

Development of fluency and English speech rhythm through individual-based, long-term speech training in the formal instruction settings

Teruaki Tsushima

ABSTRACT

The present study attempted to investigate how two adult Japanese learners of English (P1, P2) improved fluency and production of English speech rhythm through long-term, individual-based speech training under the FI (formal instruction) settings. The participants made recordings of spontaneous narratives (N=97) over the period of approximately one year. The analyses examined speed and breakdown fluency measures (e.g., articulation rate, mean length of runs, frequency and duration of between- and within-clause pauses), pitch and intensity ranges, rhythm indices (e.g., normalized variability coefficients and normalized pairwise variability of vowel duration, pitch, and intensity), and stress-related measures (e.g., the acoustic difference in duration, pitch, and intensity between stressed and unstressed vowels). The results found that, in both participants, most of the fluency measures made significant changes toward the NS (native speakers) means, indicating that they improved their ability to produce speech with a greater speed, a larger chunk of words, and fewer and shorter pauses. It was also found that the proportional difference of duration between stressed and unstressed vowels in content and mono-syllabic function words significantly became greater, indicating that they acquired a better control of duration in marking stressed syllables. Although the pitch range significantly increased through the training, both participants performed at the level comparable to that of NS in terms of the variability measures of pitch and differentiation of stressed and unstressed vowels in pitch at the outset of the training. In addition, the results did not show clear effects of the training on the use of intensity in either participant. Finally, the results indicated that both participants made accelerated improvements across most of the fluency measures and some

Development of fluency and English speech rhythm through individual-based, long-term speech ...

rhythm and stress-related measures in the middle of the training period. These results were interpreted in the framework of the L2 (second language) speech production model (e.g., Kormos, 2006). Implications for L2 pronunciation instruction were also discussed.

key words; L2 learning, speech rhythm, fluency, stress, rhythm indices

1. Introduction

1-1. Review of literature

One of the greatest challenges faced by foreign language instructors is to improve learners' pronunciation and fluency in an environment where the target language is not spoken. In the formal instruction (FI, henceforth) settings, learners generally lack the opportunities to use the target language for communicative interactions. In contrast, the study abroad (SA, henceforth) settings are generally believed to provide learners with an optimal learning environment where learners have a great deal of exposure to the target language, and ample opportunities to use it both inside and outside of classrooms. This may be one of the major reasons why many educational institutions provide students with multiple SA programs. The present case study focused on development of fluency and production of English speech rhythm in two adult Japanese learners during a one-year, individual-based speech training in the FI settings. The results were compared, where possible, with the results of the author's previous study in which development of the same abilities was examined before, during, and after a five-month SA program (Tsushima, 2019).

Previous research has shown that it is relatively difficult for Japanese learners of English to learn to produce prosodic characteristics of English speech rhythm (Mori, Erickson, Rilliard, & Hori, 2014; Mori, Hori, & Erickson, 2014). English is generally categorized as a "stress-timed" language while Japanese as a "mora-timed" language (e.g., Roach, 1982). One of the most important differences between the two languages is how stressed and unstressed syllables are acoustically realized. In English, a stressed syllable is acoustically marked by a combination of longer duration, higher pitch, and larger intensity, while in Japanese it is primarily marked by higher pitch only (Beckman, 1986). In addition, English has unstressed syllables which are not only substantially shorter in duration but also more centralized (i.e., schwa) in quality than stressed ones, while

unstressed syllables hardly occur in Japanese. These differences make the degree of durational variability across syllables substantially larger in English than in Japanese. Previous studies have demonstrated that the differences described above influence the acoustic characteristics of English produced by Japanese speakers of English. For example, it has been reported that the syllables produced by Japanese speakers tend to be equal in duration (Bond & Fokes, 1985; Mochizuki-Sudo & Kiritani, 1991), and that the duration of function words (e.g., “in”) produced by Japanese speakers is longer in duration than native speakers of English (Aoyama & Guion, 2007). It has also been found that Japanese learners of English have more difficulty learning to control duration than pitch and intensity in marking stressed syllables (Mori, Hori, et al., 2014; Tsushima, 2015). In the present study, the acoustic differences between stressed and unstressed vowels in terms of duration, pitch, and intensity were examined over the course of the speech training.

Another line of research has attempted to characterize the language-specific rhythmic differences by using the rhythm indices. According to Ramus, Nespors & Mehler (1999), for example, the degree of durational variability of speech segments such as ΔV (standard deviation of vocalic intervals) is higher in stress-timed languages including English than syllable-timed and mora-timed languages including Japanese (Ramus, Nespors, & Mehler, 1999). In addition, Grabe and Low (2002) proposed measuring the degree of durational variability in successive speech segments. For example, a normalized pairwise variability of vocalic intervals (nPVI-V) is higher in stress-timed languages than syllable-timed and mora-timed languages. Previous research has indicated that, although there exists some counter evidence (Barry, Andreeva, & Koreman, 2009; Gut, 2012; Turk & Shattuck-Hufnagel, 2013), the rhythm indices become closer to those of native speakers as the learners become more proficient (Gut, 2009; Li & Post, 2014; Ordin & Polyanskaya, 2014; White & Mattys, 2007). Li and Post (2014), for example, showed that nVarcoV (the speech-rate-normalized variation coefficient of vocalic intervals) and nPVI-V were significantly higher in the more advanced group of L1 Mandarin learners and L1 German learners of English than the less advanced ones. The present study examined how some rhythm indices (e.g., nPVI-V, nVorcoV) changed over the course of the speech training.

The effects of FI and SA on nonnative language development have been studied in longitudinal research in which the same learners consecutively participate in FI and SA (e.g., Pérez-Vidal, 2014). Overall, the existing evidence remains equivocal about the effects of either FI or SA across different domains of phonological ability (see Pérez-Vidal 2014,

Development of fluency and English speech rhythm through individual-based, long-term speech ... for a comprehensive review). With regard to segmental production, some studies reported non-significant effect of FI before SA (Mora, 2008), while others found otherwise (e.g., Pérez-Vidal, Juan-Garau, & Mora, 2011). The effect of SA on segmental production was found to be non-significant in some studies (e.g., Avello & Lara, 2014 on vowel duration, quantity, and VOT), while it was found otherwise when some phonological processes were investigated (e.g., Avello, Mora, & Pérez-Vidal, 2012 on insertions, deletions, and substitutions). Finally, Lord (2010) suggested that a combination of FI and SA is most effective for development of segmental production as compared with FI or SA alone. Previous research which investigated the effects of FI and SA on production of prosody remains inconclusive as well. For example, Valls-Ferrer (2011) examined how English speech rhythm and fluency improved among Catalan/Spanish or Basque/Spanish learners of English during FI (six months) and SA (three months). It was found that both the rhythm and fluency significantly improved during SA, but not during FI. Gut (2009), however, found that SA was not effective in improving English rhythm in terms of a syllable ratio of stressed/unstressed syllables among German learners of English before and after nine-months of SA. The overall results may be partly explained by the learners' focus on development of fluency during SA. Avello and Lara (2014) states, "... in the absence of specific pronunciation instruction while abroad, they [the learners] could have focused on the development of other skills perceived to be more important than phonological accuracy in the task of achieving fluent communication in the L2 (p. 158)." This interpretation may suggest that instruction specifically targeted at pronunciation is necessary even when L2 learners have a great deal of opportunities to engage in communicative interactions with native and nonnative speakers outside of classrooms.

Another aspect of nonnative speech development which has attracted a great deal of attention is fluency. Previous research on fluency has been conducted in the theoretical framework of a modular model of speech production in L1 (the first language) proposed by Levelt (1989), and later adapted to L2 (the second language) speech production by Kormos (2006). In these models, speech production consists of three stages:

- 1) Conceptualization stage: a speaker generates a desired concept or intention to be expressed, and formulates an intended message called a preverbal plan.
- 2) Formulation stage: the preverbal plan will go through a series of processes, including lexical access/retrieval and syntactic/morpho-phonological encoding processes, which results in a phrase structure. The phrase structure will then receive phonetic and prosodic encoding, then produces a phonetic plan, which

consists of articulatory scores which specify how articulatory gestures are activated.

- 3) Articulation stage: the articulatory scores are executed through articulatory movements of the vocal apparatus (e.g., tongue, lips, jaw, larynx), which produces speech.

For proficient L2 speakers, at least some encoding processes in the formulation stage are assumed to be automatized and operate simultaneously with the other modules (i.e., parallel processing), although the conceptualization stage may require a certain degree of attention. For less proficient L2 speakers, on the other hand, each stage is expected to require relatively large attentional resources, which makes it difficult for each module to operate simultaneously with the other modules (i.e., serial processing). Especially, the formulation stage may require a great deal of attentional resources due to less efficient lexical access and syntactic and morpho-phonological encoding (Kormos, 2006). These differences are expected to result in more fluency breakdowns in less proficient L2 speakers.

According to Segalowitz (2010, pp.46-52), fluency consists of three interrelated domains; “*Cognitive fluency* refers to the efficiency of the speaker’s underlying processes responsible for fluency-relevant features of utterances, *utterance fluency* refers to the oral features of utterances that reflect the operation of underlying cognitive processes, and *perceived fluency* refers to the inferences that listeners make about a speaker’s cognitive fluency based on perception of the utterance fluency features of the speaker’s speech output (p.50).” Previous research has used utterance fluency as data to make inferences about cognitive fluency postulated in the speech production model (e.g., Lambert, Aubrey, & Leeming, 2020; Lambert, Kormos, & Minn, 2017). The utterance fluency consists of three major components: speed fluency (e.g., articulation rate), breakdown fluency (e.g., pause frequency), and repair fluency (e.g., self-repairs). Although the relation between the domains of the utterance fluency and the stages of the cognitive fluency may not be clear-cut, it has been suggested that between-clause pauses are primarily associated with the conceptualization stage (i.e., concept generation and content planning), while within-clause pauses with the formulation stage (i.e., lexical access and syntactic/morpho-phonological coding), and that the speed fluency is associated with all the stages above (Kormos, 2006; Lambert et al., 2020; Lambert et al., 2017).

