

An exploratory case study: improvement of the ability to produce the rhythmic properties of English in a spontaneous speech task through long-term, individual-based speech training.

Teruaki Tsushima

ABSTRACT

The present paper reports an exploratory case study which investigated how two adult Japanese learners of English (J1, J2) improved the ability to produce the rhythmic properties of English in a spontaneous speech task through long-term, individual-based, speech training. It specifically examined the developmental changes in the rhythm indices (i.e., nPVI-V) and the stress-related acoustic measures (e.g., a proportion of unstressed to stressed vowels) in duration, pitch, and intensity. It also examined how the rhythm measures were related to fluency measures (e.g., articulation rate) during the training period. Spontaneous speech data were obtained in a spontaneous narrative task over the training period of 12 months ($N=88$) and 15 months ($N=117$), respectively. The rhythm measures were obtained by acoustically analyzing the sound data. It was found that, overall, the rhythm indices, the stress-related acoustic measures, as well as the fluency measures, made substantial changes toward those of native speakers, suggesting that the training was effective in improving the ability to produce the rhythmic properties of English, and to speak with greater fluency in the spontaneous speech task. It was found, however, that both participants, particularly J1, had more difficulty learning to modify duration to distinguish stressed and unstressed syllables than pitch and intensity. It was also found that, in J2, most of the fluency measures showed weak to moderate correlations with the rhythm indices and the stress-related acoustic measures, suggesting that the ability to speak with greater fluency might be related to the ability to manipulate the rhythmic properties of speech. Teaching implications of the present findings will

An exploratory case study: improvement of the ability to produce the rhythmic properties of...
be discussed.

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

1. Introduction

1-1. Review of literature

Previous research in second language (L2) learning has indicated that acquisition of the prosodic properties of a target language is important to increase comprehensibility to native as well as nonnative listeners (e.g., Derwing, Munro, & Wiebe, 1998). Hahn (2004), for example, showed that international teaching assistants' speeches were more comprehensible to and more easily recalled by U.S. undergraduate listeners when primary stresses were correctly placed than otherwise. However, it is very difficult, if not impossible, for L2 learners to successfully modify the prosodic features of their speech to approximate those of the target language (e.g., Tsushima, 2015). Many adult Japanese learners of English, for instance, have failed to acquire the ability to produce the rhythmic properties of English after many years of English learning (e.g., Tsushima, 2017). The present case study attempted to examine how two Japanese university students learned to produce the rhythmic properties of English during long-term, individual-based speech training. The present study is part of a larger study designed to examine the development of the learners' productive abilities during a three-year academic program which includes a five-month study abroad period in the latter half of the second year. The present article concerns preliminary data up to the end of the first semester of the second year, just before the study abroad period begins. During the training period, the participants received individual-based speech production training, in addition to regular classes offered by the academic program.

The difficulty Japanese learners of English may experience in learning the English speech rhythm may be partly due to the fact that the two languages have distinctly different rhythmic properties. Traditionally, Japanese is classified as a mora-timed language where the duration of a rhythmic unit, a mora, tends to be held constant, while English is classified as a stress-timed language where the duration of adjacent stressed syllables tends to be equal (e.g., Ladefoged, 1975). Although these claims in terms of the isochrony of the temporal units failed to obtain convincing acoustic evidence (Derwing et al., 1998),

the two languages have phonological and acoustic properties that may underlie the perceived rhythmic differences. First of all, Japanese has a simple syllable structure with a consonant and a vowel (i.e., CV), while English allows for a more complex syllable structure, with a consonant cluster at the beginning and/or the end of a syllable (e.g., “*Christ*”). The difference in the syllable structure is expected to make syllable durations more variable in English than in Japanese. Second, the two languages show different acoustic correlates of lexical stress/accent in both production and perception. For instance, duration, as well as fundamental frequency (f_0) and amplitude is correlated with a lexical stress in English, while only f_0 is correlated with lexical accent in Japanese (e.g., Beckman, 1986). Finally, Japanese lacks reduced vowels, while in English they occur in unstressed syllables of lexical items and function words (e.g., prepositions, determiners). They are generally short in duration, low in amplitude and f_0 . These properties of Japanese may give rise to a perceptual impression of what may be called a staccato rhythm, with a repetition of syllables with similar durations. For English, on the other hand, the properties may be linked with the perceptual impression of a recursive rhythm of strong and weak syllables.

Previous research has indicated that the rhythmic properties of Japanese mentioned above influence the prosodic characteristics of English spoken by Japanese learners of English (Mori, Hori, & Erickson, 2014; Tsushima, 2015, 2016). Mori *et al.* (2014), for example, examined pitch, duration, and intensity of vowels in English sentences where monosyllabic content and function words alternated with the recursive rhythmic pattern. It was found that the Japanese speakers (JS henceforth) predominantly used high-low pitch and strong-weak intensity to differentiate the content and function words, but that the durational differences in JS were much smaller than those of native speakers of English (NS henceforth). On the other hand, NS used durational differences to differentiate stressed and unstressed vowels as well as pitch and intensity. It was also found that JS were not able to reduce the vowel quality of function words (i.e., schwa). In training studies, Tsushima (2015, 2016) examined how individual-based, long-term speech training modified the ability of adult Japanese learners of English to use pitch, intensity, and duration to differentiate stressed and unstressed syllables in a sentence reading task. It was found that JS were able to increase the differences in pitch and intensity, while they had a great deal of difficulty modifying durational differences and the quality of reduced vowels (i.e., the first and second formant).

To account for the observed rhythmic differences among languages, previous research has proposed rhythm indices (e.g., Grabe & Low, 2002; Ramus, Nespore, & Mehler,

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

1999). One group of the indices measure the variation of vocalic or consonantal intervals in an entire speech sample (e.g., the variation coefficient of vocalic intervals; VARCO-V). Another group of the indices measure the variation of vocalic or consonantal intervals of adjacent syllables (i.e., a pairwise variability index, or PVI). The present study used the normalized pairwise variability index of vowel intervals (i.e., nPVI-V) because it has been successfully applied to acquisition of speech rhythm in some L2 studies (e.g., Li & Post, 2014; Ordin & Polyanskaya, 2014). The normalized pairwise variability index measures the averaged variations in adjacent vowel intervals in a speech sample, normalized for speaking rate. Grave and Low (2002), for example, found that nPVI-V was higher in stress-timed languages (e.g., British English, Dutch, and German) than syllable-timed languages (e.g., French and Spanish) and a mora-timed language (e.g., Japanese). The higher values of nPVI-V in English than Japanese, in particular, may be due to the differences in the phonological and phonetic/acoustic factors mentioned above. Especially, English has stressed syllables with full vowels and unstressed syllables with reduced vowels, both of which tend to alternate in a sentence. On the other hand, although Japanese has contrastive vowel length (i.e., the long and short vowel), the sentences are primarily made up of syllables (or morae) of similar vowel durations.

