

A pilot study on the rhythmic properties of English produced by Japanese learners before and after a five-month study abroad period.

Teruaki Tsushima

ABSTRACT

The present study was designed as a pilot study to investigate whether and how the rhythmic properties of sentences produced by adult Japanese learners of English (JS, N=7) changed before and after a 5-month study abroad period. The participants read and recited a set of 15 English sentences which had different rhythmic patterns. Some of the sentences had an alternation of stressed and unstressed syllables in content and function words, and others included a succession of stressed syllables in content words (i.e., a stress clash). The rhythm index called “normalized pairwise variability index of vocalic intervals (nPVI-V)” was used to analyze the changes in the rhythmic properties of the sentences. The results showed that nPVI-V increased to approximate that of the native speakers of English (NS, N=8), although such improvement was observed only in JS with a higher proficiency level. NS also evaluated the degree of accentedness in terms of the rhythmic properties among JS’s productions. It was found that, commensurate with the results on nPVI-V, the evaluation scores increased to varying degrees before and after the study abroad period. The two measures, however, were found to be significantly, but only weakly correlated ($r=.386$). Another aim of the study was to evaluate the validity of nPVI-V as a measure of the rhythmic properties in L2 learners. The results suggested that nPVI may have a number of limitations, including difficulty with estimating the native level of nPVI-V, and with controlling for the prosodic, phonological, and phonemic properties in the speech materials used for a comparison between JS and NS. Teaching implications were discussed.

key words; speech rhythm, duration, stress, reduced vowels, rhythm indices, L2 learning

A pilot study on the rhythmic properties of English produced by Japanese learners before and...
study abroad

1. Introduction

Previous research (e.g., Flege, Munro, & MacKay, 1995) has amply demonstrated that the second language (L2) learners' phonetic (perceptual and productive) abilities are constrained by the phonetic and phonological properties of the first language (L1). The influence of L1 on L2 has been found not just at a segmental level (e.g., phonemes), but also at a prosodic level (e.g., rhythm and intonation). For example, many instructors teaching English in Japan would agree that the learners speak English with the rhythmic patterns that have characteristics of the Japanese rhythm. Specifically, the length of a syllable is more or less equal, with the pitch range being relatively small, which may give an impression of a staccato rhythm. The issue of great interest and importance for L2 researchers and foreign language instructors is how learners learning an L2 with a different rhythm acquire the ability to produce the rhythmic properties of the target language. The present study investigated whether and how the rhythmic properties of English spoken by adult Japanese L2 learners of English improved through exposure to native English during a study abroad period (SA, henceforth).

English and Japanese have traditionally been categorized as languages which represent different rhythm classes (Abercrombie, 1967). English is classified into a stress-timed language in which the length of the adjacent stressed syllables (i.e., foot) is relatively isochronous. On the other hand, Japanese is a mora-timed language in which the length of a mora is relatively equal. Extensive research which examined the duration of those units (i.e., foot, syllable, and mora), however, did not find evidence that such isochrony existed in the speech signal (Dauer, 1983; Roach, 1982). The following research sought to find acoustic differences between different rhythm classes by using a number of rhythm indices. Ramus, Nespors & Mehler (1999) proposed measuring the degree of durational variability and proportions of speech segments (e.g., vowels, consonants, syllables) over a set of utterances. The rhythm indices include %V (the proportion of vocalic intervals), ΔV (the standard deviation of vocalic intervals), and ΔC (the standard deviation of consonantal intervals). The Japanese, which has a mora-timed rhythm, is characterized by a relatively large %V and small ΔC with less complex syllable structures. The following research proposed a set of rhythm indices which measure the degree of durational variability in successive speech segments (e.g., vowels, consonants,

and syllables). For example, a normalized pairwise variability of vocalic intervals (nPVI-V), proposed by Grabe and Low (2002), is calculated based on a sum of the absolute difference in duration between each pair of successive segments divided by the mean duration of each pair. The study found that stress-timed languages (e.g., English, German) showed a higher value of nPVI-V than syllable-timed languages (e.g., French, Spanish) and mora-timed languages (e.g., Japanese). The difference in nPVI-V is primarily due to the fact that a stress-timed language generally has both full vowels and shortened/spectrally reduced vowels, which make the durational contrast between the adjacent vowels higher.

Previous research has attempted to investigate whether the rhythm indices can be used to describe the process of learning the rhythmic properties of the target language in L2 acquisition. The available evidence to date is equivocal as to the validity of the rhythm indices and their applicability to research in L2 acquisition. Some studies have found that the rhythm indices become more approximate to those of native speakers in the more advanced than the less advanced learners (Gut, 2009; Li & Post, 2014; Ordin & Polyanskaya, 2014; White & Mattys, 2007). For example, Li & Post (2014) examined the development of English speech rhythm among L1 Mandarin learners and L1 German learners as compared with native English speakers. It was found that VarcoV (the speech-rate-normalized variation coefficient of vocalic intervals) and nPVI-V were significantly higher in the more proficient group and the native control group than the less proficient group in both languages. On the other hand, some studies failed to find significant differences in the values of certain rhythmic indices between a group of more experienced and less experienced L2 learners (Dellwo, Diez, & Gavaldà, 2009; Guilbault, 2002). In addition, Gut (2009) found no significant effect of 6-month pronunciation training or a 9-month study abroad program on the changes of the syllable ratio (the average ratio of adjacent syllable pairs of stressed and unstressed syllables). Based on these equivocal results, some researchers have criticized the idea of using the rhythm indices in the study of L2 rhythm acquisition (Barry, Andreeva, & Koreman, 2009; Gut, 2012; Li & Post, 2014; Turk & Shattuck-Hufnagel, 2013).

It has been pointed out that these results might be due to a number of factors, including speaking style (e.g., reading/spontaneous speech), speech materials (e.g., the degree of syllable complexity in the test materials or produced speech), speaking rate, and segmentation procedure (Gut, 2009, pp. 86-87). In addition, the rhythmic properties of L1 and the target language may influence the degree of changes in the values of the

A pilot study on the rhythmic properties of English produced by Japanese learners before and... rhythm indices in question. Finally, the proficiency levels of L2 learners should also play an important role.

The present study attempts to extend the previous research by providing further data on L2 learning of rhythmic properties and on the validity of using the rhythm indices for L2 research. It was intended to be an exploratory study which would provide pilot data for future research with a larger sample size. The rationale of the study was the following. First, it examined the changes in nPVI-V before and after a 5-month study abroad program using a within-subject design. The participants were adult Japanese learners of varying proficiency levels, learning English as a foreign language at college. Among the studies which have investigated L2 learning of rhythm properties, only a few studies have used a longitudinal design (Dellwo et al., 2009; Guilbault, 2002; Gut, 2009). Given possible individual variability in the process of learning rhythmic properties, it is important to make individual-based observations over different points in development. In addition, the study attempted to observe whether and how the initial proficiency level of the learners affected the degree of learning during the study abroad program.

Second, the study examined data on native speakers' evaluation of L2 rhythmic properties. To the best of the author's knowledge, little data is available as to how the acoustic data on the rhythmic properties of L2 speech are related to native speakers' evaluation of those properties. One of the problems of the data based solely on acoustic analyses of speech is that even if there are significant acoustic differences between two sets of data, it does not guarantee that native listeners are able to recognize the differences in listening. If both sets of data are positively correlated, it might increase the validity of the results based on the rhythm index.

Third, following Mori, Hori, & Erickson (2014), the test sentences had two different rhythmic patterns (to be described in more details in Method section). Specifically, one set of sentences had an alternation of a stressed syllable of a content word, and unstressed syllables of a content word or function words, which created an alternating rhythm of strong and weak syllables in a sentence. Another set of sentences included a succession of stressed syllables in content words (i.e., a stress clash). In the stress clash situation, the speakers were expected to have difficulty following the alternating rhythm of strong and weak syllables. One of the purposes of the present study was to examine whether and how the rhythmic patterns of the test sentences influenced the observed rhythmic index. Specifically, it was of interest to examine whether there were differences in the rhythm index between the two types of rhythmic patterns among Japanese speakers, as

compared with native speakers.

Finally, the present study used nPVI-V as a rhythm index for analyzing the data. First, the rhythmic difference between English and Japanese is expected to be reflected in the difference in the values of nPVI-V. One of the important prosodic differences which give rise to two distinct rhythms is how stressed syllables are acoustically realized. In English, a stressed syllable is realized not just by higher f_0 and larger intensity, but also longer duration, while in Japanese, lexical accent is marked primarily by higher f_0 (Beckman, 1986). In addition, unstressed syllables in English are shorter in duration, and the vowel is centralized (i.e., schwa), while there's no unstressed syllables in Japanese. Given these differences, production of English syllables by Japanese learners of English is influenced by the prosodic characteristics of Japanese. For example, it was found that the English syllables produced by Japanese speakers are relatively equal in duration as compared with native speakers English (Bond & Fokes, 1985; Mochizuki-Sudo & Kiritani, 1991). In addition, the duration of function words produced by Japanese speakers is longer than that of native speakers of English (Aoyama & Guion, 2007). Moreover, previous studies have found that Japanese learners of English have a great deal of difficulty controlling the relative duration of English stressed and unstressed syllables (Mori, Hori, & Erickson, 2014; Tsushima, 2015). Therefore, when the English sentence consists of stressed and unstressed syllables with content and function words, the relative duration of adjacent syllables or vowels (which nPVI-V measures) was expected to be a good measure of how well the Japanese learners of English produce the rhythmic properties of English.

