiwan Journal of TESOL*, *17*(1), 57-86.

Jang, T. Y. (2009a). Rhythm metrics of spoken Korean. *Language and Linguistics*, *46*, 169-185.

Jang, T. Y. (2009b, November). Automatic assessment of non-native prosody using rhythm metrics: Focusing on Korean speakers' English pronunciation. *Proceedings of the 2nd International Conference on East Asian Linguistics*. Vancouver, BC.

Jung, Y. J., & Rhee, S. C. (2018). Acoustic analysis of English lexical stress produced by Korean, Japanese and Taiwanese-Chinese speakers. *Phonetics and Speech Sciences*, *10*(1), 15-22.

Kim, H., & Chung, H. (2016). An analysis of the rhythm and reduced vowels of Korean adult learners of English. *Studies in Foreign Language Education*, *30*(3), 109-132.

Kim, S. (2021). Foreign accentedness rating and rhythm in L2 English. *The Mirae Journal of English Language and Literature*, *26*(4), 101-120.

Kim, S. A. (2008). Rethinking the dichotomy between syllable-timed vs. stress-timed languages with particular reference to Korean L1 speakers' English. *The Journal of Studies in Language*, *24*(3), 473-493.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, *35*(2), 162-179.

Kwon, H. (2007). A phonetic analysis of advanced Korean EFL learners' English reduced vowels. *Korean Journal of English Language and Literature*, *7*(4), 547-564.

Ladefoged, P. (1975). *A course in phonetics*. New York, NY: Harcourt Brace Jovanovich.

Lee, H. B., Jin, N., Seong, C., Jung, I., & Lee, S. (1994, September). An experimental phonetic study of speech rhythm in standard Korean. *Proceedings of the 3rd International Conference on Spoken Language Processing* (pp. 1091-1094). Yokohama, Japan.

Lee, O. H., & Kim, J. M. (2005). Syllable-timing interferes with Korean learners' speech of stress-timed English. *Speech Sciences*, *12*(4), 95-112.

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition*, *36*(2), 223-255.

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

Mok, P., & Lee, S. I. (2008, July). Korean speech rhythm using rhythmic measures. *Proceedings of the 18th International Congress of Linguists (CIL18)*. Seoul, Korea.

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533-544.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

Sa, J. J. (2015). Rhythm metrics approach on English speech by Korean learners of English. *Journal of British and America Studies*, *35*, 235-254.

Song, J. J. (2006). *The Korean language: Structure, use and context*. London, UK: Routledge.

Trouvain, J., & Gut, U. (2007). *Non-native prosody: Phonetic description and teaching practice*. Berlin, Germany: Walter de Gruyter.

White, L., & Malisz, Z. (2020). Speech rhythm and timing. In C. Gussenhoven, & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 166-179). New York, NY: Oxford University Press.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501-522.

Zhang, J., & Lee, S. (2019). Acquisition of English speech rhythm by Chinese learners of English at different English proficiency levels. *Phonetics and Speech Sciences*, *11*(4), 71-79.

• **Wook Kyung Choe,** Corresponding author
Assistant Professor, College of General Education
Dong-A University
37 Nakdong-Daero 550beon-gil, Saha-gu, Busan 49315, Korea
Tel: +82-51-200-1022
Email: regina1004@dau.ac.kr
Fields of interest: Phonetics, Phonology, L2 acquisition

Appendix

The top of the world is a wintry wonderland. Icebergs float in the cold Arctic Ocean. In winter, the temperature often falls to −30°F and the sun never rises. The ocean is surrounded by frozen ground. There are few people or trees, but to polar bears, the Arctic is home.

Polar bears have thick fur, huge paws and other features that make them well prepared for life in their harsh environment. In fact, they need the Arctic sea ice for survival. But climate change is causing larger and larger areas of summer sea ice to melt. Experts say that if warming patterns continue, the Arctic could be free of summer sea ice by 2050. That may cause two-thirds of the world's 20,000 polar bears to be gone by then too.

Polar bears can't survive for long on land. Seals are their main source of food. The bears hunt for seals in openings in the sea ice. Polar bears need the ice to get to their prey. In summer, the polar bears that live on land eat very little and wait for the sea ice to return.