Previous research in L2 attempted to use some rhythm indices to examine development of the ability to produce the rhythmic properties of the target language. Some studies, which used nPVI-V in particular, have shown that the rhythm indices changed to become closer to those of NS as the learners' proficiency levels increased (Ordin & Polyanskaya, 2015). For example, Ordin et al. (2015) measured several rhythm indices, including nPVI-V, on the sentences produced by French (i.e., syllable-timed) and German (stressed-timed) learners of English with different proficiency levels (i.e., beginners, intermediate, and advanced). It was found that, in both groups, nPVI-V was higher at a more advanced than less advanced proficiency level, indicating that more advanced learners produced English sentences with greater durational variability among adjacent vowels. Other studies, however, failed to find significant increases in the rhythm indices according to different proficiency levels (Dellwo, Diez, & Gavalda, 2009; Guilbault, 2002). In fact, some researchers cautioned against using the rhythm indices in L2 rhythm acquisition research, as they can be potentially affected by 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).

More recent studies have investigated whether and how nPVI-V of adult JS changed toward those of NS during speech training or a study-abroad period (Tsushima, 2016, 2017). Tsushima (2017), for example, examined nPVI-V of Japanese university students before and after a five-month study abroad period. In the study, they were asked to memorize and read a set of fifteen sentences. One set of the sentences were composed of an alternation of stressed and unstressed syllables (i.e., the recursive rhythmic pattern), while another included a succession of stressed syllables (i.e., stress clash). At the pretest, nPVI-V averaged over the sentences was significantly higher in NS (60.8) than JS (45.8). However, the significant differences between JS and NS were found only in the sentences with the recursive rhythmic pattern. The results suggested that whether nPVI-V can differentiate native and nonnative speakers depends on particular phonological and phonetic characteristics of the target sentence. It was also found that, when the analyzed sentences were limited to those which showed the significant differences in the pretest, the participants with a higher proficiency level (N=4) showed substantial increase in nPVI-V before and after the study-abroad period (i.e., from 51.6 to 61.0). Among these students, but not those with a lower proficiency level, evaluation of the accentedness by NS also improved substantially (i.e., from 3.7 to 4.6 on a six-point scale) on the average. The study also found that the observed increase in nPVI-V was mainly due to increased vocalic duration of stressed syllables in content words and decreased vocalic duration of unstressed syllables in content and function words.

Another criticism of using nPVI-V in L2 speech development is that using only the durational measures may miss the influence of the other critical acoustic properties, namely, pitch and intensity, on measuring the English rhythm (cf., Nolan & Asu, 2009). Tsushima (2016) explored the use of the pairwise variability measure on pitch and intensity, in addition to that of duration. Three adult Japanese learners of English received long-term, individual-based speech training. Diagnostic tests were administered seven times during the entire training period (i.e., approximately 11 months). The pairwise variability of vowels, nPVI-V-D (duration), nPVI-V-P (pitch), and nPVI-V-I (intensity), were calculated based on speech samples elicited through a reading task. The analyzed sentence had a recursive rhythmic pattern of stressed and unstressed syllables. It was found that all three rhythm indices changed substantially toward those of NS across the tests. Although it is difficult to generalize the results of these case studies due to the small number of participants, the direction of the changes was commensurate with what had been found in the previous research with larger data sets (Gut, 2009; Li & Post, 2014; Ordin & Polyanskaya,

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

2015; White & Mattys, 2007). Overall, however, the available evidence remains sparse as to how the rhythm indices or some other acoustic measures can characterize the development of the ability to produce the rhythmic properties of the target language through L2 speech training.

1-2. Rationale and specific research questions

The first purpose of the present study was to examine the effects of individual-based, long-term training on the ability to produce the rhythmic properties of English in a spontaneous speech task. Two Japanese university students received individual speech production training for 12 and 15 months, and made recordings of their spontaneous speeches about 150 times and 200 times, respectively, during the training period. Then, the acoustic properties relevant to production of English speech rhythm were analyzed. As a spontaneous speech task requires online lexical, syntactic, and discourse processing in addition to articulatory planning and execution, it was expected that the rate of improvement would be much slower than what had been found in the previous research which used a reading task.

The second purpose of the present study was to investigate whether and how the rhythm indices (i.e., nPVI-V) and the stress-related acoustic measures (e.g., the durational proportion of stressed to unstressed syllables) could be used to characterize L2 learners' development of the ability to produce the rhythmic properties of English. The rhythm indices used in the study were nPVI-V in pitch and intensity, as well as duration. The present study specifically explored the use of nPVI-V on the sentences produced in a spontaneous speech task. In the analysis, the rhythm measures (e.g., nPVI-V) were averaged over a number of sentences with different phonological and phonetic characteristics. Accordingly, the rhythm indices were expected to be less likely to be influenced by some particular characteristics of a single sentence, as might be the case with sentences in a reading task. On the other hand, a spontaneous speech would involve varying length of sentences or phrases, which might negatively influence the reliability of the measure. The present study also examined how the pitch range and intensity range changed in relation to the changes in the rhythm measures and the stress-related acoustic measures.

The third purpose of the study was to examine how the fluency of the learners' speech improved during the training period. Previous research has indicated that it is relatively difficult to improve the fluency in the context of formal instruction where the target language is not spoken as a primary means of communication outside of classrooms

(i.e., the EFL environment). Valls-Ferrer and Mora (2014), for example, compared several fluency measures of 27 Catalan/Spanish adult learners of English during a six-month period of formal instruction and the following three-month period of study abroad. It was found that none of the fluency measures (e.g., the speech rate, the articulation rate, the phonation-time ratio, and the mean length of runs) significantly improved during the formal instruction period, while a majority of the measures did during the study abroad period. The present study attempted to examine whether long-term, individual-based instruction in the EFL environment is able to significantly improve the fluency of the learners' spontaneous speech. In addition, the present study attempted to examine whether and how changes in the fluency measures were related to those in the rhythm measures and the stress-related acoustic measures during the training period. It was expected that the ability to speak with a certain level of fluency would form a basis of the ability to manipulate the rhythmic properties of speech according to the prosodic structure of a phrase or a sentence.

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

- 1) How did the pairwise variability indices in duration, pitch, and intensity (i.e., nPVI-V) change during the training period?
- 2) How did the stress-related acoustic measures change during the training period?
- 3) How did the pitch and intensity range change during the training period?
- 4) How did the fluency measures change during the training period? How, if any, were the changes in the fluency measures related to those of the rhythm measures and the stress-related measures?

2. Method

2-1. Participants

Two adult learners of English at a private university in Tokyo participated in the study. Both of them belonged to an academic program that includes a five-month study abroad period in the latter half of the sophomore year. The first participant, a 20-year-old female at the time of entry into the project (J1 henceforth), joined the study abroad program at the beginning of the sophomore year. She received the speech training from the end of July in the freshman year and until the end of July in the sophomore year (i.e., approximately twelve months). The second participant, a 19-year-old female at the time of entry into the project (J2 henceforth), joined the study abroad program at the beginning

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

of the freshman year, and received the training from the end of April in the same year until the end of July in the sophomore year (i.e., approximately fifteen months). The proficiency level of their English could be categorized as pre-intermediate at the beginning of the training period, and as intermediate at the end of the period, based on their TOEIC scores. Both of them were highly motivated to study English, especially to improve their speaking and pronunciation skills.