Second, nPVI-V has successfully discriminated between less proficient and more proficient learners of L2 in previous studies (e.g., Li & Post, 2014; Tsushima, 2016). In a recent pilot study, Tsushima (2016) used nPVI-V, along with VarcoV, to examine how the rhythm indices changed during an individual-based, long-term pronunciation training (i.e., 11 months) among three adult Japanese learners of English. The sentence analyzed (i.e., "Age is an important factor in learning to pronounce.") was part of a passage in a diagnostic test, and was read by the participants. The results showed that, in all the participants, nPVI-V were much lower than that of native speakers in the initial test, while they increased to approximate those of the native toward the end of the training period. Further analyses showed that the increase in nPVI-V was mostly due to an increase in duration of stressed syllable of content words (e.g., "port" in "important"), and to decrease in duration of unstressed syllables in function words (e.g., "an", "in"). It was

A pilot study on the rhythmic properties of English produced by Japanese learners before and... expected that, using similar speech materials and the reading task, analogous developmental changes, if any, would be observed in the values of nPVI-V before and after SA.

Specific research questions asked in the present study were the following.

- 1) Did the rhythm index shown by Japanese learners of English significantly differ from that of native speakers of English in the initial test? Did they differ according to the rhythmic patterns of the sentence?
- 2) Did the rhythm index shown by Japanese learners of English change to approximate that of native speakers of English before and after the study abroad period and as a function of the rhythmic patterns of the sentence? Did the degree of improvement differ according to the proficiency levels of the participants?
- 3) Did the native speakers' evaluation of the rhythmic properties of the learners' productions change before and after the study abroad period and as a function of the rhythmic patterns of the sentence?

2. Method

2-1. Participants

Participants were seven Japanese students (four females, three males; JS, henceforth) in a sophomore year of a private university in Tokyo at the time of the study. They belonged to a program where students go to study abroad in Sydney, Australia, for about five months in the latter half of the sophomore year. The proficiency levels of their English could be categorized as pre-intermediate to intermediate based on the TOEIC scores and the author's observations. To prepare for the study abroad, they took a 90-minute English conversation class five days a week in the first half of the sophomore year. During their stay, they took English classes which focused on all four English skills five days a week. They stayed with their host families where they had a plenty of opportunities to communicate in English.

Eight native speakers of English also participated in the study as native speaker controls and also evaluators of JS's productions. All of them were male, and were in their 20's to 50's. Five of them acquired English as the most predominantly used language in USA, two in New Zealand, one in Canada, and one in England. Except for one person, all of them were English instructors at a private company, who had had extensive experiences teaching English to Japanese learners. Five of them (three from USA, one

from England, and one from New Zealand) served as evaluators.

2-2. Target Sentences

As briefly described in Introduction, sentences with two types of rhythmic patterns were created. For the first type, a sentence had an alternation of a stressed syllable in a content word and unstressed/reduced syllables in a content word and/or function words. Sentences of this type were further divided into two types. In one type (AS-N, henceforth), the sentence began with a proper noun (i.e., the name of a person (e.g., Bob)), while in another (AS-I, henceforth) the sentence began with a pronoun, "I". In the former, the first syllable (i.e., the proper name) is normally stressed in NS's productions as it is interpreted as carrying new information in the sentence. In the latter, whether "I" is stressed or not depends on the prosodic structure of the particular sentence. For example, in the first sentence of this type (AS-I: #1¹⁾), "I" is normally stressed to create an alternation of a strong and weak syllable, while in AS-I: #2 and AS-I#3, it is not normally stressed.

AS-N:

- (1) Jack has a lot of donuts in his basket now.
- (2) Bob is starting his job in Japan this year.
- (3) Bob is starting his dentist's office this year.
- (4) Dan has to phone the police in the town now.
- (5) Jack has a lot of hotels in the city now.
- (6) Dan has to call the captain at his house now.

AS-I:

- (1) I took five exams for the final today.
- (2) I found the jacket in the closet today.
- (3) I taught him how to fix the machine just now.

In the second type, a sentence had a succession of stressed syllables of content words, creating so-called a stress clash (called SC sentences, henceforth). Here is a list of SC sentences. The italicized parts indicate where the stress clash occurs.

A pilot study on the rhythmic properties of English produced by Japanese learners before and...

SC:

- (1) Bob starts his *part-time job* at college this year.
- (2) I taught him how to *cook nice duck legs* just now.
- (3) Dan has to *buy guide maps* for his students now.
- (4) Jack has a lot of *pet cats* in his house now.

In the reading list used for data acquisition, the order of the sentences was randomized such that the sentence of the same rhythm type would not occur consecutively. It should also be noted that words starting with an approximant (except for /l/) were avoided because segmentation of the prevocalic portion was known to be very difficult to perform.

2-3. Data Acquisition

In the reading task, JS were asked to read fifteen sentences in a fixed order across the participants. They were shown a list of the sentences and were allowed to practice reading them until they felt comfortable doing so. Then, they were recorded producing the sentences. When they made a mistake, they were allowed to read it again. After a few minutes' break, the recitation task began. In the task, the participants saw a sentence on a PC monitor, and were asked to memorize it. When they clicked on the monitor, the sentence disappeared, and then they were recorded reciting the sentence. When they made a mistake or were not able to complete the sentence, they were allowed to get the sentence back on the monitor and memorize it again. NS read the list in a fixed order at two speaking rates. First, they were asked to read them at a normal rate. After a short break, they were asked to read them at a slower rate in such a way that they would read them for beginning learners of English. When they made a mistake, they were allowed to read the sentence once again.

The recording session was followed by the evaluation session after a short break. NS were asked to evaluate the rhythmic properties of the sentences produced by JS. The five sentences in AS-N were subjected to evaluation, with 70 productions (five target sentences x seven participant x two data sets (before and after SA)) evaluated per evaluator. Only the limited number of sentences were chosen for evaluation because the number of times for the evaluation appeared appropriate to prevent the evaluators from losing concentration due to fatigue. The sentences in AS-N were chosen because relatively clear differences in nPVI-V were found between JS and NA, as well as before and after

SA in JS. Within each target sentence, 14 sentences were pseudo-randomized on the condition that 1) the utterance of the same speaker was not presented consecutively, 2) the utterances of the same data set (i.e., before and after SA) were not presented more than two times in succession. The 14 sentences were normalized in terms of average intensity. NS were instructed to evaluate a degree of foreign accent by paying attention to the timing, pitch and loudness. The latter two were included in the instruction as it was considered very difficult to separate the three properties in evaluating the language rhythm. An evaluation score was given on a scale of six (i.e., “1”=“the sentence is strongly accented.”; “7”=“the sentence has no discernible foreign accent”). They listened to one sentence twice, but could listen more if requested. They were also asked to disregard the adverbial phrase at the end of each sentence (e.g., “just now”, “this year”). After the evaluation, a brief post-hoc interview was held. The evaluators were asked about the criteria they primarily used for evaluation. They also reported about any notable characteristics of a particular speaker’s speeches.

The recording sessions were held in a quiet but non-sound-proofed research room at the author’s university or at the private language school. Their utterances were recorded at a resolution of 16 bits with a sampling rate of 44.1 Hz by a PCM recorder with a high-quality microphone placed approximately 20 cm from the mouth of a speaker. The recorded sounds were low-pass filtered at 10,000 Hz, normalized and analyzed using sound analysis software, *Praat* (Boersma & Weenink, 2014).

2-4. Analyses Procedure

The recorded speech sound was segmented through visual inspection of wave forms and wideband spectrograms, following standard criteria (e.g., Payne, Post, Astruc, Prieto, & Vanrell, 2012; Peterson & Lehiste, 1960). First, a boundary of a vocalic and consonantal portion was placed at the point of zero crossing at the start or end of pitch periods. A glottalized portion at the beginning of a vowel, if any, was excluded from the vocalic portion. Second, the postvocalic /ɹ/ and /w/ were included in the vocalic portion (e.g., “(a)r” in “start”). Third, the boundary of the vocalic portion following /l/ was marked at the release of constriction (e.g., “leg”), which is normally indicated by a sudden rise in the first formant.