Previous research on the effects of SA on development of fluency has mostly supported the conclusion that SA, but not FI, has significant effects on improvement of

Development of fluency and English speech rhythm through individual-based, long-term speech ... fluency (Freed, 1995; Lennon, 1990; Towell, Hawkins, & Bazergui, 1996; Valls-Ferrer & Mora, 2014). For example, Valls-Ferrer & Mora (2014) examined development of the utterance fluency among Catalan/Spanish learners of English during 6 months of FI, followed by 3 months of SA, and another 6 months of FI. It was found that four out of six fluency measures (i.e., speech rate, mean length of runs, phonation-time ratio, and pause duration ratio) significantly improved during SA, but not in FI before SA. They stated, "For many learners, the SA context provides the kind of intensive and extensive practice necessary for the automatization of proceduralized knowledge which speeds up lexical access and retrieval and makes phonological, morphological and syntactic encoding fast and efficient, ... (p. 131)."

A combination of the results of the previous research on the effects of FI and SA on phonological and fluency development appear to suggest that it is not adequate to provide the learners with rich input of the target language to improve pronunciation abilities. The observed significant improvement of fluency during SA might suggest that learners may potentially have greater resources allocated to pronunciation. However, it seems necessary for them to consciously pay attention to segmental accuracy and rhythmic properties while speaking (e.g., phonological /phonetic encoding and articulation). In this sense, specific pronunciation instruction or training maybe desirable or necessary to significantly improve learner's pronunciation in FI and SA. As a matter of fact, previous research on the nonnative speech training has found that such training is generally effective in significantly improving learners' nonnative production abilities, given adequate amount of time (cf., Saito, 2012). It is also suggested that training on the prosodic level (e.g., intonation, rhythm, stress) is important as well as that on the segmental level (e.g., vowels and consonants), especially to improve pronunciation in spontaneous speech (Celce-Murcia, Brinton, & Goodwin, 2010; Fraser, 2001). Finally, previous research suggested that it is more beneficial to use pronunciation training which involves the communicative framework to improve speech intelligibility as compared with training which uses decontextualized speech materials (e.g., Saito, 2012).

The inconsistent results of the effects of FI and SA on phonological development may also be due to the fact that the research is mostly based on averaged group data. In the longitudinal research, individual variability is usually the case, and not the exception. The individual variability may manifest itself in a number of aspects of development. It includes rate of development, the timing of a spurt of development, initial ability, and temporary regression. The averaging might have missed some important individual

developmental patterns which could have been captured if the data had been analyzed on individual basis.

Given these considerations, Tsushima (2019) conducted a longitudinal case study designed to examine development of fluency and production of English speech rhythm among two Japanese learners of English (J1, J2) enrolled in an academic program with a five-month SA. It analyzed spontaneously produced narratives (N=184 for J1 and N=139 for J2) recorded over a period of 24 months for J1 and 20 months for J2, which included the period of FI plus individual speech training before SA (PreTr), SA, and FI plus individual speech training after SA (PostTr). For measurements of English speech rhythm, two sets of measures were used. First, one set (termed “rhythm measures”) included normalized pairwise variability of vowels in duration, pitch, and intensity (nPVI-V-D/P/I), and variation coefficient of duration, pitch, and intensity (VarcoV-D/P/I). Another set of measures (termed “stress-related measures”) measured the acoustic differences between stressed and unstressed vowels in content words in terms of duration, pitch and intensity (STCN-D/P/I), and between stressed vowels in content words and monosyllabic function words in terms of duration, pitch and intensity (STFN-D/P/I). For measurements of fluency, articulation rate (AR), speech rate (SR), mean length of runs (MLoR), and pause ration (PauseRat) were used. It was found that PreTr substantially improved a number of rhythm and fluency measures in both participants, although the levels of attainment in some measures was not comparable to those of the NS controls. During SA, however, J2 showed accelerated improvement in most of the rhythm and fluency measures, strongly suggesting that the amount of interaction with native and nonnative speakers in the SA setting allowed her to improve the rhythmic properties of her speech and the degree of fluency. On the other hand, J1 showed little improvement in either rhythm or fluency measures during SA but showed substantial improvement in both measures in PostTr. Given a relatively long period of plateau before and during SA, the observed “spurt” in the rhythm and fluency measures in PostTr might be interpreted as suggesting that the input and output experiences during SA formed a foundation on which J1 was able to improve the production abilities through the speech training following SA.

1-2. Rationale

One of the most important limitations of Tsushima (2019) described just above was that it lacked control data where participants received FI plus individual speech training

Development of fluency and English speech rhythm through individual-based, long-term speech ... but did not participate in SA. The present study was designed to investigate how the participants who received individual speech training in the FI settings were able to improve fluency and production of English speech rhythm. Following Tsushima (2019), the study was conducted in the form of a case study where the development of a small number of participants (N=2) was traced for a relatively long period of time (approximately 12 months). Both of them (called P1 and P2 henceforth) belonged to an academic English program for selected students with relatively advanced English abilities. In contrast to a longitudinal study with group data which typically acquire data a few to several times at most, the present case study obtained data 97 times for each of the participants during the training period, which allowed for closer examination on the shape of improvement. As described in more detail below, the procedure of the individual speech training and of data acquisition (i.e., recording of spontaneous narratives) was basically the same as used in Tsushima (2019).

1-3. Specific Research Questions

The following were the specific questions asked in the present study.

- 1) How did the fluency measures change during the training period?
- 2) How did the pitch range and intensity range change during the training period?
- 3) How did the rhythm measures change during the training period?
- 4) How did the stress-related acoustic measures change during the training period?

2. Method

2-1. Participants

The participants (P1, P2) were two female university students who joined the study in the fall of their sophomore year. They had joined the advanced English program at the beginning of the academic year (i.e., April). The program offers two years of English instruction twice a week (except during long-term vacations). One class is taught by a Japanese instructor and focuses on reading, writing, and grammar, while the other is taught by an NS instructor and focuses on English speaking skills. Both of the participants were very motivated to study English and especially improve their pronunciation and speaking skills. Based on their TOEIC scores, the CEFR level of both participants was upper A2 at the beginning of the training period and middle B1 at the end. These levels were comparable to those of the two participants (J1, J2) in Tsushima (2019).

2-2. Data acquisition

The data consisted of spontaneous narratives recorded by the participants. They were asked to make an approximately one-minute recording every day, but at least eight times a month. P1 and P2 made a total of 97 recordings each over the period of 12 months from the beginning of September in their sophomore year to the end of August in their junior year (i.e., approximately eight times a month). The length of each recording varied from approximately 50 seconds to a few minutes. The topic included daily events, class activities, memories of a trip, hobbies, an opinion on a familiar issue and others. Recordings were mostly done in a quiet environment at their homes using an iPhone with a high-quality microphone (*Zoom IQ7*) with a pop filter attached to it. Sound recording software (*Zoom Handy Recorder App*) was used with a sampling rate of 44,000 Hz and 16 bits of resolution. The sound file was saved in a wav format and sent to the author via email. It was then low-pass filtered (under 8,000 Hz) and normalized for average intensity (70 dB) on sound analysis software, *Praat* (Boersma & Weenink, 2014). When they made a recording during a training session at the author's office, the recordings were made using a high-quality microphone (*Audio-Technica AT4040*) in a quiet environment or in a sound-treated recording room.

Data acquisition from native speakers of English (NS, henceforth) was conducted at a university in Edinburgh, England. The participants (N=13) were native speakers of English from England or North America and were English instructors at the university. They were asked to talk about their life events for approximately one minute. Another set of data was obtained from two female Japanese speakers in their freshman year who belonged to a general English communication class. They engaged in communication activities mostly with their Japanese peers using a textbook twice a week (90 minutes each). Based on their TOEIC scores, the CEFR level of both participants was around middle A2. They made weekly recordings ten times per semester at the author's office with the instruments described above. Averaged data obtained in the first semester (N=20) were used as the Japanese control (JCtr, henceforth) data.

2-3. Acoustic data analyses

Following standard segmentation criteria (Payne, Post, Astruc, Prieto, & Vanrell, 2012; Peterson & Lehiste, 1960) as much as possible, the sound waves were segmented into a consonant and vowel portion through visual inspection of the wave forms and wideband spectrograms, and perceptual confirmation of the segmented portions on *Praat*.

Development of fluency and English speech rhythm through individual-based, long-term speech ...

Using *Praat* scripts, vowel durations, *f0* and *dB* at the mid-point of a vowel portion were measured. Following Tsushima (2019), vowels equal to or longer than 300 ms were excluded from the analyses. Altogether, 8,713 and 9,266 vowels were analyzed for P1 and P2, respectively.

2-4. Fluency measures

The present study used a set of speed fluency measures (i.e., speech rate (SR); articulation rate (AR); mean length of runs (MLoR)), and a set of breakdown measures (i.e., pause duration rate (PauseRat); frequency/duration of between-clause pauses (PauseFreqBC, PauseDurBC); frequency/duration of within-clause pauses (PauseFreqWC, PauseDurWC)). Following Valls-Ferrer & Mora (2014: p.120), a pause was defined as a silent interval equal to or longer than 300 ms.