2-2. Data Acquisition

As soon as the training began, the participants were assigned a task to record their English narrative about what happened or what they did on the day (called “daily recording”). They were encouraged to make a recording every day, but required to do it at least eight times a month. J1 made 152 recordings during the period (12.7 times per month on the average), while J2 made 205 recordings (13.6 times per month on the average). Each recording lasted approximately one minute on the average, but varied from about 45 seconds to up to more than three minutes. At the initial stage of the training period, they were asked to describe basic daily activities (e.g., get up, have a meal, leave home, attend a class, arrive home). After they became used to talking about them, they were encouraged to focus on one particular event or activity and tell a story in more details with a beginning and ending. When they became more advanced, they were occasionally given a familiar topic such as their hobby, hometown, favorite artist, etc. Recordings were done basically at home, using their mobile phone with a high-quality microphone (*Zoom IQ7*). Then, the sound file was electronically sent to the author for the acoustic analysis that followed. In some occasions, they made a recording in a sound-proof recording room in the author’s office, using a high-quality microphone (*Audio-Technica AT4040*).

2-3. Acoustic Data Analyses

The recorded sound file (16 bits of resolution with a sampling rate of 44.1 kHz) was first low-pass filtered at 8,000 Hz using “noise reduction” feature of sound analysis software, *praat* (Boersma & Weenink, 2014), and then normalized for intensity (*75db*). Using *praat*, the speech sound was segmented into a vowel and consonant portion through visual inspection of wave forms and wideband spectrograms, following standard criteria (e.g., Machač & Skarnitzl, 2009; Payne, Post, Astruc, Prieto, & Vanrell, 2012; Peterson & Lehiste, 1960). When the speech sample exceeded 60 seconds, only the sentences which

started within the first 60 seconds were analyzed. F0 and intensity (*dB*) were measured at the mid-point of the vowel portion, using a *Pratt* script. When the measurements by means of the script were not complete, a measurement was conducted manually.

Altogether, 4339 vowels (J1) and 8438 vowels (J2) were analyzed. The mean duration was 177*ms* in J1 (*SD*=112), and 188*ms* for J2 (*SD*=120). As the relatively large *SD* indicated, the distribution of the vowel duration was skewed toward a longer duration, with a maximum value of 988*ms* in J1, and 913*ms* for J2. These long vowels were mostly due to a disfluent speech where the speakers disproportionately lengthened a word (e.g., “I”, “is”). It was decided to exclude vowels longer than 350*ms* (approximately over 90 percentile) from the analyses of duration.

2-4. Rhythm indices

The present study used a normalized pairwise variability index of vowels in duration (nPVI-V-D henceforth), pitch (nPVI-V-P henceforth), and intensity (nPVI-V-I henceforth). First of all, it should be noted that the last pair of vowels in each sentence was excluded from calculation of the rhythm indices to eliminate potential effects of sentence-final lengthening and/or sentence-final pitch changes. The nPVI-V values were calculated for a succession of CV sequences without a silent period of 300*ms* or longer (i.e., a pause). So the shortest sequence on which the value was calculated was a pair of a CV sequence (e.g., “I think”) followed by a pause. The sequence was also terminated at the end of a syntactically determined sentence, whether the next sentence began with or without a pause (e.g., “I got up at seven.”). For nPVI-V-D, the absolute durational difference of adjacent vowels was calculated for each pair, which was divided by the mean duration of both vowels to normalize for a speaking rate. Then, the normalized values for all the pairs in an entire narrative were averaged, and then multiplied by 100. For nPVI-V-P, f0 values were first transformed into *mel* values such that the numeric difference in *mel* corresponds to a perceptual difference in pitch. Then, the difference in *mel* between adjacent vowels was averaged over all the pairs in the entire narrative. For nPVI-V-I, the difference in *dB* between adjacent vowels was averaged over all the vowel pairs in the entire narrative.

2-5. Measures of the acoustic difference in duration, pitch, and intensity between stressed and unstressed vowels

Each syllable in a speech was given a code of 1) a stressed syllable of a content word

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

(e.g., “break” of *breakfast*), 2) an unstressed syllable of a content word (e.g., “fast” of *breakfast*), or 3) a syllable of a function word (e.g., “in”). For words of more than two syllables, only a syllable with a primary stress was coded as stressed (e.g., “na” in *combination*). For a monosyllabic content word (e.g., *have*, *get*), either 1) or 2) was given depending on the prosodic structure of the phrase or sentence. For example, “got” in “I got up at seven.” was coded as “an unstressed syllable of a content word,” while “got” in “I got a prize.” was coded as “a stressed syllable of a content word.” Function words included prepositions (e.g., *in*), articles (e.g., *a*, *the*), and *to*-infinitives (i.e., *to*). Although not considered as function words, “be-verbs” (e.g., *is*, *was*) were included in this category as the vowels are generally unstressed and reduced with shorter duration, lower pitch, and smaller intensity. Pronouns (e.g., *him*), auxiliary verbs (e.g., *can*), conjunctions (e.g., *if*) and others were excluded from the analysis.

For the durational difference between the stressed and unstressed vowel of content words (STCN-D henceforth), the proportion of all the unstressed vowels to all the stressed vowels in an entire narrative was calculated and multiplied by 100. For the pitch difference between the stressed and unstressed vowel of content words (STCN-P henceforth), the mean *mel* of the unstressed vowels averaged over all the stressed vowels in an entire narrative was subtracted from that of the stressed vowels. For the intensity difference between the stressed and unstressed vowel of content words (STCN-I henceforth), the mean *dB* of all the unstressed vowels in an entire narrative was subtracted from that of the stressed vowels. The stress-related acoustic measures of stressed vowels of content words and function words in duration (STFN-D henceforth), pitch (STFN-P henceforth), and intensity (STFN-I henceforth) were calculated in the same way.

2-6. Pitch range and intensity range

The pitch range (RNG-P henceforth) and the intensity range (RNG-I henceforth) were calculated using a sentence as a unit of analysis. A difference between the minimal and maximum value in each sentence was averaged over all the sentences in an entire narrative.

2-7. Fluency measures

All the fluency measures were calculated following the definitions provided in Vallas-Ferrer & Mora (2014: p.120). Speech Rate (SR henceforth) is “total number of words produced in a given speech sample divided by the amount of total time required to pro-

duce the speech sample (including pause time) expressed in minutes.” Articulation Rate (AR henceforth) is “total number of words produced in a given speech sample divided by the amount of time taken to produce them (excluding pause time) expressed in minutes.” Phonation-time Ratio (PhonRat henceforth) is “percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample.” Following Vallas-Ferrer & Mora (2014), a pause was defined as a silence period equal to or longer than 300ms. Finally, Mean Length of Runs (MLoR henceforth) is “average number of words produced in utterances between pauses of 300ms and above.”