2-5. Rhythm indices

For computing nPVI-V, the absolute difference in duration for all the pairs of adjacent

A pilot study on the rhythmic properties of English produced by Japanese learners before and... vocal intervals was calculated, which was divided by the mean duration of both vocalic intervals (i.e., to control for the speech rate). The values for all the pairs were summed and divided by the number of pairs, and was multiplied by 100. It should be noted that an adverb or adverbial phrase at the end of each sentence (e.g., “now”, “just now”, “this year”, “today”) were excluded for calculation of nPVI-V to eliminate the effects of potential sentence final lengthening.

3. Results

3-1. NPVI-V as a function of the rhythm types among JS and NS

First of all, it was found the average sentence duration (the sum of the duration of consonantal and vocalic portions excluding the adverb or adverbial phrase at the end of the sentence) among JS was not substantially different between the reading task (2.38 sec.) and the recitation task (2.43 sec.). This result was somewhat surprising as the recitation task was expected to take longer than the reading task as the participants had to recall the sentences from the memory. It might be because the recitation task always followed the reading task, and JS were able to practice as much as they wanted to before the recitation task began. In addition, nPVI-V were not substantially different between the reading task (45.7) and the recitation task (46.0) either. Therefore, it was decided to use the mean of nPVI-V over the two tasks in the following analyses.

Table 1 shows nPVI-V as a function of the rhythm type and the sentence in JS and NS. First of all, nPVI-V differed greatly across the target sentences in NS, ranging from 47.1 (i.e., SC: #5) to 82.0 (i.e., AS-N#2). It should also be noted that there was a great deal of variability in nPVI-V across the native speakers within a particular sentence. For example, nPVI-V in AS-N: #2 ($M=82.1$) ranged from 53.0 to 98.1. When averaged across all the sentences, however, the range of nPVI-V was much smaller (i.e., from 52.3 to 66.7 with a mean of 60.8). Thus, it appears necessary to use an adequate number of target sentences to estimate the native level of nPVI-V.

Regarding the effects of the rhythm types, nPVI-V averaged across the sentences in NS was lower in SC (53.0) than AS-N (65.8) and AS-I (63.2), as expected. This may probably be due to the relatively lower pairwise variability of vowels in a succession of content words where the stress clash occurred. The same was found in JS with SC (39.8) being lower than AS-N (46.6) and AS-I (52.3). An ANOVA was run with Speaker (JS, NS) and Rhythm Type (AS-N, AS-I, SC) as between-subject factors²⁾. The effect of

Table 1. The normalized pairwise variability index for vocalic intervals (nPVI-V) averaged across JS and NS as a function of the rhythm type, the target sentence, and the task type (i.e., the reading and recitation task only for JS). JS=Japanese speakers; NS=Native speakers of English; Dif=Difference between JS and NS; AS-N=Alternation of stress with a name as a subject; AS-I=Alternation of stress with “T” as a subject; SC=Stress clash.

Type	Speaker	JS		NS		Dif		
		Sentence#	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>p</i>
AS-N		1	39.1	14.2	58.2	16.2	19.1	0.001
		2	48.1	16.0	82.0	7.2	33.9	0.000
		3	31.0	10.3	48.4	14.0	17.4	0.003
		4	62.3	14.1	81.2	10.5	18.9	0.001
		5	49.6	6.5	55.1	14.3	5.5	>0.10
		6	49.4	10.9	70.1	6.3	20.7	0.001
	Subtotal		46.6	14.1	65.8	17.3	19.2	
AS-I		1	56.2	18.0	72.2	9.4	16.0	0.000
		2	50.4	13.9	65.9	15.2	15.5	0.008
		3	48.3	13.5	60.8	13.1	12.4	0.034
		4	54.2	13.8	54.0	8.2	-0.2	>0.10
	Subtotal		52.3	13.4	63.2	13.1	10.9	
SC		1	37.2	13.2	53.3	7.4	16.1	0.006
		2	36.6	13.8	55.2	11.3	18.6	0.002
		3	45.1	10.9	53.0	5.9	7.9	>0.10
		4	39.4	9.3	56.5	11.5	17.2	0.003
		5	40.6	9.5	47.1	8.4	6.4	>0.10
	Subtotal		39.8	9.8	53.0	9.3	13.3	
	Total		45.8	13.4	60.8	15.0	14.9	

Rhythm Type was significant ($F(2, 324)=21.9, p=.000$), as well as that of Speaker ($F(1, 324)=78.1, p=.000$). The interaction of Rhythm Type and Speaker was not significant. In both groups, nPVI-V for SC was significantly lower than that of AS-N and AS-I ($p<.01$ with Bonferroni correction). It should be noted, however, that nPVI-V differed substantially across the target sentences in NS. It ranged from 55.1 to 82.1 in AS-N. The results indicated that, among native speakers of English, the degree of nPVI-V depends largely on the particular prosodic, phonological and lexical properties of an individual sentence. In AS-N: #2 and AS-N: #3 in NS, for example, the first part of the two sentences were identical (i.e., “Bob is starting”) and the following part, “job in Japan” and

A pilot study on the rhythmic properties of English produced by Japanese learners before and...

“dentist’s office” was different. Inspection of the data of NS found that the differences in duration of the stressed syllable in the two sentences were the major cause of the difference in nPVI-V. The vowel duration of the stressed syllable, “job” and “(Ja)pan”³ were 181 ms and 154 ms in AS-N: #2, while that of “den” and “of(fice)” were 90 ms and 114 ms. As a result, the mean “local nPVI-V” (i.e., the absolute difference in duration between the vowels of the pair divided by the mean duration of the pair) was higher in the former (e.g., “job in”=1.04; “Ja-pan”=1.26) than the latter (e.g., “den-tist’s”=.28; “of-fice”=.38).

Next, nPVI-V in each target sentence was compared between JS and NS. An ANOVA was conducted with Speaker (JS, NS) and Sentence (15 target sentences) as between-subject factors. It was found that the effect of Speaker was highly significant, $F(1, 324) = 78.1, p = .000$, showing that nPVI-V averaged across the participants and sentences was significantly higher in NS than JS. However, the mean magnitude of the difference differed greatly across the sentences. Simple-effects analyses of Speaker at each sentence were run with a significance level at $\alpha = .01^4$. The results are shown in Table 1. In AS-N, nPVI-V was significantly higher in NS than JS in five out of six sentences, with the difference ranging from 5.5 to 33.9. The largest difference was observed in AS-N: #2 (“Bob is starting his job in Japan this year.”). In this sentence, the mean local nPVI-V was higher in six of eight pairs in NS than JS. For example, the local nPVI-V for the pair, “Bob is” was .23 in JS and .91 in NS. The mean normalized duration (i.e., the duration of a vowel normalized for the total duration of the sentence) for “Bob” in NS was 171 ms, while that of JS was 116 ms. On the other hand, the mean normalized duration of “is” was 64 ms in NS, while that of JS was 101 ms. As this example shows, the higher nPVI-V in NS were largely due to a combination of longer durations of stressed syllables in vowels of content words, and shorter durations of unstressed syllables in function words. This applied to the other pairs such as “start-ing”, “his job”, “job in”, and “Ja-pan.” In contrast, the lower nPVI-V difference was observed in AS-N: #5 (“Jack has a lot of hotels in the city now.”). In this sentence, the mean local nPVI-V was actually higher in JS (1.05) than NS (.8) for “ho-tels”, and was almost the same (JS=.54; NS=.56) for “ci-ties.” For “ho-tels”, NS pronounced “ho” as /hou/, while JS pronounced it as /ho/. This made the duration of “ho” in NS relatively long (99 ms), which made the local nPVI-V relatively low. For “ci-ties”, the high vowel of the stressed syllable, which is inherently short in duration, did not produce much difference in duration between the two groups (JS=60 ms; NS=56 ms), which resulted in the very similar local nPVI-V.

For AS-I, the difference in nPVI-V between JS and NS was significant in AS-I: #1 and AS-I: #2. The difference was especially small in AS-I: #4 (“I taught him the logic in the class just now.”). In this sentence, the mean local nPVI-V was higher in NS in only four out of eight pairs (“I taught”, “taught him”, “gic in”, and “the class”). However, the mean local nPVI-V was unexpectedly higher for “him the” in JS (.58) than NS (.38). This was because the mean normalized duration of “him” in JS (102 ms) was much longer than that of NS (71 ms) as the vowel of “him” was not reduced in the former group. The difference in the local nPVI-V for “lo-gic” was not very different between the two groups.