- 1) SR (Speech Rate) : the total number of words produced in a narrative divided by the amount of total time required to produce it (including pause time) expressed in minutes.
- 2) AR (Articulation Rate) : the total number of words produced in a narrative divided by the amount of time taken to produce it (excluding pause time) expressed in minutes.
- 3) MLoR (Mean Length of Runs) : the average number of words produced in utterances between pauses of 300 ms and above.
- 4) PauseRat (Pause Duration Ratio) : the total length of pauses divided by the total amount of speaking time (including pause time).
- 5) PauseFreqBC (Frequency of Between-Clause Pauses per 100 words) : the number of between-clause pauses divided by the number of words times 100.
- 6) PauseDurBC (Duration of Between-Clause Pauses) : the average length of between-clause pauses in seconds.
- 7) PauseFreqWC (Frequency of Within-Clause Pauses per 100 words) : the number of within-clause pauses divided by the number of words times 100.
- 8) PauseDurWC (Duration of Within-Clause Pauses) : the average length of within-clause pauses in seconds.

2-5. Pitch range and intensity range

For the pitch range (RNG-P) and the intensity range (RNG-I), an absolute difference between the minimal and maximum value of *mel* and *dB* among all the vowels in a

sentence was calculated. Then, the values were averaged over all the sentences in an entire narrative.

2-6. Rhythm measures

2-6-1. Normalized pairwise variability indices (nPVI-V-D/P/I)

These indices indicate local variability of the prosodic acoustic properties (i.e., duration, pitch, and intensity). Specifically, they measure the amount of difference in the acoustic property between a pair of adjacent vowels, averaged over all the pairs in the whole narrative. In order to minimize potential effects of sentence-final lengthening and pitch/intensity changes, a pair of vowels which included the last vowel in a sentence were excluded from analyses. In addition, a pair of vowels which straddled a syntactically determined sentence boundary or a pause (i.e., a silent period of 300 ms or longer) were excluded.

- 1) nPVI-V-D: a normalized pairwise variability index of vowel in duration: the absolute durational difference of a pair of adjacent vowels divided by the mean duration of both vowels, averaged over all the pairs in the passage, and multiplied by 100.
- 2) nPVI-V-P: a normalized pairwise variability index of vowel in pitch: the pitch difference in *mel* of a pair of adjacent vowels, averaged over all the pairs in the passage.
- 3) nPVI-V-I: a normalized pairwise variability index of vowel in intensity: the intensity difference in *dB* of a pair of adjacent vowels, averaged over all the pairs in the passage.

2-6-2. Normalized variability coefficient indices (nVarcoV-D/P/I)

These indices indicate global variability of the prosodic acoustic properties. Specifically, they measure the amount of variability among all the vowels in the narrative.

- 1) nVarco-V-D: the standard deviation of vowel durations averaged over all the vowels in the narrative, divided by the mean.
- 2) nVarco-V-P: the standard deviation of *mel* values averaged over all the vowels in the narrative, divided by the mean.
- 3) nVarco-V-I: the standard deviation of *dB* values averaged over all the vowels in the narrative, divided by the mean.

2-6-3. Stress-related measures

These measures indicate the degree of difference in the prosodic acoustic properties between stressed and unstressed syllables. The following codes were assigned to each syllable in the narrative; (1) stressed syllable of a content word (e.g., “prob” in problem), (2) unstressed syllable of a content word (e.g., “lem” in problem), and (3) a syllable of a monosyllabic function word (e.g., in). A monosyllabic content word (e.g., “went”, “home”) was assigned either (1) or (2), depending on the prosodic/informational structure of the phrase or sentence. Monosyllabic function words included prepositions (e.g., at), articles (e.g., the), to-infinitives, and be-verbs (e.g., is). The other categories, including pronouns, auxiliary verbs, conjunctions were excluded from the analyses.

- 1) STCN-D: the average duration of all the unstressed vowels in content words (e.g., “tor” in “factor”) divided by that of stressed vowels (e.g., “fac” in “factor”), multiplied by 100.
- 2) STCN-P: the average *mel* of all the unstressed vowels in content words subtracted from that of stressed vowels.
- 3) STCN-I: the average *dB* of all the unstressed vowels in content words subtracted from that of stressed vowels.
- 4) STFN-D: the average duration of all the vowels of monosyllabic function words (e.g., “in”) divided by that of stressed vowels in content words (e.g., “fac” in “factor”), multiplied by 100.
- 5) STFN-P: the average *mel* of all the vowels of monosyllabic function words subtracted from that of stressed vowels in content words.
- 6) STFN-I: the average *dB* of all the vowels of monosyllabic function words subtracted from that of stressed vowels in content words.

2-7. Data analyses

The period of the training (i.e., 12 months) was divided into four sub-periods (i.e., T1, T2, T3, and T4). Each sub-period (i.e., 3 months) contained approximately 24 recorded narratives. Non-parametric *t*-tests (i.e., Mann-Whitney *U* tests) were conducted, when necessary, to examine whether the observed values of a particular variable (e.g., SR) in one sub-period were significantly different from those of the other sub-periods or those of NS.

2-8. Speech training

For both participants, individual speech training was conducted by the author once a week for two semesters (except during semester breaks). The length of one training session varied from 45 minutes to 90 minutes. P1 and P2 had 32 and 27 sessions, respectively. The training followed Unit 1 through Unit 8 of a pronunciation textbook, *Clear Speech* (Gilbert, 2012). The textbook allows the learners to gain knowledge about syllable, English vowels, acoustic differences between stressed and unstressed syllables, stress rules, content/structure words, emphasis on peak vowels in focus words, and deemphasis of structure words. It also provides them with a lot of practice to learn the skills related to the topics above. The training session included supplementary activities such as reading aloud, retelling a story, and making a short speech, describing a set of toy items. In all of these activities, a focus was given to improve the learners' ability to pay attention to the prosodic properties of English while improving fluency. In addition to this, the author gave feedback on each of the recorded narratives such that the participants could learn how to modify their pronunciation.

3. Results

3-1. Fluency measures

Figure 1 shows how SR (i.e., speech rate), which reflects both speed and breakdown fluency, changed across the sub-periods. It is shown that, in both participants (i.e., P1 and P2), SR increased substantially from T1 through T4, and that the rate of increment was relatively large between T2 and T3. It should be noted, however, that SR of NS was substantially higher than those of T4, indicating that there was a lot more room for improvement even at the end of the training (see Table 1). It should also be noted that the highest SR (i.e., 111.4) shown among J1 and J2 in Tsushima (2019) was almost the same as those shown by P1 and P2 at T4, indicating that they reached almost the same level of fluency as those who participated in SA.

Table 1 shows means (M) and standard deviations (SD) of all the fluency measures across the sub-periods in P1 and P2, as well as those of NS and the Japanese controls (JCtr, henceforth). First of all, it should be noted that all the fluency measures of P1 and P2 at T1 were intermediate between those of JCtr and those of NS. This was not surprising as P1 and P2 had started the training in the second semester of their sophomore year and had already been in the advanced English program for half a year,

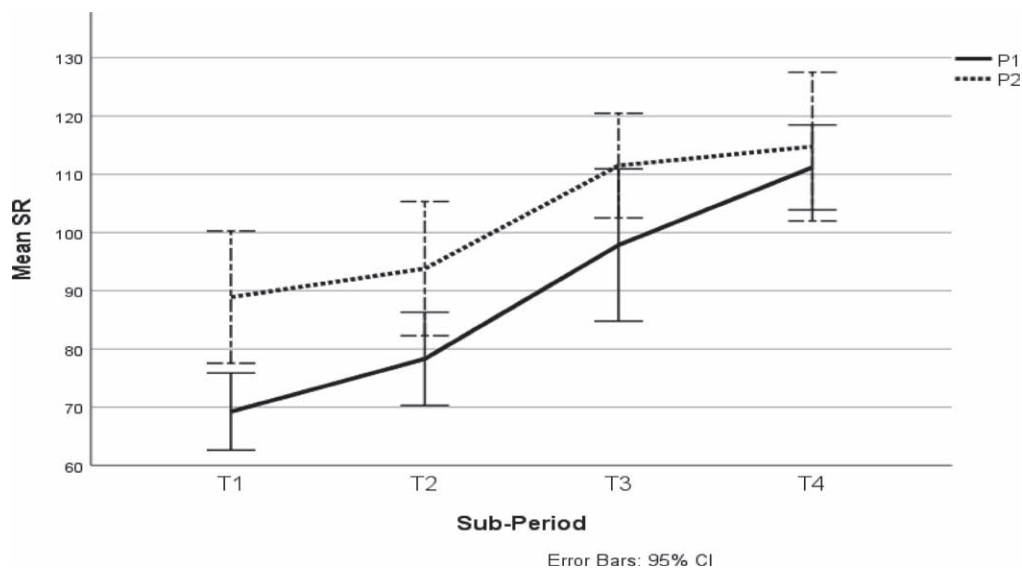


Figure 1. Mean (M) and one standard deviation (SD) of SR (Speech Rate) averaged over narratives recorded within each of the sub-periods. P1/P2=Japanese participants.

while JCtr' data were obtained in the first semester of their freshman year. Starting from T1, most of the fluency measures of P1 and P2 became substantially closer to those of NS, indicating that their overall fluency improved a great deal through the training. As shown in Figure 1, a number of the measures showed a relatively greater improvement between T2 and T3 than the other pairs of the sub-periods. AR (i.e., speed fluency), which reflects the speed of speaking without regard to pauses, increased between T2 and T3 ($p=.000$), and continued to increase through T4 in P1, while it showed a substantial increase only between T2 and T3 in P2 ($p=.000$). MLoR (i.e., speed fluency), which reflects the length of word chunks bounded by pauses, also increased significantly between T2 and T3 in both participants ($p=.000$). The result indicated that, in these sub-periods, the participants became able to speak faster and with substantially larger chunks of words without pausing. P1 reached the same level as the highest value recorded among J1 and J2 in AR (i.e., 147.7) but fell short in MLoR (J2=5.9), while P2 reached comparable levels in both measures.