2-8. Data analyses

The entire period of training was divided into four sub-periods with a length of three months each, in order to examine potential longitudinal changes in the data. All the measures described above were averaged within each sub-period. A total length of training for J1 was 12 months, which was divided into four sub-periods (termed T1, T2, T3, T4, respectively), with 89 narratives analyzed. Below is how each sub-period corresponded to the academic schedule.

T1: summer vacation and the 1st month of the 2nd semester (freshman)

T2: the rest of the 2nd semester

T3: spring vacation and the 1st month of the 1st semester (sophomore)

T4: the rest the 1st semester

A total length of training for J2 was 15 months, which was divided into five sub-periods (i.e., from T1 to T5), with 117 narratives analyzed. Below is how each sub-period corresponded to the academic schedule.

T1: the 1st semester (freshman)

T2: summer vacation and the 1st month of the 2nd semester

T3: the rest of the 2nd semester

T4: spring vacation and the 1st month of the 1st semester (sophomore)

T5: the rest the 1st semester

The data obtained from the participants (JS) were compared to those of native speakers of English (NS). Unfortunately, data of NS performing the same task were not available at the time of writing. Instead, it was decided to use the following two sets of data. The first set, obtained for the author’s previous study, came from NS (N=6, all males from U.S. (N=3), Canada (N=1), England (N=1), and New Zealand (N=1)) who performed a task of telling a story based on a sequence of pictures (i.e., four pictures describ-

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

ing a boy scout leader and children going camping). Another set came from NS (N=2, both female from Canada and U.S.A.) who were hosts of a podcast program called *Culips ESL Podcast* (<https://esl.culips.com/>). This is a program for learners of English where the hosts have conversations about a variety of topics including daily activities, English expressions, and intercultural understanding. The portion of the recording where the speaker spoke about some event (i.e., getting late for work), or explains some idea (i.e., veganism, linking, alternative medicine) was used for the present analysis. Two sets of recording for each NS were used for the analysis. The task of NS was similar to that of JS in that it involved spontaneous speech about events, activities, and ideas toward a potential listener. The first 30 seconds of their speech data were acoustically analyzed using the same criteria described above. The same measures as were used for JS were calculated and averaged over all NS.

2-9. Training sessions

The present pronunciation training primarily followed a pronunciation textbook, *Clear Speech*, which emphasizes the importance of learning the prosodic aspect of speech, especially the English speech rhythm and the distinction between stress and unstressed syllables (Gilbert, 2001). Over the course of the training period, the participants studied the first eight units which included a topic such as the notion of syllable, how to pronounce English vowels, the notion of stress, how to pronounce stressed/unstressed/reduced vowels, and the notion of sentence focus. In addition to studying the textbook, the participants were engaged in a variety of speaking activities relevant to the content of the text book. The activities included 1) reading aloud, 2) retelling a story, 3) making a short speech about a given topic, 4) producing a sentence given a key word, 5) describing pictures in a vocabulary picture book, and describing a set of toy items (e.g., cars, animals). In all of the activities, they were asked to pay attention to the rhythmic properties of English. In addition, the participants and the author occasionally reviewed the recording they had recently made by looking at the acoustic properties of speech on *praat* software. Based on the review, the participants were asked to modify the rhythmic properties of the speech.

J1 had one 90-minute practice session per week except during a break between the semesters. She had a total of 32 practice sessions during the training period. J2 had two or three sessions per week during the freshman year, and had one or two sessions per week during the sophomore year. J2 had a total of 80 sessions during the training period.

3. Results

3-1. Fluency measures

Table 1 shows that, in J1, all the fluency measures except for the phonation-time ratio (PhoneRat), substantially increased to become closer to the NS means between T2 and T3. For J2, all the fluency measures showed a substantial increase toward the NS means between T2 and T4. These results indicated that the ability to speak with a greater speed and the ability to produce words without pausing substantially improved during these sub-periods in both participants. It should also be noted that all the fluency measures

Participant			T1	T2	T3	T4	NS		
J1	SR	<i>M</i>	48.2	41.7	67.8	57.2	149.2		
		<i>SD</i>	8.1	5.5	13.3	7.9	36.4		
	AR	<i>M</i>	106.8	113.9	137.4	126.6	193.0		
		<i>SD</i>	11.2	7.9	12.2	10.3	41.2		
	PhoneRat	<i>M</i>	45.1	36.7	49.1	45.1	77.2		
		<i>SD</i>	5.0	4.9	6.5	4.4	6.1		
	MLoR	<i>M</i>	2.2	1.8	3.1	2.4	7.3		
		<i>SD</i>	0.4	0.2	0.7	0.4	3.2		
		<i>N</i>	22	20	20	26	10		
				T1	T2	T3	T4	T5	NS
	J2	SR	<i>M</i>	57.6	57.3	68.5	82.6	77.7	149.2
			<i>SD</i>	6.1	8.1	10.3	10.1	9.9	36.4
AR		<i>M</i>	105.0	107.5	119.8	131.8	127.2	193.0	
		<i>SD</i>	9.1	18.0	9.5	10.6	12.5	41.2	
PhoneRat		<i>M</i>	55.2	53.7	57.0	62.6	61.2	77.2	
		<i>SD</i>	6.8	5.2	5.6	5.2	5.8	6.1	
MLoR		<i>M</i>	2.6	2.5	2.8	3.8	3.6	7.3	
		<i>SD</i>	0.4	0.4	0.7	0.7	1.4	3.2	
		<i>N</i>	23	26	24	22	22	10	

Table 1. Mean (*M*) and one standard deviation (*SD*) of fluency measures averaged over narratives recorded within each of the four sub-periods. J1/J2=Japanese participants; NS=Native speakers of English; SR=Speech Rate; AR=Articulation Rate; PhoneRat=Phonation-time Ratio; MLoR=Mean Length of Runs.

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

were far from those of NS even at the last training sub-period, showing that the participants still had a great deal to improve in terms of fluency.

3-2. Normalized pairwise variability index for duration, pitch, and intensity

As is shown in Table 2, the normalized pairwise variability of vowel duration (i.e., nPVI-V-D) in J1 increased substantially toward the NS mean (57.3) between T2 and T4. J2 showed a substantial increase in nPVI-V-D between T2 and T3. It should be noted that, for both participants, nPVI-V-D at the last sub-period (T4 and T5) was not very different from the NS mean. As for the normalized pairwise variability of pitch, nPVI-V-P in J1 increased substantially between T1 and T2, and became comparable to the NS mean (28.9) at T2. Inspection of the data in the first four months found that nPVI-V-P increased from 14.6 (the first month), 15.6 (the second month), 20.9 (the third months), and to 30.7 (the