In SC, nPVI-V was significantly higher in NS than JS in three out of five sentences. For SC: #1 (“Bob starts his part-time job at college this year.”), the mean local nPVI-V was higher for NS than JS in six out of seven pairs, including the consecutive stressed syllable of content words, “time job” (JS=.34 and NS=.52). The mean normalized duration of “time” and “job” for JS was 101 ms and 135 ms, while that of NS was 102 ms and 175 ms, respectively, showing that the duration of “job” in NS was longer than that of JS. For SC: #2 (“I taught him how to cook nice duck legs.”), the mean local nPVI-V was higher for NS than JS in all the eight pairs, including “cook nice” (JS=.52 and NS=.67), “nice duck” (JS=.37 and NS=.57), and “duck legs” (JS=.16 and NS=.40). The mean normalized durations of “nice duck legs” were 145 ms, 102 ms, and 100 ms in JS, and 150 ms, 85 ms, and 126 ms in NS, showing that NS had an alternation of a long and short syllable even in the stress clash situation. For SC: #5 where there was no significant difference between JS and NS (“I bought a bike lock system in the city today.”), nPVI-V for “bike lock” was much larger in JS (.62) than in NS (.26), which inflated nPVI-V for JS. The mean normalized durations of “bike lock” were 183 ms and 95 ms in JS and 158 ms and 130 ms in NS. This was probably because the loan word for “bike” in Japanese is pronounced as /baiku/ where /bai/ has two morae, whereas /lo/ in the loan word, /lokku/, has one mora.

In sum, it was found that nPVI-V for SC was significantly lower than AS-N and AS-I in both NS and JS, as expected. However, nPVI-V showed a great deal of variability within each of the rhythm types across the target sentences. It was also found that nPVI-V was significantly higher in NS than JS in at least half of the sentences in each of the rhythm types. However, the degree of the difference varied greatly across the target sentences within the rhythm type. The degree and direction of the difference may be affected by phonological characteristics of a content word and the difference in vowel reduction between JS and NS. First, the difference (NS>JS) tends to be relatively large

A pilot study on the rhythmic properties of English produced by Japanese learners before and... when the pair involves a stressed syllable which has a low vowel closed by a voiced consonant (e.g., “Bob”, “Dan”, “(Ja)pan”, “job”, “(ex)am”). In contrast, the difference tends to be relatively small when a pair involves a stressed syllable which has a high vowel closed by a voiceless consonant (e.g., “ci(ty)”, “fix”, “sys(tem)”), and when a pair involves a stressed syllable of the English word which corresponds to two morae in its loan word in Japanese (e.g., “guide” [gai], “start [ta:]”, “donuts [do:]”). In addition, the direction of the difference tends to be reversed (i.e., JS>NS) when the vowel of an unstressed syllable of a disyllabic content word (e.g., “hotel”) is pronounced as a diphthong by NS, but mistakenly pronounced as a single vowel by JS. Finally nPVI-V was found to be higher for JS when a vowel in one of a succession of function words is reduced by NS but not by JS (e.g., “him in”).

3-2. Reliability and item analysis of the target sentences

An item analysis was conducted on ten target sentences in which the difference in nPVI-V between JS and NS was significant, including AS-N: #1, #2, #3, #4, #6, AS-I: #1, #2, SC: #1, #2, and #4. The other sentences were excluded from further analyses. The values of nPVI-V produced by seven NS (N=70) were submitted to the reliability analysis. The result showed that the inter-item correlations were especially low or even negative for AS-N: #2. Inspection of the data found that the participants whose mean nPVI-V were in the lower half showed very high values for this particular target sentence. Therefore, it was decided to exclude this sentence from further analyses. As a result, the cronbach’s α was .86, indicating that the reliability of the target sentences could be considered good as a measure of nPVI-V.

3-3. Comparison of nPVI-V before and after the study abroad period

Figure 1 shows that, averaged across the rhythm types and sentences, the participants whose nPVI-V were relatively high before SA showed an increase in nPVI-V toward the native mean (i.e., 63.9), while those with lower nPVI-V did not. It is notable that nPVI-V (i.e., 61.5) of JS1 approached the native mean after SA. In contrast, nPVI-V of JS5 actually decreased and became more deviant from the native average. It was decided to divide the Japanese participants into the advanced group (JS1, JS2, JS3, and JS4, called AG, henceforth) and the basic group (i.e., JS5, JS6 and JS7, called BG, henceforth). The division of the participants corresponded to the differences in terms of TOEIC scores and the proficiency levels in speaking, based on the author’s observation.

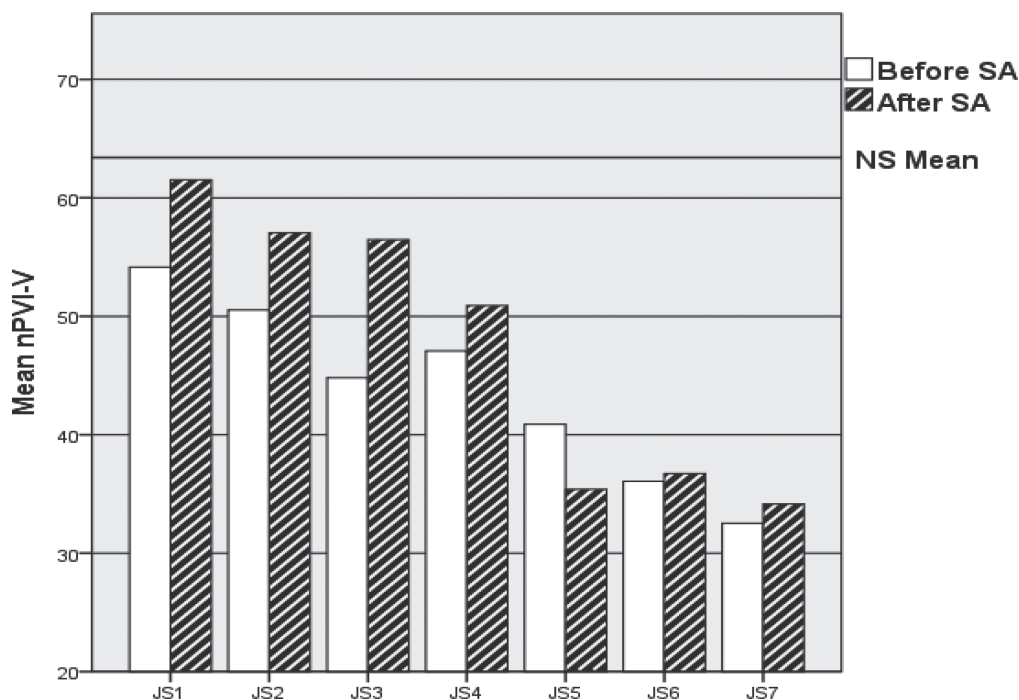


Figure 1. Mean nPVI-V averaged across the rhythm types and the sentences as a function of Japanese participants and time (i.e., before and after the study abroad period). JS=Japanese speakers, NS=native speakers.

Table 2 shows that nPVI-V slightly decreased before and after SA in BA. That of AS-N: #1 actually decreased by 7.3 points in AS-N: #1. In contrast, nPVI-V in AG showed varying degrees of a positive increase toward the native mean across the sentences in AS-N and AS-I. The results suggested that the rhythmic patterns of the alternation of stressed and unstressed syllables in content and function words in JS became more approximate to those of NS before and after SA. In SC, however, only one of three sentences showed a substantial increase in nPVI in AG. This suggested that the participants in the group had relatively more difficulty modifying the rhythmic patterns in the sentence with a stress clash.

The next analysis examined how the local nPVI-V in the sentence changed before and after SA in each participant group. First, the mean normalized vocalic durations of syllables in AS-N: #2 are shown in Table 3-1. Both before and after SA, the durations of the stressed syllables of content words (e.g., “Bob”, “start”, “job”, and “(Ja)pan”) were relatively longer in AG than BG. In addition, those of the unstressed function words (e.g., “is”, “ing”, “his”, “in”, “Ja(pan)”) were relatively shorter in AG than BG. When the vocalic

Table 2. Mean nPVI-V averaged across the participants as a function of group (i.e., advanced and basic), time (i.e., before and after the study abroad period), rhythm types, and the sentences. Dif=the difference between Before SA and After SA.

		JS						NS
Group		Basic			Advanced			
	Time	Before SA	After SA	Dif	Before SA	After SA	Dif	
Type	Sentence#	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
AS-N	1	31.0	23.8	-7.3	45.2	50.1	4.9	58.2
	2	35.3	35.0	-0.3	57.7	70.9	13.2	82.0
	3	28.5	26.5	-2.0	32.9	45.4	12.5	48.4
	4	54.3	56.8	2.5	68.3	76.1	7.8	81.2
	6	43.5	43.7	0.2	53.9	63.4	9.5	70.1
	Subtotal	38.5	37.1	-1.4	51.6	61.2	9.6	68.0
AS-I	2	44.2	43.7	-0.4	55.0	62.9	7.9	65.9
SC	1	31.1	32.9	1.8	41.8	51.7	9.9	53.3
	2	26.5	26.5	0.0	44.1	43.9	-0.2	55.2
	4	34.2	30.1	-4.1	43.3	44.5	1.3	56.5
		30.6	29.8	-0.8	43.1	46.7	3.7	55.0
	Total	36.7	35.6	-1.1	49.4	57.0	7.6	63.9

Table 3-1. Mean normalized vocalic durations of a syllable averaged across the participants as a function of group (i.e., advanced and basic) and time (i.e., before and after the study abroad period) in AS-N: #2.