PauseRat (i.e., breakdown fluency), which reflects the amount of pausing, showed substantial decrease from T1 to T4 in both participants. However, the rate of decrease was much smaller between T3 and T4. Both participants reached the same level as the lowest value recorded among J1 and J2 (i.e., 25.0). PauseFreqBC and PauseDurBC (i.e.,

Participant	Measure		T1	T2	T3	T4	NS	JCtrl
P1	SR	<i>M</i>	69.2	78.3	97.8	111.7	172.3	42.3
		<i>SD</i>	6.6	8.0	13.1	7.0	23.9	8.7
	AR	<i>M</i>	129.0	127.6	136.2	152.6	197.4	122.4
		<i>SD</i>	9.7	10.5	11.5	7.2	24.9	10.1
	MLoR	<i>M</i>	2.8	3.1	4.8	4.9	15.7	2.0
		<i>SD</i>	0.5	0.4	1.0	0.5	13.4	0.2
	PauseRat	<i>M</i>	46.1	38.7	27.8	26.8	13.0	61.6
		<i>SD</i>	5.5	3.3	5.5	3.1	6.0	10.7
	PauseFreqBC	<i>M</i>	21.8	21.8	17.7	17.5	4.8	
		<i>SD</i>	4.0	3.1	2.8	2.1	2.5	
	PauseDurBC	<i>M</i>	1.49	1.12	0.91	0.78	0.56	
		<i>SD</i>	0.32	0.17	0.14	0.12	0.24	
	PauseFreqWC	<i>M</i>	12.8	9.9	3.3	2.1	2.6	
		<i>SD</i>	6.2	3.8	2.8	1.3	1.8	
	PauseDurWC	<i>M</i>	0.68	0.60	0.55	0.52	0.60	
		<i>SD</i>	0.22	0.16	0.18	0.26	0.17	
		<i>N</i>	22	24	24	27	13	20
	P2	SR	<i>M</i>	88.9	93.8	111.5	114.8	172.3
<i>SD</i>			11.4	11.5	9.0	12.8	23.9	8.7
AR		<i>M</i>	136.4	134.4	145.1	145.5	197.4	122.4
		<i>SD</i>	9.9	11.3	9.7	12.7	24.9	10.1
MLoR		<i>M</i>	4.1	4.5	5.6	6.0	15.7	2.0
		<i>SD</i>	0.9	0.8	0.7	1.2	13.4	0.2
PauseRat		<i>M</i>	34.7	30.3	23.2	20.7	13.0	61.6
		<i>SD</i>	6.9	4.8	2.8	3.6	6.0	10.7
PauseFreqBC		<i>M</i>	18.9	18.2	14.5	14.3	4.8	
		<i>SD</i>	3.9	3.1	2.3	2.5	2.5	
PauseDurBC		<i>M</i>	1.05	0.97	0.78	0.72	0.56	
		<i>SD</i>	0.18	0.19	0.12	0.11	0.24	
PauseFreqWC		<i>M</i>	5.7	3.5	2.5	1.9	2.6	
		<i>SD</i>	4.4	3.0	1.7	1.5	1.8	
PauseDurWC		<i>M</i>	0.77	0.66	0.57	0.55	0.60	
		<i>SD</i>	0.28	0.24	0.20	0.15	0.17	
		<i>N</i>	26	24	24	23	13	20

Table 1. Mean (*M*) and one standard deviation (*SD*) of fluency measures averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS= native speakers of English; JCtrl=Japanese controls; SR=speech rate; AR=articulation rate; MLoR=mean length of runs; PauseRat; pause ratio; PauseFreqBC=frequency of between-clause pauses; PauseDurBC=duration of between-clause pauses; PauseFreqWC=frequency of within-clause pauses; PauseDurWC=duration of within-clause pauses.

frequency and duration of between-clause pauses) significantly decreased from T1 through T4 ($p=.000$), while the decrease was pronounced between T2 and T3 in the frequency measure in both participants. PauseFreqWC and PauseDurWC (i.e., frequency and duration of within-clause pauses) also decreased significantly between T1 and T2 in

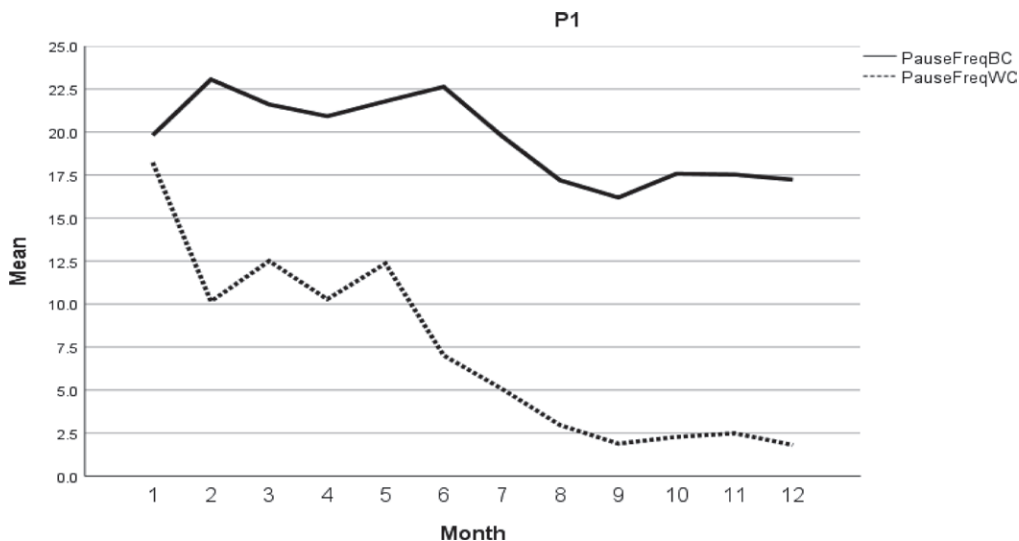


Figure 2. Mean (M) of PauseFreqBC (frequency of between-clause pauses) and PauseFreqWC (frequency of within-clause pauses), averaged over narratives recorded within each month in P1.

both participants. In P1, both of these measures were not significantly different from those of NS at T4. In P2, they were not significantly different from those of NS even at T1. Thus, the difference between the participants and NS at the end of the training was largely due to that of frequency and duration of between-clause pauses rather than those of within-clause pauses.

To further investigate how the frequency of between- and within-clause pauses changed, the data are plotted as a function of months in Figure 2 (P1) and Figure 3 (P2). As shown in Figure 2, PauseFreqBC and PauseFreqWC in P1 decreased substantially from the 6th to 9th month, and 5th to 9th month, respectively, which corresponds to the sub-periods of T2 and T3. In P2, a substantial drop in PauseFreqBC was observed between the 5th and 7th month, while the magnitude of the drop was much less in PauseFreqWC. It should also be noted that, in both participants, a substantial decrease in PauseFreqWC was evidenced in the first two months of the training, indicating that both participants were able to substantially reduce the frequency of within-clause pauses during the beginning of the training period. It was also found that both participants were able to reduce the frequency of between-clause pauses in a certain limited period in the middle of the training period, and that P1 was able to reduce the frequency of within-clause pauses almost simultaneously with that of between-clause pauses.

The following examples illustrate how the between- and within-clause pauses

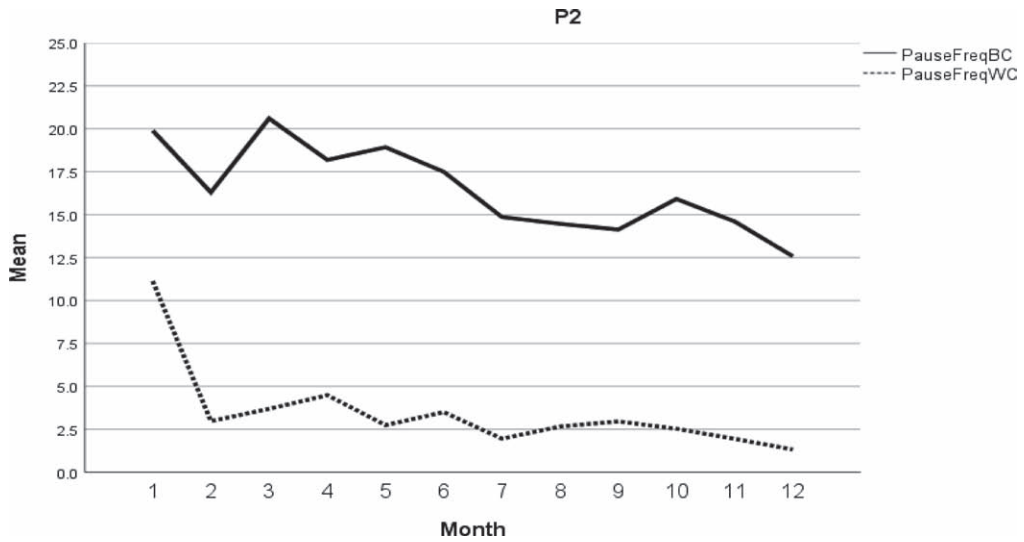


Figure 3. Mean (M) of PauseFreqBC (frequency of between-clause pauses) and PauseFreqWC (frequency of within-clause pauses), averaged over narratives recorded within each month in P2.

decreased across the training. The sentences, (P1S1), (P1S2), and (P1S3), were drawn from P1’s narratives recorded in the 1st, 5th, and 9th month. “///” indicates a between-clause pause, while “//”, a within-clause pause. A value in the parenthesis shows the duration in seconds.

(P1S1) I think /// (1.58) the restaurant uses good stuff /// (6.34) it was // (.327) very (.902) // nice point.