Participant			T1	T2	T3	T4	NS		
J1	nPVI-V-D	<i>M</i>	48.6	47.8	51.7	55.1	57.3		
		<i>SD</i>	8.3	11.1	7.5	8.2	9.0		
	nPVI-V-P	<i>M</i>	16.9	28.2	31.0	29.1	28.9		
		<i>SD</i>	5.5	7.0	6.4	5.9	8.9		
	nPVI-V-I	<i>M</i>	2.7	3.2	3.0	3.4	3.1		
		<i>SD</i>	0.7	1.1	0.5	0.7	0.6		
		<i>N</i>	22	20	20	26	10		
			T1	T2	T3	T4	T5	NS	
	J2	nPVI-V-D	<i>M</i>	49.0	50.0	55.7	54.9	55.7	57.3
			<i>SD</i>	8.8	11.5	10.6	6.6	6.9	9.0
nPVI-V-P		<i>M</i>	30.2	25.9	26.8	30.4	29.2	28.9	
		<i>SD</i>	5.9	5.0	4.0	4.4	5.4	8.9	
nPVI-V-I		<i>M</i>	3.0	3.2	3.8	4.1	3.9	3.1	
		<i>SD</i>	0.6	0.6	0.8	0.9	0.7	0.6	
		<i>N</i>	23	26	24	22	22	10	

Table 2. Mean (*M*) and one standard deviation (*SD*) of the normalized pairwise variability index for vowels averaged over narratives recorded within each of the four sub-periods. J1/J2=Japanese participants; NS=Native speakers of English; nPVI-V-D=the normalized pairwise variability index of duration for vowels; nPVI-V-P=the normalized pairwise variability index of pitch for vowels; nPVI-V-I=the normalized pairwise variability index of intensity for vowels.

fourth month). The observed increase at the beginning of the training period matched the author's perceptual impression. Namely, at the beginning of the training period, J1 produced a succession of syllables with relatively flat pitch, which is typically found in the beginning learners of English. It was observed that J1 was able to relatively quickly learn to make use of pitch changes to mark stressed syllables during the following sub-period. In J2, nPVI-V-P was at the same level as the NS mean even in T1, and did not change substantially throughout the training period. Inspection of the data in the first four months showed that nPVI-V-P was already high in the first month (i.e., 29.2). As for the normalized pairwise variability of intensity, nPVI-V-I in J1 increased substantially from T1 and T2, where nPVI-V-I was greater than that of the NS mean. In J2, nPVI-V-I was already close to the NS mean at T1. In sum, it was found that nPVI-V-D increased substantially to approximate the NS mean through the training period, and that nPVI-V-P and nPVI-V-I reached the NS mean in the early stage of the training period in both participants.

3-3. Pitch and intensity range

The pitch range increased substantially between T1 and T3 in J1, while it increased from T1 through T5 in J2. In J1, the maximum pitch increased from 228*mel* in T1 to

Participant			T1	T2	T3	T4	NS	
J1	RNG-P	<i>M</i>	33.6	57.4	68.9	63.7	95.4	
		<i>SD</i>	8.8	8.7	11.6	14.5	43.4	
	RNG-I	<i>M</i>	6.5	7.4	8.7	8.9	12.1	
		<i>SD</i>	1.7	1.8	1.6	1.9	2.3	
		<i>N</i>	22	20	20	26	10	
				T1	T2	T3	T4	T5
J2	RNG-P	<i>M</i>	56.5	58.8	56.1	69.6	66.6	95.4
		<i>SD</i>	9.0	11.1	11.5	19.8	15.7	43.4
	RNG-I	<i>M</i>	7.8	8.1	9.1	10.2	10.8	12.1
		<i>SD</i>	2.0	1.2	1.7	3.5	2.1	2.3
		<i>N</i>	23	26	24	22	22	10

Table 3. Mean (*M*) and one standard deviation (*SD*) of the pitch range and intensity range averaged over narratives recorded within each of the four sub-periods. J1/J2=Japanese participants; NS=Native speakers of English; RNG-P=pitch range (in *mel*) ; RNG-I=intensity range (in *dB*).

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

242*mel* in T4 while the minimum pitch decreased from 195*mel* to 179*ms*. In J2, the maximum pitch increased from 254*mel* in T1 to 259*mel* in T4 while the minimum pitch decreased from 198*mel* to 189*mel*. The range of intensity also increased toward the NS mean throughout the training period in both participants. It should be noted, however, that the values at the last sub-period still fell short of the NS mean for both measures. The results indicated that both participants became able to use a wider range of pitch and intensity through the speech training.

3-4. Stress-related acoustic measures; Comparison between stressed and unstressed vowels of content words

As is shown in Table 4, the durational proportion of unstressed to stressed vowels of content words (i.e., STCN-D) slightly decreased from T1 to T4 in J1. As the articulation rate increased (see Table 1), the averaged duration for stressed vowels of content words decreased from 169*ms* in T1 to 143*ms* in T4, and that of unstressed vowels from 179*ms* to 135*ms*. However, the proportional differences did not change substantially through the training period. J2, on the other hand, showed a sharp drop in STCN-D, especially between T3 and T4, toward the NS mean (74%). The averaged duration for stressed vowels of content words decreased from 178*ms* in T1 to 156*ms* in T5, while that of unstressed vowels from 169*ms* to 132*ms*, showing that the duration of unstressed vowels decreased to a greater extent than stressed vowels.

The pitch difference between stressed and unstressed vowels (STCN-P) started with a negative value (-5.7*mel*) in J1, showing that the average pitch was higher for unstressed than stressed vowels. STCN-P, however, increased constantly toward the NS mean (15.4 *mel*) throughout the training period. While the pitch height of stressed vowels stayed at the same level, that of unstressed vowels decreased from 299*mel* in T1 to 287*mel* in T4. STCN-P of J2 started from a much higher value than that of J1 at T1, and increased to the same level as the NS mean at T4. Especially, the duration of unstressed vowels decreased from 314*mel* in T1 to 303*mel* in T4.

The intensity difference between stressed and unstressed vowels (STCN-I) also started with a negative value (-0.2) in J1, but increased to the level of the NS mean (1.5) at T4. STCN-I of J2 was already higher than the NS mean at T1, but further increased to the level much higher than the NS mean in the rest of the training period.

In sum, the results indicated that J1 became able to use pitch and intensity to differentiate stressed and unstressed vowels in content words, but had difficulty modifying du-

Participant			T1	T2	T3	T4	NS	
J1	STCN-D	<i>M</i>	107.8	93.4	92.3	94.8	74.4	
		<i>SD</i>	20.8	16.5	20.0	12.1	11.1	
	STCN-P	<i>M</i>	-5.7	1.9	6.2	10.1	15.4	
		<i>SD</i>	7.6	12.5	12.8	8.5	16.4	
	STCN-I	<i>M</i>	-0.3	0.9	1.3	1.5	1.5	
		<i>SD</i>	1.1	1.9	0.9	1.0	1.0	
	<i>N</i>	22	20	20	26	10		
			T1	T2	T3	T4	T5	NS
J2	STCN-D	<i>M</i>	95.6	96.9	92.8	80.5	85.2	74.4
		<i>SD</i>	16.9	14.6	19.2	11.5	14.4	11.1
	STCN-P	<i>M</i>	8.2	11.4	10.5	16.0	23.6	15.4
		<i>SD</i>	8.6	8.9	8.7	9.6	13.0	16.4
	STCN-I	<i>M</i>	1.7	2.0	3.1	3.8	3.2	1.5
		<i>SD</i>	1.5	1.1	1.1	1.5	1.1	1.0
	<i>N</i>	23	26	24	22	22	10	

Table 4. Mean (*M*) and one standard deviation (*SD*) of the differences in duration, pitch, and intensity between stressed and unstressed vowels in content words, averaged over narratives recorded within each of the four sub-periods. J1/J2=Japanese participants; NS=Native speakers of English; STCN-D=the proportion of durations in unstressed vowels of content words to that of stressed vowels of content words; STCN-P=the difference in pitch (in *mel*) between stressed and unstressed vowels of content words; STCN-I=the difference in intensity (in *dB*) between stressed and unstressed vowels of content words.

ration. J2, on the other hand, was able to learn to use all three properties to differentiate them.