Group	Time	Bob	is	start	ing	his	job	in	Ja	pan
Basic	Before SA	95.6	107.8	102.6	84.0	102.2	150.4	81.2	57.3	119.1
	After SA	105.0	100.3	113.0	75.0	82.8	119.6	96.4	60.9	147.1
Advanced	Before SA	123.4	98.1	127.3	72.3	76.8	179.1	63.3	41.6	118.3
	After SA	131.8	83.6	128.5	53.0	78.5	201.2	49.5	38.2	135.6
NS		186.2	58.9	116.7	58.5	60.3	187.3	49.5	32.4	150.2

durations were compared before and after SA in AG, those of the stressed syllables of content words (e.g., “Bob”, “job”, and “(Ja)pan”) increased, while those of unstressed function words (e.g., “is”, “ing”, “in”) decreased mostly to approximate the native values. As is shown in Table 3-2, the combination of these changes resulted in an increase in the mean local nPVI-V in “Bob-is”, “his-job”, and “job-in.” In BG, although the vocalic duration of the stressed syllables of content words increased for “Bob” and “(Ja)pan”, that of “job” actually decreased by as much as 30 ms. In addition, the magnitude of a decrease in the

Table 3-2. Mean local nPVI-V averaged across the participants as a function of group (i.e., advanced and basic) and time (i.e., before and after the study abroad period) in AS-N: #2.

Group	Time	Bob is	is start	start-ing	ing his	his job	job in	in Ja	Ja-pan
Basic	Before SA	0.17	0.11	0.21	0.21	0.38	0.63	0.42	0.70
	After SA	0.14	0.15	0.41	0.21	0.35	0.28	0.44	0.81
Advanced	Before SA	0.23	0.41	0.53	0.21	0.78	0.95	0.54	0.96
	After SA	0.46	0.44	0.83	0.42	0.85	1.19	0.35	1.13
NS		1.04	0.68	0.68	0.27	1.01	1.17	0.42	1.29

Table 4-1. Mean normalized vocalic durations of a syllable averaged across the participants as a function of group (i.e., advanced and basic) and time (i.e., before and after the study abroad period) in AS-I: #2.

Group	Time	I	found	the	jac	ket	in	the	clo	set
Basic	Before SA	136.6	122.8	79.0	86.5	89.9	96.8	49.5	153.1	85.9
	After SA	141.6	134.5	79.0	88.3	102.6	78.0	51.0	137.9	87.2
Advanced	Before SA	145.2	140.9	64.6	116.6	85.0	72.6	55.9	154.9	64.3
	After SA	128.2	151.1	60.0	131.7	84.4	64.0	52.2	157.7	70.8
NS		127.9	175.4	49.7	146.2	85.8	61.3	53.9	140.0	59.8

vocalic duration of unstressed function words (e.g., “is”, “ing”) was relatively small, and that of “in” and “Ja(pan)” actually increased. These resulted in a decrease in the local nPVI-V for “Bob is”, “his job”, and “job in”.

The mean normalized vocalic durations of syllables in AS-I: #2 is shown in Table 4-1. It is shown that the vocalic durations of “found”, “jac(ket)”, and “clo(set)” in NS were much longer than the others. In AG, the duration of “I” became shorter, while that of “found” and “jac(ket)” became longer, to approximate the native values. As is shown in Table 4-2, these changes resulted in an increase in the local nPVI-V in “I found”, “found the”, “the jac(ket)”, and “jac-ket”, indicating that the participants in this group became better able to produce this sentence with a more native-like rhythmic pattern. In BG, the duration of both “I” and “found” became longer, so that the rhythmic pattern of “I found” did not change. In addition, the duration of “jac(ket)” remained almost the same, while that of “(jac)ket” actually became longer. As a result, nPVI-V for “I found” became smaller, while that of “jac-ket” remained almost the same. Although those of “found the” and “the jac(ket)” became higher, they fell short of AG and NS.

Table 4-2. Mean local nPVI-V averaged across the participants as a function of group (i.e., advanced and basic) and time (i.e., before and after the study abroad period) in AS-I: #2.

Group	Time	I found	found the	the jac	jac-ket	ket in	in the	the clo	clo-set
Basic	Before SA	0.41	0.42	0.13	0.18	0.15	0.65	1.03	0.56
	After SA	0.28	0.53	0.33	0.19	0.27	0.50	0.93	0.46
Advanced	Before SA	0.16	0.75	0.58	0.34	0.45	0.37	0.94	0.81
	After SA	0.38	0.83	0.75	0.46	0.47	0.35	1.01	0.78
NS		0.38	1.12	0.96	0.51	0.41	0.27	0.86	0.77

Table 5-1. Mean normalized vocalic durations of a syllable averaged across the participants as a function of group (i.e., advanced and basic) and time (i.e., before and after the study abroad period) in SC: #1.

Group	Time	Bob	starts	his	part	time	job	at	col	lege
Basic	Before SA	101.5	134.7	102.2	98.9	118.7	123.0	79.5	53.1	88.5
	After SA	99.6	134.4	86.0	101.4	119.0	137.7	73.5	61.6	86.8
Advanced	Before SA	108.1	141.4	75.8	109.7	86.9	143.6	68.5	90.8	75.3
	After SA	111.9	134.1	64.7	108.1	85.8	162.9	55.7	88.3	88.6
NS		145.3	123.0	52.6	100.0	95.7	181.0	62.5	75.6	64.3

The mean normalized vocalic durations of syllables in SC: #2 is shown in Table 5-1. In NS, the vocalic duration of “his”, “part”, “time”, and “job” was 53 ms, 100 ms, 96 ms, and 181 ms, respectively, showing that the unstressed function word, “his”, was the shortest, and the last stressed content word, “job”, was the longest. In AG, this rhythmic pattern could be seen before SA. The relative vocalic durations of the syllables became more approximate to those of NS after SA, as that of “his” and “job” became shorter and longer, respectively. These resulted in the increase in the local nPVI-V of “start his”, “his part”, “time job”, and “job at”. In BG, the vocalic duration of “his” was almost as long as that of “part”, and that of “time” was as long as “job”, making all the durations relatively equal. After SA, however, the relative durations of these syllables became more similar to those of NS (although not as much as in AG), with a decrease in the duration of “his” and an increase in the duration of “job”. This led to a relatively slight increase in the local nPVI-V of “start his”, “his part”, and “time job”. In SC: #2 and SC#3 where nPVI-V did not substantially improve before and after SA, there was only a slight change in the rhythmic pattern of the syllables in content words involving a stress clash.

Table 5-2. Mean local nPVI-V averaged across the participants as a function of group (i.e., advanced and basic) and time (i.e., before and after the study abroad period) in SC: #1.

Group	Time	Bob starts	starts his	his part	part-time	time job	job at	at col	col-lege
Basic	Before SA	0.29	0.32	0.19	0.20	0.10	0.44	0.44	0.52
	After SA	0.34	0.44	0.23	0.18	0.14	0.60	0.33	0.38
Advanced	Before SA	0.26	0.60	0.38	0.26	0.49	0.69	0.35	0.32
	After SA	0.21	0.73	0.52	0.26	0.63	0.96	0.58	0.24
NS		0.42	0.78	0.62	0.18	0.61	0.97	0.41	0.34

In sum, nPVI-V changed substantially to approximate that of NS among the participants who had a relatively high proficiency level before SA (i.e., AG), but not among those whose proficiency level was relatively low (i.e., BG). In AG, nPVI-V increased in the sentences with alternation of stressed syllables of content words and unstressed syllables in content words and function words (i.e., AS-N, AS-I). In contrast, the sentences with a stress clash (i.e., SC) turned out to be relatively difficult to modify the rhythmic patterns. In AS-N, the increase in nPVI-V was largely due to lengthening of the stressed syllables of content words and reduction in duration of function words. In AS-I, in addition to the modifications just mentioned above, the rhythmic pattern of the first part of the sentence (e.g., "I found the") became more native-like. In SC, the rhythmic pattern of the stressed syllables involving the stress clash became more approximate to that of NS (e.g., "part-time job") only in one out of three target sentences. The results suggested that adult Japanese L2 learners with a certain proficiency level can learn to be able to modify the rhythmic properties of English sentences to approximate those of native speakers, by controlling the durations of stressed and unstressed syllables, especially when the sentences involve the alternation of stressed and unstressed syllables.