(P1S2) I had to submit the application to get // (.540) library card at CSC /// (1.64) because I could pass the qualification. ... I wrote paper /// (.856) it was about // (.456) foreign history.

(P1S3) I usually feel /// (.555) it’s very tough to earn money /// (1.98) so we will not waste money.

At the initial stage of the training (as shown in P1S1), the between-clause pause was typically long, and even a structurally simple clause (e.g., “it was very nice point”) may include multiple within-clause pauses. In the 5th month (P1S2), the speaker was able to produce a certain length of phrase without pauses (e.g., “had to submit the application”, “could pass the qualification”). However, within-clause pauses still occur occasionally, typically when the speaker had difficulty with lexical access and retrieval (e.g., “library card”, “foreign history”). In the 9th month (P1S3), within-clause pauses became much less frequent, as the speaker became able to produce a clause-size chunk of speech without

Development of fluency and English speech rhythm through individual-based, long-term speech ... pauses.

In sum, it was found that both participants became significantly better at speaking faster (i.e., AR) with larger chunks (i.e., MLoR) and fewer and shorter pauses (i.e., PauseFreqBC/WC and PauseDurBC/WC) through the training. It was also found that they showed marked improvement in a number of fluency measures between T2 and T3, and that they did not reach the NS level on all the frequency measures at the end of the training, except for the frequency and duration of within-clause pauses. Finally, both participants reached the same level as those (i.e., J1 and J2) who went to SA in Tsushima (2019).

3-2. Pitch and intensity range

As shown in Table 2, the pitch and intensity range were higher in P1 and P2 than those of JCtr at T1, as expected. The pitch range became significantly wider to approximate the NS mean through the training period in both participants. In P1, it increased substantially from T2 through T4. In P2, it increased substantially between T1 and T2, and approximated the NS mean at T4. In either participant, the RNG-P is not significantly different from that of NS at T4. It should be noted, however, that the standard deviation (*SD*) of NS is large, indicating that there was a great deal of variability among NS in the degree to which they varied pitch in their speech. With regard to the intensity range, P1 showed little improvement over the sub-periods. Although P2 showed a significant increase in RNG-I between T1 and T2 ($p < .001$), little change was observed

Participant	Measure		T1	T2	T3	T4	NS	JCtrl
P1	RNG-P	<i>M</i>	88.2	90.7	108.2	141.4	133.8	54.5
		<i>SD</i>	16.9	22.3	23.8	19.0	48.4	7.6
	RNG-I	<i>M</i>	10.2	9.9	10.7	10.6	18.4	2.8
		<i>SD</i>	2.2	2.5	5.8	1.3	2.8	0.6
		<i>N</i>	22	24	24	27	13	20
				T1	T2	T3	T4	NS
P2	RNG-P	<i>M</i>	85.6	107.5	114.0	123.5	133.8	54.5
		<i>SD</i>	19.6	15.5	13.0	16.7	48.4	7.6
	RNG-I	<i>M</i>	7.3	8.4	8.8	8.7	18.4	2.8
		<i>SD</i>	1.5	1.1	1.3	1.2	2.8	0.6
		<i>N</i>	26	24	24	23	13	20

Table 2. Mean (*M*) and one standard deviation (*SD*) of the pitch range and intensity range averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; JCtrl=Japanese controls; RNG-P=pitch range (in *mel*) ; RNG-I=intensity range (in *dB*).

from T2 to T4. In both participants, RNG-I was significantly lower than that of NS. As compared with J1 and J2, RNG-P in both participants were substantially higher than the highest value among J1 and J2 (i.e., 98.4). RNG at T4 was at the same level as the highest value among J1 and J2 (i.e., 10.4) for P1 and a little lower in P2. The overall results indicated that, through the training, both participants became able to make use of a range of pitch to a degree not different from that of NS. On the other hand, the training had negligible effects on widening the range of intensity.

3-3. Rhythm indices

Table 3 shows normalized pairwise variability indices (i.e., nPVI-V-D/P/I) and normalized variability coefficient indices (i.e., nVarcoV-D/P/I) as a function of sub-periods in both participants. These indices measure the degree of variability in the acoustic properties of all the syllables in the speech sample, regardless of the parts of speech and the stress status (i.e., stressed or unstressed). As mentioned above, the degree of variability in these acoustic properties was expected to be lower in Japanese speakers of English than those of NS. In terms of the durational indices, nVarcoV-D significantly increased between T1 and T4, while it was significantly different from that of NS at T4 in both participants ($p < .01$). On the other hand, nPVI-V-D did not show significant improvement between T1 and T4 in either participant. As a matter of fact, nPVI-D-V of both participants were relatively high at T1 as compared with J1 (47.5) and J2 (47.4) at the beginning of the training session reported in Tsushima (2019). A closer inspection of the first eight narratives revealed that, in P1, nPVI-D-V was around 40.0 in three narratives while it was unusually high (i.e., 61.3) in the first narrative largely due to dysfluency. In P2, nPVI-D-V of the first three narratives were lower than 43, but quickly increased to over 50 in the fourth narrative. The data indicated that both participants had difficulty producing the pairwise durational variability comparable to that of more proficient L2 speakers and NS at least at the very beginning of the training. Finally, both participants reached the same level of the highest values of these indices among J1 and J2 (i.e., nPVI-D-V=55.4, nVarcoV-D=50.4).

As for pitch indices, both nPVI-V-P and VarcoV-P increased significantly from T1 to T4 in both P1 and P2 ($p < 0.01$), indicating that they became able to use a greater degree of pitch difference in adjacent vowels as well as among overall vowels. In both participants, however, nPVI-V-P was actually higher than the NS mean while those of VarcoV-P were not significantly different from the NS mean. It should be noted that NS

Participant	Measure		T1	T2	T3	T4	NS	JCtrl
P1	nPVI-V-D	<i>M</i>	50.4	56.9	56.7	53.9	58.4	48.4
		<i>SD</i>	8.8	6.2	6.7	6.3	9.7	6.3
	nVarcoV-D	<i>M</i>	44.2	49.2	49.4	48.7	57.8	45.4
		<i>SD</i>	5.0	4.3	5.1	3.8	6.7	5.8
	nPVI-V-P	<i>M</i>	38.7	38.2	40.0	46.0	22.3	22.0
		<i>SD</i>	7.7	9.9	8.5	6.4	8.3	4.6
	nVarcoV-P	<i>M</i>	33.7	35.2	40.9	48.8	36.1	21.1
		<i>SD</i>	6.8	8.1	8.7	6.2	12.2	2.9
	nPVI-V-I	<i>M</i>	4.3	4.3	4.3	3.7	4.1	2.8
		<i>SD</i>	0.7	0.9	2.1	0.5	0.9	0.6
	nVarcoV-I	<i>M</i>	3.9	3.9	5.0	4.0	5.3	2.7
		<i>SD</i>	0.6	0.7	7.1	0.5	1.0	0.3
		<i>N</i>	22	24	24	27	13	20
				T1	T2	T3	T4	NS
P2	nPVI-V-D	<i>M</i>	51.1	52.0	55.3	53.2	58.4	48.4
		<i>SD</i>	6.8	5.9	5.7	5.6	9.7	6.3
	nVarcoV-D	<i>M</i>	46.1	48.3	51.0	50.1	57.8	45.4
		<i>SD</i>	4.6	3.3	4.0	3.3	6.7	5.8
	nPVI-V-P	<i>M</i>	34.5	42.7	42.3	46.0	22.3	22.0
		<i>SD</i>	6.3	5.8	5.2	6.0	8.3	4.6
	nVarcoV-P	<i>M</i>	34.4	41.5	41.0	44.7	36.1	21.1
		<i>SD</i>	6.6	4.9	4.1	5.9	12.2	2.9
	nPVI-V-I	<i>M</i>	3.1	3.4	3.2	3.4	4.1	2.8
		<i>SD</i>	0.5	0.5	0.5	0.5	0.9	0.6
	nVarcoV-I	<i>M</i>	2.9	3.1	3.1	3.2	5.3	2.7
		<i>SD</i>	0.4	0.4	0.4	0.4	1.0	0.3
		<i>N</i>	26	24	24	23	13	20

Table 3. Mean (*M*) and one standard deviation (*SD*) of the normalized pairwise variability index and variability coefficient for vowels averaged over narratives recorded within each sub-period. P1/P2=Japanese participants; NS=native speakers of English; JCtrl=Japanese controls; nPVI-V-D=normalized pairwise variability index of duration for vowels; nPVI-V-P=normalized pairwise variability index of pitch for vowels; nPVI-V-I=normalized pairwise variability index of intensity for vowels; VarcoV-D=variation coefficient of duration for vowels; VarcoV-P=variation coefficient of pitch (in *mel*) for vowels; VarcoV-I=variation coefficient of intensity (in *dB*) for vowels.

showed relatively high variability in both indices. As mentioned above regarding the pitch range, there appeared to be a great deal of variability in the use of pitch differences in their production of vowels among NS. It should also be noted that, as in the case of the duration indices, the initial means were much higher than those of J1 (nPVI-V-P=15.1, VarcoV-P=16.0) and J2 (nPVI-V-P=28.2, VarcoV-P=24.3).

Regarding intensity indices, neither nPVI-V-I nor VarcoV-I showed significant improvement through the training period in either participant. In P1, nPVI-V-I was higher

than the NS mean at T1, while VarcoV-I was around the same level as the NS mean in T3. In P2, both indices were significantly lower than the NS means at T4, indicating that there was some more room for improvement. The results indicated that, consistent with what was observed in the intensity range, the participants' use of intensity differences among adjacent and overall vowels did not increase through the training.