3-5. Stress-related acoustic measures; Comparison between stressed vowels of content words and unstressed vowels of function words

As is shown in Table 5, the durational proportion of unstressed vowels of function words to stressed vowels of content words (STFN-D) in J1 was over 100% at T1, and stayed in the 90% range throughout the training period. This may be partly because she occasionally used prolonged vowels for be-verbs (e.g. *is*) and to-infinitives (i.e., *to*) even at

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

the end of the training period. J2, on the other hand, showed a substantial decrease in STFND in T4. Between T1 and T4, while the duration of stress vowels decreased slightly from 178ms to 169ms, that of vowels in function words decreased greatly from 187ms to 111ms. This clearly indicated that J2 was able to learn to shorten the vowels of function words through the training.

In J1, the pitch difference between stressed vowels of content words and unstressed vowels of function words (STFN-P) increased substantially between T1 and T2, and further increased to approximate the NS mean (23.7) at the end of the training periods. While the

Participant			T1	T2	T3	T4	NS	
J1	STFN-D	<i>M</i>	109.2	119.4	90.2	102.5	53.6	
		<i>SD</i>	29.5	34.3	22.7	31.2	10.8	
	STFN-P	<i>M</i>	8.7	19.8	20.7	22.2	20.6	
		<i>SD</i>	11.5	11.6	11.1	15.7	13.8	
	STFN-I	<i>M</i>	1.2	2.4	2.0	2.8	2.0	
		<i>SD</i>	1.6	2.2	1.2	1.9	1.5	
			T1	T2	T3	T4	T5	NS
J2	STFN-D	<i>M</i>	104.9	96.7	76.5	66.6	82.3	53.6
		<i>SD</i>	23.8	23.7	22.3	16.5	14.6	10.8
	STFN-P	<i>M</i>	17.6	20.9	20.8	23.1	23.2	20.6
		<i>SD</i>	10.0	8.2	8.6	7.6	13.1	13.8
	STFN-I	<i>M</i>	1.6	2.2	3.4	4.2	3.3	2.0
		<i>SD</i>	1.2	1.0	1.1	1.5	1.8	1.5

Table 5. Mean (*M*) and one standard deviation (*SD*) of the differences in duration, pitch, and intensity between stressed vowels of content words and unstressed vowels of function words, averaged over narratives recorded within each of the four sub-periods. J1/J2=Japanese participants; NS=Native speakers of English; STFND=the proportion of durations in unstressed vowels of function words to that of stressed vowels of content words; STFN-P=the difference in pitch (in *mel*) between stressed vowels of content words and unstressed vowels of function words; STFN-I=the difference in intensity (in *dB*) between stressed vowels of content words and unstressed vowels of function words.

pitch of stressed vowels stayed at the same level, that of vowels in function words decreased from 285mel in T1 to 275mel in T4. As was the case with STCN-P, J2's STFNP was much higher than J1 in T1, and increased steadily to approximate the NS mean in T5. The pitch of vowels in function words decreased from 305mel in T1 to 295mel in T5.

The intensity difference between stressed vowels of content words and unstressed vowels of function words (STFN-I) in J1 increased to reach the NS mean (2.8) in Time 4, while that of J2 surpassed the NS mean at Time 3 (3.4), and remained high through the training period. In sum, the results indicated that, as in the case of differentiation of stressed and unstressed vowels of content words, J1 became able to differentiate the unstressed vowels of function words and the stressed vowels of content words by using pitch and intensity, but had difficulty modifying vowel duration. J2, on the other hand, learned to use all the three acoustic properties to differentiate them through the training period.

3-6. Correlations between the rhythm measures and the fluency measures

Pearson correlation analysis was conducted to explore the relations between the rhythm indices, the stress-related acoustic measures, and the fluency measures separately for two participants (two-tailed, $N=88$ in J1, $N=117$ in J2). Table 6 shows that, in J1, there

	J1				J2			
	SR	AR	PhonRat	MLoR	SR	AR	PhonRat	MLoR
nPVI-V-D	0.08	0.12	0.05	0.09	.20*	.28**	0.03	0.03
nPVI-V-P	.31**	.44**	0.08	.28**	.25**	0.12	.26**	.21*
nPVI-V-I	-0.06	0.04	-0.14	-0.10	.46**	.47**	.23*	.26**
STCN-D	-0.10	-0.14	-0.02	-0.14	-.36**	-.27**	-.28**	-.26**
STFN-D	-0.19	-.25*	-0.07	-.29**	-.52**	-.59**	-0.18	-.34**
STCN-P	0.17	0.19	0.10	0.14	0.09	-0.02	0.17	-0.01
STFN-P	.24*	.30**	0.10	0.14	0.06	-0.07	.193*	-0.04
STCN-I	0.05	0.16	-0.08	0.02	.39**	.32**	.29**	.26**
STFN-I	0.00	0.13	-0.12	-0.08	.47**	.40**	.34**	.30**

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 6. Pearson correlation coefficients, r , among the rhythm indices, the stress-related acoustic measures, and fluency measures. Refer to the tables above for information on abbreviations.

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

were only a few rhythm indices and the stress-related acoustic measures which showed a significant correlation with the fluency measures. These measures (i.e. nPVI-V-P, STFN-P, STFN-D) showed a substantial increase or decrease during the training periods. It should be noted that, although they were statistically significant, the correlations were mostly weak, with the coefficient, r , below .30.

In J2, on the other hand, most of the fluency measures were significantly correlated with the rhythm indices and the stress-related measures except for STCN-P and STFN-P. For the duration measures, the highest correlations were found in STFN-D, whose substantial drop coincided with the sharp increase in SR and AR (i.e., between Time 2 and T4). STCN-D was also significantly correlated with all the fluency measures except for PhonRat. Finally, all the intensity measures were significantly correlated with all the fluency measures.

4. Discussion and Conclusion

4-1. Summary of the present findings

The present study was designed to examine how the rhythmic properties of English produced by two adult Japanese learners of English changed during the long-term, individual-based speech production training. It specifically examined changes in the pairwise variability among adjacent vowels in terms of duration, pitch, and intensity, as well as the degree to which stressed and unstressed vowels were differentiated in terms of the three acoustic properties. It also examined changes in the degree of pitch and intensity range, as well as that of fluency.