3-4. NS's evaluation of the rhythmic properties of JS's sentences

First of all, the analysis of reliability in the evaluation scores (ES, henceforth) showed that the reliability measure was at an acceptable level (the Cronbach's $\alpha=.78$, $N=70$). It was also found that ES and nPVI-V were significantly, but only weakly correlated in the positive direction ($r=.386$, $p=.001$). The results were not surprising because the evaluators based their evaluation on the overall rhythmic pattern of a sentence including not just duration, pitch and intensity, although nPVI-V is a measure of variability of vowel

A pilot study on the rhythmic properties of English produced by Japanese learners before and... duration alone. According to the post-hoc interview, most evaluators said that it was easy to differentiate very good from very rudimentary productions. For example, one NS said he definitely rated the production lower if the speaker put almost equal stress on content and function words. They said, however, that it was difficult to differentiate productions that might be rated somewhere in the middle range (i.e., “3” and “4” in the six-point scale). Overall, however, the results suggested that it was worth examining how ES changed before and after SA, and how they were related to the observed changes in nPVI-V.

As shown in Figure 2, the mean ES, averaged across the sentences, were higher among the participants who showed higher nPVI-V (i.e., AG (JS1, JS2, JS3, and JS4)) than those who showed lower nPVI-V (i.e., BG (JS5, JS6, and JS7)) before SA. The obvious exception was JS1. He showed the highest nPVI-V before SA (see Figure 1), but showed ES lower than the other participants in AG. According to the post-hoc interview with the evaluators, J1’s speech sounded somewhat unnaturally exaggerated. It is also shown that ES increased to varying degrees (except for JS7) across the participants before and after SA. It is notable that, after SA, JS1 and JS2 whose nPVI-V were close to

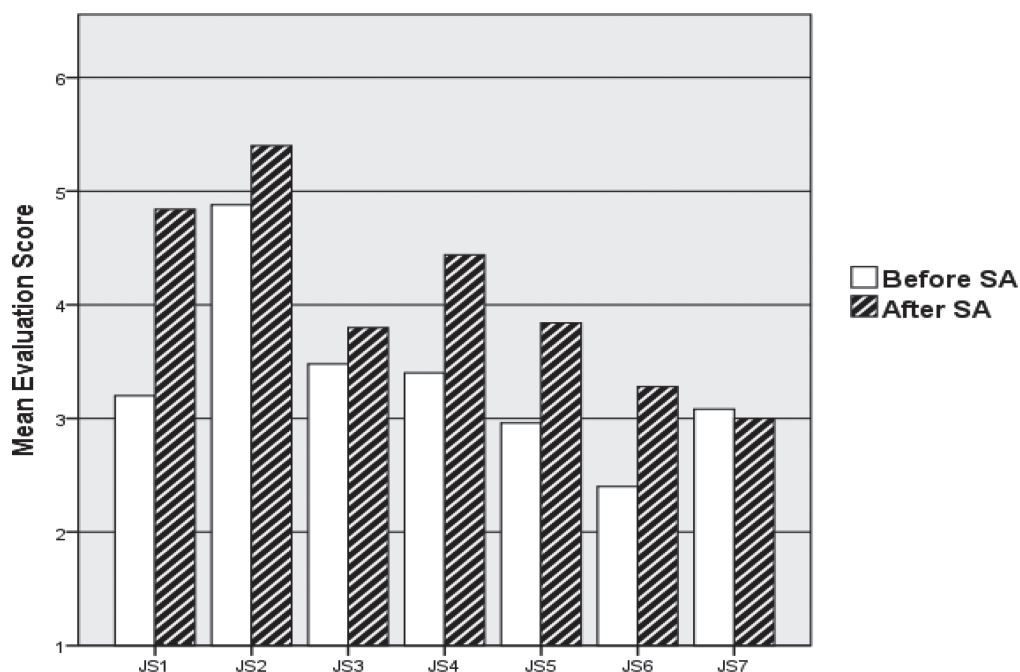


Figure 2. Mean evaluation scores (ES) averaged across the evaluators and the sentence as a function of Japanese participants and time (i.e., before and after the study abroad period). JS=Japanese speakers.

that of the average NS showed the highest ES. The exception was J3. According to the post-hoc interview, the participant's speech had peculiar pitch patterns. For two participants in BG (JS5 and JS6), the observed change in ES did not match those of ES. For JS5, ES increased while nPVI-V decreased, while for JS6, ES increased while nPVI-V remained almost the same. These discrepancies as well as the exceptions mentioned above might be the reasons why ES and nPVI-V were only weakly correlated.

Table 6 shows that, in both BG and AG, the mean ES increased to varying degrees across the target sentences. In AG, the mean ES did not differ greatly across the sentences, with the scores between four and five on the scale of six (i.e., 6="no discernible foreign accent"). The range of increment before and after SA ranged from .45 to 1.55. There was overall agreement across the sentences in the way ES and nPVI-V changed before and after SA (see Table 2 for comparison). Especially, AS-N: #2 and #3 showed the greatest gain in both ES and nPVI-V. AS-N: #3, which received the highest mean ES showed nPVI-V which was closest to the mean nPVI-V of NS (i.e., 3 points below the mean). The results suggested that the sentences with higher nPVI-V were generally evaluated as being less accented in terms of their properties of rhythm.

As was found in AG, ES after SA in BG did not differ greatly across the sentences. It is also shown that ES increased to varying degrees across the sentences. After SA, ES were between 3.07 and 3.60 in a scale of six, which were closer to the lower end. Unlike what was found in AG, there was little agreement in the way ES and SA changed before and after SA. As is shown in Table 2, nPVI-V did not increase or even decreased in some sentences. As for AS-N: #2, the decrease in nPVI-V was probably due to the decrease in

Table 6. Mean evaluation scores (ES) averaged across the evaluators and the participants as a function of group (i.e., advanced and basic), time (i.e., before and after the study abroad period), and the sentences. Dif=the difference between Before SA and After SA.

Group		Basic			Advanced		
	Time	Before SA	After SA	Dif	Before SA	After SA	Dif
Type	Sentence #	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
AS-N	1	2.93	3.60	0.67	3.65	4.55	0.90
	2	2.27	3.20	0.93	3.70	4.75	1.05
	3	2.93	3.07	0.14	3.40	4.95	1.55
	4	2.87	3.60	0.73	4.15	4.60	0.45
	6	3.07	3.40	0.33	3.80	4.25	0.45
		2.81	3.37	0.56	3.74	4.62	0.88

A pilot study on the rhythmic properties of English produced by Japanese learners before and... the vocalic duration of “job” (see Table 3-1), as mentioned above. However, the mean vocalic duration of “(Ja)pan” increased from 119 ms to 147 ms, which increased the mean local nPVI-V of “Ja-pan” from .70 to .81. The fact that the sentence finished with a better rhythm might have led the evaluators to give a higher ES. As for AS-N: #1 (“Jack has a lot of donuts in his basket now.”), the mean vocalic duration of “a” (113.4 ms) was longer than “has” (72.2 ms) before SA, while that of “a” was much shorter than “has” in NS. After SA, the mean vocalic duration of “has” was 96 ms while that of “a” was 90 ms, showing that the durations changed toward the values of NS. However, nPVI-V for “has a” became lower because the durations became more or less equal. In this case, the improvement in the rhythmic properties actually resulted in lower nPVI-V. This might be one of the reasons for the observed discrepancy between the changes in ES and nPVI-V.

In sum, the reliability of evaluation scores in terms of the rhythmic properties of JS’s productions was found to be at an acceptable level ($\alpha=.78$). Second, in AG, ES increased moderately before and after SA across the participants and the sentences, with a few exceptions. After SA, ES increased to the less accented side of the evaluation scale (i.e., between 4 and 5 on the scale of 6), indicating that the rhythmic properties became more approximate to those of NS. The changes in ES and nPVI-V before and after SA were roughly in agreement across the sentences, as the sentences with larger changes in nPVI-V also tended to show larger changes in ES. The result suggested that the changes in nPVI-V might be reflected in the evaluator’s evaluation of the rhythmic properties of the sentence. However, it was also suggested that evaluation was affected by other factors such as the idiosyncratic pitch/rhythmic pattern of the sentence. In BG, ES increased before and after SA across the participants and sentences, although to a lesser degree than AG on average. After SA, ES was still on the more accented side of the evaluation scale, indicating that the rhythmic properties of the sentences were still deviant from those of NS. The changes before and after SA in ES and nPVI-V were not in agreement, as ES increased while nPVI-V increased very little and even decreased in some cases. This was partly due to the cases where the changes in durations of syllables, which were in the direction of those of NS, resulted in a decrease in the local nPVI-V.