3-4. Stress-related measures

Table 4 shows the stress-related measures as a function of the sub-periods. These measures indicate a degree of difference in the acoustic properties (i.e., duration, pitch, and intensity) between the stressed and unstressed vowels of content words (STCN-D/P/I), and between stressed vowels of content words and vowels of mono-syllabic function words (STFN-D/P/I). In contrast to the normalized pairwise variability and normalized variability coefficients described above, these measures are sensitive to the stress status of the vowels and are only applied to content words and function words. With regard to the durational indices, both STCN-D and STFN-D in P1 were not significantly different from the NS mean even at T1. It should be noted, however, that the durational proportion of the unstressed to stressed vowels became smaller through the training, with the difference in STCN-D and STFN-D between T1 and T4 being at least marginally significant ($p=.07$ and $p=.03$, respectively). The results indicated that P1 had already acquired a sufficient level of durational control in differentiating the stressed and unstressed vowels at the beginning of the training session but attempted to further differentiate them by means of duration. As for P2, both STCN-D and STFN-D were significantly higher than the NS mean at T1, and significantly decreased from T1 to T4 ($p<.05$), where both measures were not significantly different from the NS means. The results indicated that P2 was able to learn to control duration in differentiating stressed and unstressed vowels through the training. It is notable that STFN-D decreased substantially between T2 and T3 where a host of the frequency measures (i.e., SR, AR, MLoR, PauseFreqBC, PauseFreqWC) also decreased significantly. This might suggest the de-emphasis of function words by means of duration is related to the faster rate of speech, reduction of pauses, and lengthening of chunks. In both participants, the initial measures were much lower than those of J1 (STCN-D=103.9; STFN-D=103.4) and J2 (STCN-D=100.7; STFN-D=107.4), and reached a comparable or lower level at the end of the training.

With respect to the pitch measures, the mean STCN-P and STFN-P were higher than

Participant	Measure		T1	T2	T3	T4	NS	JCtrl
P1	STCN-D	<i>M</i>	74.6	70.3	66.7	68.5	76.6	106.4
		<i>SD</i>	11.1	22.1	8.1	8.6	18.3	25.9
	STFN-D	<i>M</i>	70.7	64.2	56.0	60.3	64.7	128.4
		<i>SD</i>	16.6	18.4	13.5	11.7	16.8	31.6
	STCN-P	<i>M</i>	20.2	19.1	22.0	15.4	9.5	6.7
		<i>SD</i>	13.1	13.9	11.7	9.3	17.2	7.1
	STFN-P	<i>M</i>	34.5	19.2	21.4	17.5	11.5	13.6
		<i>SD</i>	16.3	20.3	12.4	14.4	19.4	8.7
	STCN-I	<i>M</i>	3.5	3.5	2.7	2.4	1.5	1.7
		<i>SD</i>	0.9	1.9	2.4	1.3	2.4	0.9
	STFN-I	<i>M</i>	4.3	3.6	3.0	2.3	3.1	1.2
		<i>SD</i>	1.6	1.9	2.6	1.9	2.9	1.9
		<i>N</i>	22	24	24	27	13	20
	P2	STCN-D	<i>M</i>	86.6	82.1	79.2	79.7	76.6
<i>SD</i>			11.2	9.6	9.1	9.3	18.3	25.9
STFN-D		<i>M</i>	84.5	80.4	71.6	68.8	64.7	128.4
		<i>SD</i>	16.8	16.7	9.9	11.4	16.8	31.6
STCN-P		<i>M</i>	11.6	24.8	26.7	36.2	9.5	6.7
		<i>SD</i>	13.0	11.3	12.1	9.1	17.2	7.1
STFN-P		<i>M</i>	27.1	40.2	41.8	48.8	11.5	13.6
		<i>SD</i>	15.0	12.7	10.5	12.9	19.4	8.7
STCN-I		<i>M</i>	2.2	2.3	2.4	2.8	1.5	1.7
		<i>SD</i>	0.8	0.7	0.8	0.6	2.4	0.9
STFN-I		<i>M</i>	1.9	2.9	2.8	3.0	3.1	1.2
		<i>SD</i>	1.2	1.1	1.1	1.2	2.9	1.9
		<i>N</i>	26	24	24	23	13	20

Table 4. Mean (*M*) and one standard deviation (*SD*) of the acoustic differences in duration, pitch, and intensity of vowels between stressed and unstressed syllables of content words and between unstressed vowels of content words and function words, averaged over narratives recorded within each of the sub-periods. P1/P2=Japanese participants; NS=native speakers of English; JCtrl=Japanese controls; STCN-D=proportion of vowel durations in unstressed to stressed syllables of content words; STCN-P=pitch difference (in *mel*) between unstressed to stressed vowels of content words; STCN-I=intensity difference (in *dB*) between unstressed to stressed vowels of content words; STFN-D=proportion of vowel durations in stressed syllables of content words to vowels in monosyllabic function words; STFN-P= pitch difference (in *mel*) between stressed vowels of content words and vowels in function words; STFN-I=intensity difference (in *dB*) between stressed vowels of content words and vowels in function words.

the respective NS means at T1 in both participants. As was found in the pitch range and the normalized variability indices described above, NS showed a great deal of variability in both indices, indicating that individuals differ greatly in the way they use pitch to differentiate stressed and unstressed syllables. Unexpectedly, both STCN-P and STFN-P decreased to approximate the NS mean in P1. In P2, both indices substantially increased

through the training period. The initial STCN-P and STFN-P in P1 and P2 were higher than those of J1 (STCN-P=-8.8; STFN-P=4.8) and J2 (STCN-P=9.9; STFN-P=18.8). The final measures in P1 and P2 at T4 were also much higher than those of J1 (STCN-P=11.6; STFN-P=18.6) and J2 (STCN-P=22.1; STFN-P=33.8). The results indicated that both participants had an acceptable level of ability to differentiate stressed and unstressed vowels by means of pitch at the entry of the training.

As was the case with the pitch indices, both of the intensity indices (i.e., STCN-I and STFN-I) in P1 were higher than the NS means at T1, and both of them decreased to become closer to the NS means through the training. In P2, STCN-I was higher than the NS mean, but STFN-I was significantly lower than the NS mean ($p<.01$) at T1. Both indices, however, increased significantly from T1 to T4 ($p<.01$). STCN-I and STFN-I of P1 at T1 were especially high as compared to J1 (STCN-P=-.5; STFN-P=1.0) and J2 (STCN-P=1.7; STFN-P=2.1), but those at the final sub-period in both participants were comparable to those of J1 and J2. The results indicated that both participants had comparatively less difficulty learning to use intensity in differentiating stressed and unstressed vowels.

4. Discussion and Conclusion

4-1. Summary of the major findings

The present study attempted to investigate how two adult Japanese learners of English (P1, P2) improved fluency and English speech rhythm through a long-term, individual-based speech training under the FI (formal instruction) settings where the participants had limited access and exposure to the target language. The data were based on the spontaneous narratives (N=97) the participants produced over the course of approximately one year.

1) Fluency measures (SR, AR, MLoR, PauseRat, PauseFreqBC/WC, PauseDurBC/WC)

All the fluency measures, including the speed measures (SR, AR, MLoR) and the breakdown measures (PauseRat, PauseFreqBC/WC, PauseDurBC/WC), changed significantly toward the NS means, although most of the measures did not reach the NS levels (except for PauseFreqWC and PauseDurWC). Most of the measures reached the same levels as those who participated in SA (J1, J2). In addition, the magnitude of improvement in most of the fluency measures was found to be greater between T2 and T3 (i.e., from the 4th to 9th month of the training) than the other pairs of the sub-periods.

Development of fluency and English speech rhythm through individual-based, long-term speech ...

2) Pitch and intensity ranges (RNG-P, RNG-I)

In both participants, the pitch ranges significantly increased to the same level as that of NS, and to the level higher than that of J1 and J2. On the other hand, the intensity ranges did not show substantial changes, and were significantly lower than the NS mean at the end of the training.

3) Rhythm measures (nPVI-V-D/P/I, nVarcoV-D/P/I)

The global variability measure of vowel durations (nVarcoV-D) increased significantly through the training in both participants but did not reach the NS level. On the other hand, the pairwise variability measure of vowel durations (nPVI-D-V) was already relatively high at the outset of the training and did not show significant improvement in either participant. Both global and pairwise variability measures of pitch were similar to or even higher than the NS means at the beginning of the training in both participants. Finally, neither global nor pairwise variability measures of intensity showed a significant developmental trend during the training.

4) Stress-related measures (STCN-D/P/I, STFN-D/P/I)

The durational measures (STCN-D, STFN-D) in P1 were not significantly different from the NS means at the beginning of the training but showed a further significant improvement during the training. Those in P2, being significantly different from the NS means at T1, improved significantly to reach the NS level at the end of the training. The pitch measures (STCN-P, STFN-P) were not significantly different from, or substantially higher than, the NS means at the initial sub-period. In P1, the measures decreased to approximate the NS means, while in P2, the measures further increased to greatly deviate from the NS means. As for the intensity measures (STCN-I, STFN-I), P1 showed the decline toward the NS means, as was found in the pitch measures. In P2, both indices significantly increased through the training, although STCN-I was higher than the NS mean at T1. Most of the measures at the beginning of the training were substantially lower (in terms of duration) or higher (in terms of pitch and intensity) than those of J1 and J2, but turned out to be comparable at the end of the training.

4-2. Limitations of the study

First of all, the present study is a case study with only two participants. As is the case with any case study, the present results have quite limited generality to other populations. Second, a number of measures at the entry into the speech training, especially, the stress-related measures, were high compared with the freshmen controls

and even those in the SA program, which merits special caution in interpreting the results. Next, the present study lacks the data on NS's evaluation of fluency, comprehensibility, and accentedness of the participants' narratives. So it is not clear how the modifications made in terms of the utterance fluency or the acoustic properties of their speech rhythm are related to the NS's perceived fluency, or perceived comprehensibility and accentedness. Future research is certainly required to investigate this point. Finally, the present results are based on a relatively short (approximately one-minute) narrative task which is repeated a great number of times. The participants might have developed some specialized ability to perform the task through the task repetitions. Thus, it might be the case that the present results might not be generalizable to other speaking situations which require turn-takings or a spontaneous production task with a longer stretch of time.