The results found the following. First of all, most of the fluency measures changed toward the native means in both participants. Especially, both showed a substantial increase in speaking rate (AR and SR) in the middle of the training period. The results clearly showed that both participants became able to speak at a faster speaking rate, and to produce a longer succession of words during the training. Second, in both participants, the normalized pairwise variability for vowels in all the three properties approximated the NS means. The results indicated that the participants became better able to vary all three acoustic properties between adjacent vowels through the training period. Third, the durational proportion of unstressed to stressed vowels in content words, and that of unstressed vowels in function words to stressed vowels in content words, decreased toward the NS means in both participants, although the degree of the decrease was much larger

in J2 than in J1. However, the proportion in the last training sub-period was much higher than the NS means, indicating that the participants still had difficulty shortening the unstressed vowels. In both participants, the difference in pitch between stressed and unstressed vowels in content words approximated, but fell a little bit short of the NS mean, while that of stressed vowels of content words and unstressed vowels in function words reached the NS means. In both types of the contrast between stressed and unstressed vowels, the difference in intensity increased to match the NS means in both participants. The results indicated that both participants became able to differentiate stressed and unstressed vowels by means of pitch and intensity through the training period. Finally, the fluency measures were significantly correlated with most of the rhythm indices and the stress-related acoustic measures in J2, while only a few correlations showed significance in J1.

4-2. Limitations of the present study

First of all, as this is a case study with only two participants, the present results may not be generalizable to other L2 language learners. It should be noted, however, that the general direction of the changes in the measures were consonant with those reported in previous research with larger data samples (Ordin & Polyanskaya, 2015). Second, the speech data from NS were not adequate as the controls, because the number of data sets was not large enough, and because they were not obtained using the same task as JS. The author is planning to obtain new sets of data from NS in Australia where the study abroad takes place. Third, at present, no control data are available from L2 learners learning English without receiving the individual-based training, so that it is difficult to evaluate the degree of the effects of the training. Data acquisition on the controls is currently underway. Finally, it should be noted that the present data will be supplemented with the data on NS's evaluation of the participants' narratives in terms of accentedness and comprehensibility.

4-3. Rhythm measures

One of the purposes of the present study was to explore the use of the pairwise variability indices (i.e., nPVI-V-D, nPVI-V-P, and nPVI-V-I) for spontaneously produced data. First of all, it was found that nPVI-V-D increased to approach the NS means in both participants, as their proficiency levels improved during the training period. The results were in agreement with the previous research which showed a significant increase among the

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

learners with different levels of proficiency (Ordin & Polyanskaya, 2015), before and after a study abroad period (Tsushima, 2017), and during speech production training (Tsushima, 2016). As for the pairwise variability indices for pitch and intensity, JI showed a substantial increase toward the NS means in both measures, in line with the observed increase in the other measures, including the stress-related acoustic measures (i.e., STCN-P/I, STFN-P/I) and those of pitch/intensity ranges (i.e., RNG-P/I). These findings suggest that nPVI-V-P/I reflects the participants' increased ability to vary pitch and intensity among adjacent vowels spontaneously in producing English sentences.

However, the present analysis found the following problems with respect to using the pairwise variability indices for analysis with spontaneous speech. The first problem particularly concerns the use of nPVI-V-D. A spontaneous speech contains a sizable number of lengthened vowels apparently due to dysfluency. In the present analysis, the vowels which were longer than $350ms$ (over .90 percentile of the data) were excluded. However, exclusion of such data might have led to missing the true nature of the timing control skills especially at the early stage of training. Second, the pairwise variability indices may reflect variability due to "incorrect" English rhythm, as well as "correct" one. In the case of duration, the syllables which are normally long are often short, and vice versa, especially when the fluency level is low. For example, in a phrase, "in the class", the vowels of the function words, "in the" could be longer than that of the content word, "class." Even in this case, the value of nPVI-V-D can be as large as when the phrase is produced in natural English rhythm. Finally, as a spontaneous speech contains a lot of pauses especially in the early stage of training, the number of words between the pauses can be small, as shown by the mean length of runs (MLoR). As a result, calculation of the pairwise indices was often based on a small number of pairs (e.g., "I went to"). In these cases, it is not clear if the index reflects the prosodic structure of a longer stretch of words or phrases. In addition, it is not clear how the indices are influenced by pre-pausal lengthening. For these reasons, the present analysis found it difficult to use the pairwise variability indices to characterize the English rhythm of L2 learners using spontaneous production data.

The advantages of using the stress-related acoustic measures (i.e., STCN-D/P/I, STFN-D/P/I) are the following. First, as the vowel durations are averaged over a sizable number of vowels in a speech sample, the values are less influenced by extreme values than in the case of pairwise variability measures. Second, "incorrect" use of the acoustic property is reflected in the values of the measures. For example, when the duration of a stressed vowel is shorter than that of an unstressed vowel, the value will exceed 100%,

while the pairwise variability measure only reflects the variability between the two. In addition, the stress-related acoustic measures were better correlated with the fluency measures than the pairwise variability measures (see 3-6), suggesting that the former better reflects the observed improvement over the course of the training period. One possible disadvantage of the stress-related acoustic measures is that they use only a limited set of syllables in a sentence, namely, stressed/unstressed syllables of content and function words. It could be argued that these syllables may make up a large portion of the syllables produced by L2 learners. Yet, they may fail to capture some acoustic characteristics of global rhythm production. In addition, there may be a difficulty of determining the criteria to decide whether a particular syllable should be coded as “stressed” in a certain context, as a decision may be influenced by a number of contextual factors in a spontaneous speech.

4-4. Effects of training

The preliminary results presented above strongly indicate that the individual-based speech production training was effective in improving the ability of the two participants to produce the rhythmic properties of English in a spontaneous speech task. It was found that, in J1, the level of achievement relative to the NS means was much greater for pitch and intensity than duration, indicating that she had greater difficulty improving the ability to modify duration. In the training, the notion of syllable stress was introduced to J1 in the third week (T1) using the textbook (i.e., Unit 3 of *Clear Speech*, “Word Stress and Vowel Length”). She was encouraged to use all the three acoustic properties (i.e., duration, pitch, and intensity) to mark the stressed syllable, and to pay attention to them when she made a speech in the daily recording. Then, she practiced modifying these properties in reading and spontaneous speech practice. The data on STFN-P/I indicated that J1 became able to sufficiently differentiate stressed vowels of content words and unstressed vowels of function words using pitch and intensity in the 4th to 6th month of the training (T2). The ability to differentiate stressed/unstressed vowels of content words using pitch and intensity substantially improved, but more slowly than that of content and function words. However, the ability to use duration to distinguish the stressed and unstressed vowels improved only slightly through the training periods.

The overall findings on J2 indicated that she was able to improve the ability to use duration, as well as pitch and intensity, through the training. However, it was indicated that the improvement of the ability to use pitch and intensity preceded that of duration.