4. Discussion and Conclusion

4-1. Summary of the present findings

The present study attempted to investigate whether and how the rhythmic

properties of the sentences produced by adult Japanese learners of English changed before and after a 5-month study abroad program. Specifically, it examined nPVI-V (i.e., rate-normalized pairwise variability of vowels) in a set of sentences read and recited by seven L2 learners of English at two tests before and after SA, which were compared with those produced by eight native speakers of English. It also examined the effects of rhythm types (i.e., alternating stressed and unstressed syllables and a stress clash) and of the proficiency levels of the L2 learners on nPVI-V at both tests. First, the results showed that, in general, nPVI-V was significantly higher in NS than JS in the initial test, as expected. It was found, however, that the magnitude of the difference between JS and NS was influenced by factors such as inherent duration of phonemes in the stressed syllables in a particular target sentence. Second, nPVI-V of sentences which had an alternating rhythm of stressed and unstressed syllables improved to approximate that of NS before and after SA in learners with a higher proficiency level (i.e., AG) while it didn't in those with a lower proficiency level (i.e., BG). Finally, the participants' productions were evaluated by NS at both tests in terms of the degree of accentedness with respect to the rhythmic properties. The results showed that the evaluation scores improved to varying degrees in both AG and BG in most of the participants and target sentences. It was also shown that the evaluation scores and nPVI-V were significantly, but only weakly correlated. In addition, the results showed that the changes in evaluation scores were roughly in agreement with those of nPVI-V in AG, but not in BG.

4-2. Limitations of the present study

The present study has the following limitations. First, the number of participants is small (JS=7, NS=8) so that the results may not be generalized to a larger population. For both NS and JS, a relatively large variability in nPVI-V was observed across individuals and the target sentences. A study with a reasonable number of speakers is clearly needed to reliably estimate the mean and variability of the population in question. Second, the reading and recitation task were used to elicit the participants' productions. It is not clear whether and how the present results are generalizable to a task where the speaker's attention is more focused on the syntactic and lexical aspects of speech than the phonetic aspects. In future study, the interaction of the task variables and the rhythm measures should be explored to fully understand the development of L2 rhythm. Finally, the present study did not include a control group of L2 learners who received the formal instruction during the same period as the SA period. Inclusion of the control group is clearly required

A pilot study on the rhythmic properties of English produced by Japanese learners before and... in future study to assess the effects of SA on the observed changes in the rhythmic properties.

4-3. Validity of nPVI-V as a measure of the rhythmic properties of L2 productions

One of the purposes of the present study was to evaluate the validity of nPVI-V as a measure of the rhythmic properties of L2 productions. Overall, the present results suggest that researchers should be cautious about using nPVI-V especially when it is compared with that of NS. First of all, nPVI-V produced by NS significantly differed across the rhythm types. It was found that, on average, the sentences with alternating stressed and unstressed syllables showed a significantly higher nPVI-V than those which included a consecutive stressed syllables (i.e., a stress clash). This was because nPVI-V is sensitive to the number of a succession of stressed and unstressed syllables in the sentence. Second, nPVI-V produced by NS varied across the target sentences within the rhythm type. Specifically, it was found that nPVI-V was influenced by the phonemic characteristics of vowels in the stressed syllables. For example, nPVI-V tended to be higher in a pair of a stressed and unstressed syllable where the vowel in the stressed syllable is inherently longer (e.g., low vowels, closed by a voiced consonant) than otherwise (e.g., high vowels, closed by a voiceless consonant). These findings indicate that the rhythmic and phonemic properties of speech samples provided by L2 learners and NS should be reasonably held constant. Third, nPVI-V produced for a particular sentence varied a great deal among NS. Given a small sample size, it is difficult to determine the sources of this variability. It might be because of the language backgrounds as NS in the present study included speakers from a variety of countries including US, Canada, England, Australia and New Zealand. Or it might be because of some idiosyncratic ways of speaking. At any rate, it appears necessary to use large sample data from a single linguistic community to estimate the native mean and the range of nPVI-V.

The more pressing issue is to control for the factors that may influence the degree of difference in nPVI-V between L2 speakers and NS. It was found that the degree of difference is influenced by the phonemic characteristics of vowels in the stressed syllables. As far as Japanese L2 learners of English are concerned, the degree of the difference tends to increase when the vowel in the stressed syllable is inherently longer (e.g., “Bob”, “job”, “(Ja)pan”) than otherwise (e.g., “ci(ty)”), as described just above. It was also found that the difference tends to be smaller when the stressed vowel of the English word corresponds to two morae in the Japanese loan word (e.g., “guide”, “start”, “donuts), as JS

tend to apply the pronunciation of the loan word and increase the duration of the vowel. These findings suggest that there may exist a number of language-specific ways in which nPVI-V itself and the difference between that of L2 and NS might be affected depending on the prosodic and phonological characteristics of each language. One solution might be to carefully devise a set of sentences with the prosodic and phonological properties that are expected to produce a reasonable amount of difference between L2 and NS. Or one should obtain very large sample data from each of the languages such that the effects of these factors are averaged out.

Another problem of nPVI-V is that the local nPVI-V may become more deviant from that of NS even though the duration of each of the vowels becomes closer to that of NS. This happens especially in a pair of unstressed function words where the relative duration of vowels in a pair is initially reversed (e.g., “has a”: long-short in NS, but short-long in JS; see 3-4 for more details). When the duration of the longer vowel decreases and becomes similar to the other vowel (thus becoming closer to the NS value), the local nPVI-V decreases. To avoid this problem, it might be possible to eliminate pairs with a succession of unstressed syllables in computing nPVI-V.

4-4. Changes in nPVI-V before and after the study abroad program

It was found that nPVI-V increased before and after the study abroad program at least for L2 learners above a certain level of proficiency. The results are compatible with previous research which used similar sentence elicitation procedure and found that nPVI-V was significantly higher among the more proficient L2 learners and native speakers than that of the less proficient learners (Li & Post, 2014). They are also commensurate with the results of the study which found that nPVI-V among Japanese L2 learners of English increased to approximate that of native speakers of English through individual-based, long-term pronunciation training (Tsushima, 2016). However, the present results are not compatible with the previous research which showed that a rhythm index (i.e., syllable ratio) did not significantly improve after a 9-month study abroad (Gut, 2009). These discrepancies might be due to a combination of factors that have been discussed so far. It is clear that the present results should be replicated in future research with larger sample data with students with different proficiency levels.

It was found that the major source of an increase in nPVI-V was a combination of an increased vocalic duration of stressed syllables in content words and a reduced vocalic duration of unstressed syllables in content words and function words, although the degree

A pilot study on the rhythmic properties of English produced by Japanese learners before and... of the change tended to be greater in more proficient than less proficient learners. Previous research found that syllable durations produced by JS are more or less equal reflecting the “mora-timed” rhythm of Japanese (Bond & Fokes, 1985; Mochizuki-Sudo & Kiritani, 1991), and that the duration of function words produced by Japanese learners of English is relatively longer than that of native speakers of English (Aoyama & Guion, 2007). It was also found that Japanese learners of English have particular difficulty modifying syllable durations as compared with pitch and intensity (Mori et al., 2014; Tsushima, 2015). The present results suggested that, in spite of the potential difficulties facing the learners, JS were able to modify the timing control of speech in production of English sentences through exposure to native speakers and language instruction in classrooms.

4-5. Changes in native speakers’ evaluation of the L2 learners’ sentence productions before and after the study abroad program

The present study found that NS’s evaluation on the degree of accentedness in the rhythmic properties of JS’s productions more or less increased before and after SA. It was found that the reliability was at an acceptable level (the Cronbach’s $\alpha=.78$). This might be because the evaluators were asked to evaluate the rhythm of the sentence by paying attention to more than one acoustic properties relevant to speech rhythm (i.e., pitch, intensity, and duration). It is possible that evaluators might differ in the degree to which they weigh each of the acoustic properties in their evaluation. It is even possible that evaluations of the same evaluator might be influenced by different acoustic properties from one production to another. Because of these limitations, it might be reasonable that evaluation scores and nPVI-V were only weakly correlated ($r=.386$, $p=.001$). In future study, it is necessary to analyze the data on pitch and intensity and how they change before and after SA. In addition, in order to systematically examine how the changes in duration of vowels are related to their evaluations, it might be necessary to acoustically manipulate L2 learners’ speech such that only information in terms of duration is available to evaluators by making pitch and intensity constant.