4-3. Fluency

Previous research which examined the effectiveness of FI (i.e., formal instruction settings) as compared with SA (i.e., study abroad settings) indicated that FI alone is not effective in significantly improving L2 fluency (e.g., Valls-Ferrer & Mora, 2014). The present results clearly indicated that individual-based speech training under the FI settings is effective in improving fluency. The results were consistent with the findings of the previous study (Tsushima, 2019) where J1 and J2 were able to significantly improve fluency before the SA period. What was surprising was that both of the present participants were able to reach the same or even higher level of achievements as J1 and J2 within one year of the speech training. One reason might be that P1 and P2 had already reached the higher level of fluency than J1 and J2 at the entry of the speech training program. For example, the initial SR was 69.2 and 88.9 for P1 and P2 respectively, while it was 48.1 and 58.1 in J1 and J2. However, the degrees of increment in SR during one year in the present study (i.e., P1 (42.5), P2 (25.9)) were comparable or even higher than those of the same amount of period including SA (i.e., J1 (34.0), J2 (23.3)). Another possible factor related to the comparable achievement of P1 and P2 might be the length of the sentences and words produced in the narrative. Although not reported in the present paper, Flesch-Kincaid readability scores, which reflect sentence length and word length (i.e., the number of syllables), were recorded for all the narratives. The Flesch-Kincaid grade levels in the latter period of SA were higher in J1 (4.96) and J2 (4.52) than P1 (3.72) and P2 (3.95), indicating that J1 and J2 produced longer and possibly

Development of fluency and English speech rhythm through individual-based, long-term speech ... more complex sentences, using words with longer syllables. Despite these possible factors, the results may be interpreted as underscoring the effectiveness of the long-term, individual-based speech training on improvement of fluency.

In both P1 and P2, most of the fluency measures made relatively rapid improvements between T2 and T3, preceded and followed by relatively slow or little improvement between T1 and T2, and T3 and T4 (with the exception of AR in P1 and of PauseFreqWC in P2). This pattern of development appears to fit into “the plateau effects” often reported in the other domains of skill acquisition (e.g., Recharads, 2008). As a matter of fact, this pattern of development was observed during the same sub-periods in the previous study (i.e., Tsushima, 2019). For example, the across-the-board rapid improvements among the fluency measures occurred between the 2nd and 3rd sub-periods, and between the 2nd and 4th sub-periods in J1 and J2, respectively. The results indicate that, even with the individual training, it takes more than half a year for the improvements of L2 fluency to manifest themselves. As most of the fluency measures fell short of those of NS at the end of the speech training, it would be interesting in future research to investigate how much continued speech training is necessary to bring about the next period of rapid progress. At any rate, the present results underline the importance of long-term, continued practice of spontaneous speaking in L2 speech learning.

4-4. Rhythm indices and stress-related measures

Previous research on phonological development during FI and SA has suggested that specific pronunciation instruction might be required to improve L2 learners’ phonological ability even in the SA settings (e.g., Avello & Lara, 2014). The present results were consistent with the conclusion that the speech training was effective in improving the ability to manipulate duration, which has been shown to be the most difficult acoustic property to acquire among duration, pitch, and intensity. For example, the global variability coefficients significantly improved in both participants, indicating that they became able to use a wider range of vowel durations in their speech. Although the pairwise variability measures did not show significant difference between the beginning and end of the training, the earliest narratives showed relatively low nPVI-V, indicating that the duration of pairwise vowels was relatively equal at the outset of the training. As for the use of duration to differentiate stressed and unstressed vowels, P2 showed a significant decrease in STCN-D and STCN-F between the first and last sub-period.

Especially, STFND showed a marked improvement between T2 and T3, as shown by the fluency measures discussed above. This indicated that P2 became able to de-emphasize shorten structure words such as prepositions (*in, on, at*) by shortening the vowel durations. This might have contributed to the increase of nPVI-V between T2 and T3 (from 52.0 to 55.3).

With regard to pitch, it was difficult to evaluate the effectiveness of the training as more than half the measures were approximate to or higher than the NS means in the first sub-period in both participants. It was indicated that, as compared with the control data (JCtr) which may represent the beginners' performance, P1 and P2 had acquired a higher level of ability to make use of pitch at the outset of the training. Still, the pitch range and pitch-related rhythm measures substantially increased through the training, indicating that both participants attempted to make use of a wider variety and wider range of pitch in their speech. As for intensity, the effects of the training appeared to be much less robust than those on duration and pitch.

It was also found that there was relatively a large amount of variability among NS in the rhythm measures, the stress-related measures, as well as the pitch and intensity range. The finding indicated that there was a great deal of individual variability in the way NS speakers varied pitch and made use of pitch in differentiating stressed and unstressed vowels in their speech. Inspection of the individual data revealed that some speakers spoke with relatively flat intonation within a fairly narrow pitch range, while others spoke with an alteration of high-pitched and low-pitched syllables using a wider range of pitch. In the former case, a phrasal intonation pattern (i.e., prolonged relatively flat intonation followed by rising intonation at the end of phrase) appears to reduce the lexical stress-related pitch differences, which resulted in lower nPVI-D-P, STCN-P, and STFND-P. It was also noticeable that the NS appeared to make use of especially duration to compensate for the relative lack of pitch variation to mark stressed syllables. This observation was supported by the results of the correlation analyses on STCN-D with STCN-P, and on STFND-D with STFND-P on the NS data. It was shown that the correlation between STCN-D and STCN-P and between STDN-D and STFND-P was significant (.644; $p < .025$, .668; $p < .025$, $N=13$), indicating that the NS speakers who produced a smaller degree of pitch differences between stressed and unstressed vowels produced a larger degree of durational differences, and vice versa. It was also found that the correlation of STCN-D with STCN-I was marginally significant (.478; $p = .098$), while that of STDN-D and STFND-I was significant (.611; $p < .05$). Although further data analyses on these points

Development of fluency and English speech rhythm through individual-based, long-term speech ... should be made, the observation is consistent with the previous research that indicate that NS of English make use of a combination of all three acoustic properties (i.e., duration, pitch, and intensity) to differentiate stressed and unstressed syllables (Beckman, 1986). Given the magnitude of individual variability in the NS data, future research should be directed toward determining the acceptable levels of each acoustic measure in terms of foreign accentedness or comprehensibility through examination of evaluation data from NS listeners. As duration and the other properties may compensate for each other, some kind of composite measures (or scores) might be desirable as an index of the acceptability.

4-5. Effects of FI and SA on development of fluency and production of English speech rhythm

The present results support the conclusion that individual-based, long-term speech training in the FI settings is effective in improving fluency and production of English speech rhythm. At the same time, however, it was found that P1 and P2 achieved comparable levels in terms of fluency and production of English rhythms as those of J1 and J2 at the end of the training. This finding makes the results of Tsushima (2019) equivocal as to the effect of SA on J1 and J2's improvements of the speech production ability. In order to delineate the effects of FI and SA and a combination of both, a more systematic group-based study is certainly required. In future, it would be interesting to compare the development of the speech production ability between students of the academic program with SA and those in the advanced English program without SA during the same period of time (in their sophomore and junior year).

4-6. Interpretation of the present results in the L2 production model

How can the present results be interpreted in the framework of the L2 production model? As shown in Table 4 and Figure 2 and 3, the present study analyzed between-clause and within-clause pause frequency and duration. As mentioned above, previous study has suggested that the within-clause pauses are primarily concerned with the formulation stage, while the between-clause pauses with the conceptualization stage (Kormos, 2006; Lambert et al., 2017; Saito, Ilkan, Magne, Tran, & Suzuki, 2018). It has also been suggested that speed measures reflect all stages of the production processes (Kormos, 2006). First of all, as shown in Figure 2 and 3, the within-clause pause frequency dropped substantially between the 1st and 2nd month in both participants. This may be

due to the reduction in the pauses within structurally simple phrases (e.g., It was // very // nice point). As the participants learned to use basic lexical items (e.g., “went to university”, “studied accounting”), the lexical access and the syntactic encoding might have become efficient. This interpretation was consistent with the slight drop in the within-clause pause duration.

Second, it was found that both between-clause pause frequency and within-clause pause frequency dropped substantially between T2 and T3, although the magnitude of the change was more pronounced in P1 than in P2. The drop in the breakdown measures was associated with the substantial increase of speed measures (AR, MLoR) in both participants. The results suggested that the processes in all three stages achieved greater efficiency during these periods. In the conceptualization stage, the participants might have become faster in creating a concept and organizing information structure of the message to be conveyed in the following clause. In the formulation stage, they might have become more efficient in lexical access and its associated syntactic encoding. These improvements might have enabled them to produce a larger chunk of words (i.e., a longer phrase/sentence), reflected in the longer MLoR (i.e., the mean length of runs). In addition, this more efficient processes might have allowed the participants to use greater attentional resources for the phonological and phonetic encoding in the formulation stage, and to pay attention to the acoustic properties of speech (e.g., duration, pitch, and intensity of vowels) in the execution/monitoring stage. The last point was consonant with the observed substantial drop in the durational difference between the stressed vowels in content words and the vowels of mono-syllabic function words (STFN-D) in both participants. The hypothesis for the greater speed and efficiency in the conceptualization and formulation stage is also supported by the concomitant drop in the between-clause pause and within-clause pause duration between T2 and T3 in both participants. The drop in the pause durations might also suggest that some parallel processing between the conceptualization stage and the formulation stage, and among encoding processes of different domains became operative during these periods.