An exploratory case study: improvement of the ability to produce the rhythmic properties of...

As was found in J1, J2 became able to sufficiently differentiate stressed vowels of content words and unstressed syllables of function words using pitch and intensity at T2. The ability to use pitch to differentiate stressed and unstressed vowels of content words became comparable to the NS mean a little later (i.e., T4), although that of intensity was already comparable to the NS mean in T1.

J2's data on the stress-related acoustic measures (STCN-D and STFND) indicated that she became able to modify the durational proportion of unstressed to stressed vowels in the 7th and 9th months (i.e., T3) for the contrast of content and function words, and in the 10th and 12th month (i.e., T4) for the contrast of content words. It should be noted that she learned about "de-emphasizing structure words" in the textbook (i.e., Unit7, H: De-emphasizing structure words: Reductions) in the 6th month. Then she was trained on reducing the function words in terms of duration, pitch, and intensity, through reading practice and spontaneous speech practice in the face-to-face training sessions. J2 was also encouraged to pay attention to reducing function words when making a spontaneous speech for "daily recording." This focused training continued until the 8th month (i.e., the end of T3). It appears that the training specifically focused on de-emphasizing function words led to J2's substantial improvement of the ability to shorten the vowels of function words. In the case of J1, the same unit was studied in the 8th month (i.e., the end of T3), but the intensive training focused on de-emphasis of function words was conducted only for a short period of time for a lack of practice time.

It was found that most of the fluency measures increased through the training period in both participants, indicating that the ability to spontaneously speak English with a greater speed (AR, SR), and to produce the greater number of words between pauses (MLoR) improved. The results indicated that, when individual-based training with practice using spontaneous production tasks is conducted for a prolonged length of time, it is possible to improve the fluency even when the learners are in the environment where the target language is not spoken as a native language. It is notable that the period of the substantial increase in the fluency in J2 (i.e., from T2 to T4) coincided with the period where the ability to modify duration of stressed and unstressed vowels greatly improved. This might suggest that, as the ability to produce a longer stretch of words and phrases improved, the ability to reduce unstressed syllables of function words improved as well. It might further be surmised that the ability to produce a phrase with a certain minimal length might be a necessary condition to learn to form a prosodic structure that contains a succession of function and content words. For example, if the learner has to pause after

a phrase, “I want to”, it is impossible to learn how to reduce the infinitive, “to.” It is not until the learner can produce a whole sentence (e.g., “I want to finish the job by tomorrow.”) that it is possible to learn how to reduce function words (e.g., “to”, “the”, “by”).

4-5. Teaching implications

Previous research suggested that it generally takes a long time for Japanese learners of English to modify the rhythmic properties of their production, especially duration, when a sentence reading task is used for data elicitation (Tsushima, 2015, 2016). The present results suggest that it takes even longer to improve such an ability in spontaneous production. When speaking spontaneously, speakers are required to utilize online resources to deal with lexical, syntactic, and discourse planning, in addition to articulatory planning and execution. Learners have to improve the ability to manipulate the acoustic properties of speech, as well as the ability to speak more fluently with respect to speed and the length of words between pauses. When the mean length of runs becomes longer, the learners will be faced with a more demanding task of building a prosodic structure of longer phrases with a larger number of feet, as well as coping with the other aspects of the online planning and execution. Therefore, if pronunciation instruction is included in a curriculum, it is recommended that spontaneous speech practice with a focus on the rhythmic properties of speech be conducted continuously throughout the period.

It is also recommended that the teaching materials include the following topics as are presented in the textbook used in the present study (i.e., *Clear Speech*); the notion of syllable, quality of stressed vowels, the notion of stress, the acoustic differences between stressed and unstressed syllables (including schwa), the notion of content and function words, and the notion of sentence focus. It may be noted that there are some other textbooks which emphasize instruction on the prosodic aspects of English including rhythm (e.g., Cook, 2000; Gilbert, 2012; Grant, 2010).

4-6. Concluding remarks

The overall results support the conclusion that long-term, individual based speech production training can significantly improve the ability of Japanese learners of English to produce the rhythmic properties of English in a spontaneous speech. It was indicated, however, that they may have more difficulty learning to modify duration of vowels than pitch and intensity, but that the ability to modify duration to differentiate stressed and unstressed vowels can be improved if speech training specifically focused on reducing

An exploratory case study: improvement of the ability to produce the rhythmic properties of...
vowels is conducted for a prolonged period of time.

Ongoing analysis of the data are currently underway to examine how the two participants further improves the ability to produce the rhythmic properties of English while interacting with native as well as nonnative speakers of English in the study abroad environment. The findings of the analysis will be reported in the near future.

Acknowledgement

The present research was supported by a research grant from Tokyo Keizai University in the academic year of 2017.

References

- Beckman, M. E. (1986). *Stress and Non-Stress Accent* (Vol. 7). Dordrecht, Riverton: Foris Publications.
- Boersma, P., & Weenink, D. (2014). Praat: doing phoetics by computer [Computer program]. Version 5.4, retrieved 4 October 2014 from <http://www.praat.org/>.
- Cook, A. (2000). *American Accent Training* (2nd ed.). Hauppauge, NY: Barron's.
- 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 Comunicacion Social, Santiago de Cuba. Retrieved from www.pholab.uzh.ch/static/volker/.../DellwoEtAl_2009_SICS.pdf
- Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48 (3), 393-410. doi: 10.1111/0023-8333.00047
- Gilbert, J. B. (2001). *Clear speech from the start: basic pronunciation and listening comprehension in North American English: student's book*. Cambridge, UK; New York: Cambridge University Press.
- 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.
- Grant, L. (2010). *Well Said: Pronunciation for Clear Communication* (3rd ed.). Boston: National Geographic Learning Cengage.
- 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.
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38 (2), 201-223. doi: 10.2307/3588378
- Ladefoged, P. (1975). *A Course in Phonetics* (2 ed.). San Diego, New York, Chicago, Washington,

- D.C., Atlanta, London, Sydney, Toronto: Harcourt Brace Jovanovich, Publishers.
- Li, A., & Post, B. (2014). L2 Acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition*, 36 (02), 223-255. doi: doi: 10.1017/S0272263113000752
- Machač, P., & Skarnitzl, R. (2009). *Principles of Phonetic Segmentation*. Praha: Epocha Publishing House.
- Mori, Y., Hori, T., & Erickson, D. (2014). Acoustic correlates of English rhythmic patterns for American versus Japanese speakers. *Phonetica*, 71 (2), 83-108.
- Nolan, F., & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica*, 66 (suppl 1-2) (1-2), 64-77.
- 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>
- 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. doi: doi: <http://dx.doi.org/10.1121/1.4923359>
- 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
- 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., Nespore, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73 (3), 265-292.
- 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.
- Tsushima, T. (2017). A pilot study on the rhythmic properties of English produced by Japanese learners before and after a five-month study abroad period. *Journal of Humanities & Natural Sciences* (141), 39-69.
- 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/Philadelphia: John Benjamins Publishing Company.
- 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>