4-6. Implications for the effects of study abroad on L2 phonetic development

A study abroad period offers students ample opportunities to access meaningful input and to produce context relevant output in daily interactions with native speakers, which is generally considered to provide an optimal learning context for L2 development,

including that of phonetic abilities. However, previous research found that measurable progress can be observed only in limited areas of L2 abilities (cf. DeKeyser, 2014). For example, a large-scale longitudinal study (Pérez-Vidal, 2014), which investigated linguistic and nonlinguistic development of Catalan/Spanish learners of English before and after SA over the course of 2.5 years, showed that significant improvement was found in oral fluency (e.g., speech rate, pause duration rate, Valls-Ferrer & Mora, 2014) and oral accuracy (e.g., grammatical, lexical, and pragmatic accuracy, Juan-Garau, 2014), but neither in segmental production (e.g., acoustic differentiation of nonnative vowels, Avello & Lara, 2014) nor perception of nonnative sounds (e.g., discrimination of nonnative consonants, Mora, 2014). Avello and Lara (2014, p.158) suggested that the lack of significant improvement in terms of phonetic ability might be due the communicative context of SA, where the learners' attention may focus on enhancing fluency and grammatical and lexical accuracy in oral communication, but not necessarily on accuracy in pronunciation. The present study found that significant gain in production of English rhythm was observed among the participants with a higher level of proficiency in English. It might be surmised that the learners with a certain level of oral fluency and oral accuracy were able to pay attention to and internalize the overall prosodic features of English in oral interaction with native speakers of English. On the other hand, those with a lower level of proficiency might not have afforded to pay attention to the phonetic aspects of speech, while focusing extensively on oral accuracy and fluency. If this might be the case, the present results might imply that L2 learners might be advised to reach a certain level of proficiency before SA such that they have enough online resources to pay attention to various aspects of speech including phonological properties of the target language. In this respect, Mora (2014, pp. 190-191) suggests that L2 learners need specific phonetic training on the nonnative segmental contrasts prior to the SA period. As regards speech rhythm, such training might provide the learners with opportunities to learn about the rhythmic structure of the target language (e.g., stressed and unstressed syllables) and the critical acoustic properties (e.g., pitch, duration, intensity), and to learn how to produce the rhythmic properties of the target language by appropriately manipulating the relevant acoustic properties. However, whether and how the training prior to the SA period influences the degree of improvement during the SA period remains an important empirical question that should be addressed in future research.

A pilot study on the rhythmic properties of English produced by Japanese learners before and...

4-7. Teaching implications

The present study provided the following implications for teaching the rhythm of English to Japanese learners. First of all, the present results confirmed that, in order to improve the rhythm of the learners' productions, it is important to focus on the contrast between stressed and unstressed syllables. It is important to teach the learners to use not just pitch and intensity, but also duration and vowel quality to differentiate the stressed and unstressed syllable. As practice materials, one can use disyllabic lexical items (e.g., "basket", "Japan"), as well as lexical items with a single stressed syllable in combination with a function word (e.g., "John is"). It is also necessary to teach how to reduce the duration and vowel quality in unstressed syllables of content words and function words, which Japanese students find very challenging. Second, the present results suggested that it is important to be cautious about the sentence materials used in the practice. It is advisable to use the vowels in stressed syllables that are inherently long (e.g., low vowels) so that learners can learn to lengthen the duration. It is also advisable to avoid using the lexical items which have corresponding loan words with a two-mora vowel (e.g., "donuts"). Third, the previous research, together with the present results, suggests that it generally takes a long time for learners to modify the rhythmic properties of their production, especially duration (Tsushima, 2015, 2016). So it is advised that instruction should continue for months or even more than a year, until learners become able to produce the rhythmic properties of English that are approximate to those of native speakers in spontaneous speech. Finally, it may be noted that several textbooks are available which put priority on learning the prosodic aspects of English (e.g., Cook, 2000; Gilbert, 2012).

4-8. Concluding remarks

The present study was designed as a pilot study to investigate whether and how the rhythmic properties of sentences produced by Japanese learners of English changed before and after a 5-month study abroad program. Another aim of the study was to evaluate the validity of using a rhythm index, nPVI-V, to measure and analyze the rhythmic properties of L2 learners as compared with those of native speakers of the target language. The results found that the rhythmic properties, as measured by pairwise variability of vocalic durations, became more approximate to those of the native speakers of English before and after the study abroad period, although such changes were limited

to those with more proficient learners. The results also suggested that nPVI-V may have a number of limitations if it is used as a measure to describe and analyze development of rhythm in L2 learners. Among these, the first issue involves difficulty with estimating the native level or range of the rhythm index. The second issue involves difficulty with controlling for the prosodic, phonological, and phonemic properties in the speech materials with which to compare L2 learners with native speakers. With a larger sample size for both L2 learners and native speakers of English, and further refinement of speech materials, the limitations mentioned in the present paper might be alleviated to some extent. Such a study is currently underway, and it is hoped that the new study will provide further data to provide answers to questions that remain unanswered in the present study.

Acknowledgement

The present research was supported by a research grant from Tokyo Keizai University in the academic year of 2016. The author would like to thank an anonymous reviewer for helpful comments on the earlier version of the manuscript.

Note

- 1) This notation is used henceforth to refer to the target sentence in the reading list.
- 2) A test of normality, using a Shapiro-Wilks test, showed that the data met the normality conditions for the ANOVA.
- 3) In this notation, the syllable in the parenthesis is the unstressed syllable, while the syllable outside of it is the stressed syllable.
- 4) The significant level was set at the lower value in an attempt to reduce Type I error.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Aoyama, K., & Guion, S. (2007). Prosody in second language acquisition. In O. S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: in honor of James Emil Flege*. Amsterdam; Philadelphia: John Benjamins Pub. Co.
- Avello, P., & Lara, A. R. (2014). Phonological development in L2 speech production during study abroad programmes differing in length of stay. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp.137-166). Amsterdam: John Benjamins Pub. Co.
- Barry, W., Andreeva, B., & Koreman, J. (2009). Do Rhythm Measures Reflect Perceived Rhythm? *Phonetica*, 66 (suppl 1-2), 78-94.
- Beckman, M. E. (1986). *Stress and non-stress accent* (Vol.7). Dordrecht, Riverton: Foris Publica-

- A pilot study on the rhythmic properties of English produced by Japanese learners before and...
- tions.
- Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from <http://www.praat.org/>.
- Bond, Z. S., & Fokes, J. (1985). Nonnative patterns of English syllable timing. *Journal of Phonetics*, *13*, 407-420.
- Cook, A. (2000). *American Accent Training* (2nd ed.). Hauppauge, NY: Barron's.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51-62.
- DeKeyser, R. M. (2014). Research on language development during study abroad: Methodological considerations and future perspectives. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 313-325). Amsterdam: John Benjamins Pub. Co.
- Dellwo, V., Diez, F. G., & Gavalda, N. (2009). *The development of measurable speech rhythm in Spanish speakers of English*. Paper presented at the XI Simposio Internacional de Comunicación Social, Santiago de Cuba. Retrieved from www.pholab.uzh.ch/static/volker/.../DellwoEtAl_2009_SICS.pdf
- Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *J Acoust Soc Am*, *97* (5 Pt 1), 3125-3134.
- Gilbert, J. B. (2012). *Clear Speech: Pronunciation and Listening Comprehension in North American English: Student's Book* (4th edition. ed.). Cambridge: Cambridge University Press.
- Grabe, E., & Low, E.-L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.
- Guilbault, C. P. G. (2002). *The acquisition of French rhythm by English second language learners*. (Ph. D.), University of Alberta, Alberta, Edmonton.
- Gut, U. (2009). *Non-native speech: a Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt: Peter Lang.
- Gut, U. (2012). Rhythm in L2 Speech. *Speech and Language Technology: Volume 14/15, Special Edition. Rhythm, Melody and Harmony in Speech. Studies in Honour of Wiktor Jassem* (pp. 83-94). Poznan, Poland: Polskie Towarzystwo Fonetyczne.
- Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 87-110). Amsterdam: John Benjamins Publishing Companies.
- Li, A., & Post, B. (2014). L2 Acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition*, *36* (02), 223-255. doi: 10.1017/S0272263113000752
- 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.
- Mora, J. C. (2014). The role of onset level on L2 perceptual phonological development after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 167-194). Amsterdam: John Benjamins Pub.

- Co.
- Mori, Y., Hori, T., & Erickson, D. (2014). Acoustic Correlates of English Rhythmic Patterns for American versus Japanese Speakers. *Phonetica*, 71 (2), 83-108.
- Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, 42, 244-257. doi: <http://dx.doi.org/10.1016/j.system.2013.12.004>
- Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. d. M. (2012). Measuring Child Rhythm. *Language & Speech*, 55 (2), 203-229. doi: 10.1177/0023830911417687
- Pérez-Vidal, C. (2014). *Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam: John Benjamins Pub. Co.
- Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32 (6), 693-703.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73 (3), 265-292.
- Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73-79). London: Edward Arnold.
- Tsushima, T. (2015). A case study: How do Japanese learners of English learn to use acoustic cues to differentiate stressed and unstressed syllables in pronunciation training? *Journal of Humanities & Natural Sciences*, 137, 33-57.
- Tsushima, T. (2016). Effects of individual-based, long-term pronunciation training on improvement of the ability to differentiate stressed and unstressed syllables in content and function words. *The Journal of Communication Studies*, 44, 27-58.
- Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, 4 (1), 93-118. doi: 10.1515/lp-2013-0005
- Valls-Ferrer, M., & Mora, J. C. (2014). L2 fluency development in formal instruction and study abroad: The role of initial fluency level and language contact. In C. Pérez-Vidal (Ed.), *Language Acquisition in Study Abroad and Formal Instruction Contexts* (pp. 111-136). Amsterdam: John Benjamins Pub. Co.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35 (4), 501-522. doi: <http://dx.doi.org/10.1016/j.wocn.2007.02.003>