Further analyses of the data are currently underway regarding the locations of both types of pauses in terms of the syntactic structure and the following lexical items. The results are expected to provide meaningful information about the relation between the breakdown measures and the underlying mechanisms of L2 speech production.

4-7. Implications for teaching pronunciation

First of all, the present study indicated that L2 learners in the FI settings achieved the same level of fluency and production of English rhythm as those who participated in SA as far as spontaneous narrative production was concerned. It was shown, however, that it took a relatively long period of time for the learners to show substantial improvement in both fluency and production of English rhythm. After a relatively rapid progress at the beginning stage of the training, both participants appeared to experience a period of plateau where they showed little or relatively slow progress in many aspects of fluency and rhythm development. The analogous pattern of development was observed in the participants in the previous study who participated in SA. It is recommended that L2 learners as well as instructors be ready for and patient during the period of little improvement (i.e., "plateau"). During this period, L2 learners presumably gain lexical knowledge and knowledge of syntactic and phonological structure, and practice putting such knowledge into practice in the task of spontaneous production.

Second, the present results may suggest that the spontaneous narrative production is an effective practice method to improve fluency and production of English speech rhythm. Previous research has suggested that the development L2 fluency involves two interrelated processes: automatization of syntactic, morpho-phonological, and phonetic encoding processes and the use of prefabricated language units called formulaic language (e.g., DeKeyser, 2007). DeKeyser (2007) suggested that, in order to improve the automatization of the encoding processes, it is of importance to provide L2 learners with the opportunity to practice repeating a certain set of syntactic and morphophonological rules (i.e., procedural knowledge of the encoding processes), and that the practice should be held under more and more demanding conditions where learners are required to express real ideas and intentions. In the narrative production practice repeated over a long period of time, L2 learners have ample opportunities to use a variety of lexical items and their associated syntactic and phonological properties and structures while completing the corresponding encoding processes. In addition, they can proceed from more basic to more advanced rules according to their own pace of progress. Although the level of difficulty in terms of the content of narratives was not controlled, the participants in the present study appeared to choose increasingly more difficult topics during the course of the training. For example, they started out from describing the basic daily activities (e.g., get up, go to the university, studied, have lunch, etc.). Then they started to describe an interesting event in detail with a beginning and end of the story. Toward the end of the

training period, they expressed their opinions on familiar issues (e.g., “Students should study accounting in college.”). Finally, recent technological advancement in speech recording has made it possible for learners to record their productions with sufficiently high sound quality using their mobile phone and submit them to instructors online for evaluation and feedback. Given these benefits, it is recommended that L2 instructors interested in improving students’ fluency and pronunciation consider the task of spontaneous narrative production as one of the methods of practice.

Finally, the present results suggest that, in improving pronunciation especially in spontaneous speech, it is important to give learners specific instruction to pay attention to pronunciation while speaking. As the production model described above suggests, improvement of pronunciation may involve modifications in at least the following representations and processes: 1) phonological representations of segments and prosodic structure, 2) phonological encoding processes (e.g., stress assignment) 3) phonetic encoding processes (i.e., commands regarding articulatory movements), and 4) monitoring of own productions and articulatory movements (i.e., self-monitoring). The present results suggest that the processes prior to the phonological encoding processes (i.e., the conceptualization stage, lexical access/retrieval, and syntactic encoding processes) may take up a great deal of attentional resources especially at the beginning of training. As these processes become speedier and more efficient, relatively greater attentional resources might be available for the pronunciation-related processes. To make the phonological, phonetic encoding and monitoring operate sufficiently, a certain degree of conscious attention appears to be required. To cope with the limited attentional resources at the early part of the training, it is recommended that relatively easier acoustic properties be targeted for modification, and that learners be encouraged to pay attention to them during speaking tasks. For example, when the learner has difficulty with production of English rhythm (e.g., flat pitch and equal syllable length regardless of the accent status of the syllable), targeting pitch would be more desirable than duration.

4-8. Concluding remarks

The present study attempted to examine how the two Japanese learners of English were able to improve fluency and production of English rhythm through individual-based speech training in the FI settings. It was found that both of them reached the level of achievements comparable to those who studied English in the SA settings. As described in 4-2 above, the case study with only two participants has a number of limitations,

Development of fluency and English speech rhythm through individual-based, long-term speech ... especially with regard to generalizability of the present findings to larger populations. In order to extend the present case study, a study with a group of participants is currently underway, where 13 students who belong to an academic program with a 5-month SA attend a class which focuses on improving fluency and pronunciation. As in the present study, their spontaneous narrative productions are being obtained regularly. This study will provide further data on whether and how the group-based instruction in the FI settings before SA is effective, as well as on whether and how their fluency and English rhythm productions will further improve during SA.

A case study might be extended by examining the effects of task complexity within a learner. Previous research has indicated that a degree of task difficulty affects speed fluency, breakdown fluency, as well as self-repair in spontaneous speech tasks (Lambert et al., 2020). Lambert, Kormos & Minn (2020), for example, used three task types (i.e., opinion, narration, and instructions) in spontaneous narrative tasks. It was presumed that, as compared with picture-based narration and instruction tasks, opinion tasks required greater attention in creation of concept plans (i.e., the conceptualization stage), leaving less attentional resources for the encoding processes in the formulation stage. Consistent with the predictions, it was found that the speakers produced more between-clause and within-clause pauses and engaged in more self-repairs in the opinion tasks than the others. It would be interesting to systematically vary the task difficulties within one participant during the speech training period. Detailed examination of fluency measures (i.e., speed, breakdown, and self-repair) and of contexts of pause occurrences might provide meaningful data that have important suggestions for the speech improvement process in the framework of the L2 speech production model.

Acknowledgement

The present research was supported by a research grant from Tokyo Keizai University in the academic year of 2019 (19–19). The author greatly appreciates the work of Jason Pipe in collecting the data from native speakers of English, and the participants' long-term and whole-hearted dedication to this research project.

References

- 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 Publishing Company.
- Avello, P., Mora, J. C., & Pérez-Vidal, C. (2012). Perception of FA by non-native listeners in a study abroad context. *Research in Language*, 10 (1), 63-78.
- Barry, W., Andreeva, B., & Koreman, J. (2009). Do Rhythm Measures Reflect Perceived Rhythm? *Phonetica*, 66 (suppl 1-2) (1-2), 78-94. Retrieved from <http://www.karger.com/DOI/10.1159/000208932>.
- Beckman, M. E. (1986). *Stress and non-stress accent* (Vol.7). Dordrecht, Riverton: Foris Publications.
- Boersma, P., & Weenink, D. (2014). Praat: doing phoetics by compter [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.
- Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching Pronunciation: a Course Book and Reference Guide* (2nd ed.). New York: Cambridge University Press.
- DeKeyser, R. M. (2007). *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*. New York: Campridge University Press.
- Fraser, H. (2001). *Teaching Pronunciation: a Handbook for Teachers and Trainers*: Department of Education Training and Youth Affairs (DETYA).
- Freed, B. (1995). What makes us think that students who study abroad become fluent? In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad*. (pp.123-148). Amsterdam: John Benjamin's.
- 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.
- 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. In *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.
- Kormos, J. (2006). *Speech Production and Second Language Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Lambert, C., Aubrey, S., & Leeming, P. (2020). Task preparation and second language speech production. *TESOL Quarterly*. doi: 10.1002/tesq.598
- Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language processing. *Studies in Second Language Acquisition*, 39, 167-196.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*,

- Development of fluency and English speech rhythm through individual-based, long-term speech ...
- 40 (3), 387-417.
- Levelt, E. J. M. (1989). *Speaking; from Intention to Articulation*. Cambridge: A Bradford book: the MIT press.
- 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
- Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals*, 43 (3), 488-503.
- 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. (2008). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau, & A. Bel (Eds.), *A Portrait of the Young in the New Multilingual Spain* (pp. 241-263). Clevedon: Multilingual Matters.
- Mori, Y., Erickson, D., Rilliard, A., & Hori, T. (2014). Effects of vowel duration, intensity, and articulation rate on judgments of naturalness and intelligibility of Japanese learners' English. *Journal of the Phonetic Society of Japan*, 18 (2), 30-41.
- Mori, Y., Hori, T., & Erickson, D. (2014). Acoustic Correlates of English Rhythmic Patterns for American versus Japanese Speakers. *Phonetica*, 71 (2), 83-108. Retrieved from <http://www.karger.com/DOI/10.1159/000365594>.
- 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 Publishing Company.
- Pérez-Vidal, C., Juan-Garau, M., & Mora, J. C. (2011). The effects of formal instruction and study abroad contexts on foreign language development: The SALA project. In C. Sanz & R. P. Leow (Eds.), *Implicit and Explicit Conditions, Processes and Knowledge in SLA and Bilingualism* (pp. 115-138). Washington DC: Georgetown University Press.
- 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. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10585517>.
- Reichardt, J. C. (2008). *Moving Beyond the Plateau: From Intermediate to Advanced Levels in Language Learning*. New York: Cambridge University Press.
- 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.
- Saito, K. (2012). Effects of instruction on L2 pronunciation development: a synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 46 (4), 842-854. doi: 10.1002/tesq.67
- Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*

- tics*, 39, 593-617.
- Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. New York: Routledge.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17 (1), 84-119.
- 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. (2019). An exploratory case study on development of fluency and English speech rhythm through individual speech training and a study abroad program. *Journal of Humanities & Natural Sciences*, 145, 29-63.
- 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. (2011). *The development of oral fluency and rhythm during a study abroad period*. (Ph. D.). Universitat Pompeu Fabra, Barcelona.
- 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 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>