The interaction of micro- and macro-rhythm measures in English and Japanese as first and second languages: A case study of two Japanese EFL students

Jun Nagao¹, Marta Ortega-Llebaria²

¹Gifu Shotoku Gakuen University, Japan
²University of Pittsburgh, US

Abstract

L2 rhythm has often been measured as the duration variation of vowel and consonant intervals using Varco and PVI measures (micro-rhythm). In the present study, in addition to micro-rhythm, we examined macro-rhythm (i.e., the duration variation of intervals between F0 events), and vowel duration in relation to three phonological patterns (i.e., content vs. function words, stressed and unstressed vowels, and final sentence lengthening) in a series of attempts by two Japanese students of English to imitate approximately 2-minute segments of English TED talks. In general, both participants obtained native-like Varco and nPVI scores for vowels and consonants in their last imitation attempt. In contrast, only one of the participants related the variation of vowel duration to stress. The same participant obtained macro-rhythm scores closer to those of the English TED talk. A debriefing interview indicated that neither participant showed conscious awareness of their changes in macro-rhythm. These results suggest that Japanese speakers may need to learn stress and micro-rhythm patterns before implementing macro-rhythm, and that the often-used micro-rhythm measures of rhythm need to be complemented with both macro-rhythm measures and phonological patterns’ measures in order to capture L2 rhythmic patterns more accurately. Pedagogically, imitation seems to work for some learners to learn macro-rhythm patterns.

Index Terms: Japanese EFL learners, acquisition of English prosody, micro-rhythm, macro-rhythm, case study

1. Introduction

In nature, the sensation of rhythm comes from the regular repetition of auditory and visual stimuli, e.g., sound patterns in time, and shape and color patterns in space. The more regular this repetition is, the stronger the sensation of rhythm. In speech, Ramus and colleagues [1, 2] demonstrated the psychological reality of rhythm by showing that infants were able to discriminate between language pairs that belonged to different rhythm classes—namely stress-timed, syllable-timed, and mora-timed classes. However, finding the acoustic correlates of rhythm classes has proven to be elusive at best [3]. We believe that part of the problem resides in examining rhythm only at one timescale. Following previous research [4, 5, 6, 7], we propose that rhythm in speech operates simultaneously at different timescales, and that the sensation of rhythm comes from the coupling of the different rhythmic levels. For example, there is a micro-rhythm operating at roughly the syllable/word timescale which is captured by duration measures of vocalic and consonantal intervals such as %V and ΔC [1, 2, 3], and a macro-rhythm operating above the word level (i.e., Accentual Phrase (AP), Intermediate Phrase, Intentional Phrase), which is captured by F0 measures such as peak-to-peak duration intervals and MacR_V. Var [4, 8]. While there is a prolific body of literature on L2 and cross-linguistic micro-rhythm [e.g., 1, 2, 3, 10], to our knowledge, only one study has examined macro-rhythm across languages [4], and no study has yet examined both micro- and macro-rhythm measures across languages and in L2.

To fill in this gap, we examined micro-rhythm and macro-rhythm measures in (a) two spaced attempts by two Japanese EFL students at imitating approximately 2-minute segments of English TED talks, (b) the English TED talk segments themselves, and (c) the Japanese renditions of these TED talk segments. Moreover, the Japanese-American English comparison constitutes an interesting arena to test micro- and macro-rhythm measures because these languages differ widely in these measures. Micro-rhythm is stronger in English than in Japanese because all English content words must have one stressed syllable creating a regular and strong micro-rhythm, while in Tokyo Japanese, words may or may not have a pitch accent creating an irregular and weaker micro-rhythm. As for macro-rhythm, each AP in Japanese has the same fixed pitch contour. Because each breath group has one or several accentual phrases with this fixed pitch contour, macro-rhythm becomes very regular and rhythmic. In contrast, in English, the melody of a sentence comes from the sequence of pitch-accented stressed syllables. Since pitch accents have different F0 shapes and can be separated by an unpredictable number of unstressed and unaccented syllables, English macro-rhythm becomes irregular and weak. Thus, for these Japanese students, approximating English rhythm means increasing micro-rhythm by learning stress while decreasing macro-rhythm by unlearning Japanese Accentual Phrases.

These students’ imitation attempts were part of the 2020 English Recitation Contest at Gifu Shotoku Gakuen University, Japan. After selecting an approximately 2-minute English speech sample, memorizing it, and rehearsing it for 200 hours, the students performed their recitation in the actual contest. These two students were selected for this study because, despite the fact that both showed enormous improvement, one clearly sounded more (American English) native-like than the other according to the contest panel of judges.

Given the above rhythmic differences between languages, participants, and imitation attempts, we wondered (a) whether micro- and macro-rhythm measures could capture these differences and (b) whether these students were aware of their improvements in micro- and macro-rhythm.
2. Method

2.1. Procedure

In this paper, we are reporting findings based on the productions of two students who participated in the 2020 English Recitation Contest at Gifu Shotoku Gakuen University followed by oral interviews with them. In this contest, students recite an approximately 2-minute native English speech sample that they choose from a TED talk, movie, or YouTube video. The participants spent around 200 hours imitating the original speech. Students are instructed to sound like the original speakers, mainly Americans, and to convey emotion. In their practice, the two students used an application on their smartphone called Audipy. With this application, users can adjust the speed of the original audio freely, and the two students were advised to start “shadowing” with 0.5x speed so they would be better able to pay attention to the prosody. Over time, they gradually adjusted the speed to 0.75x or faster. Moreover, the two regularly met with their instructors (including the first author) to receive feedback and ask questions about their pronunciation.

2.2. Participants

The two students were chosen for the study because 1) they had recorded their practice from the day they chose their speech, 2) the instructor (first author) observed a noticeable improvement in their pronunciation through the process of preparing for the contest, and 3) one sounded more native-like than the other according to the contest judges, despite similar practice regimes, sparking curiosity in the first author.

Student 1 (S1) is a Japanese male, who was a 22-year-old third-year university student at that time. He spent over 200 hours preparing for the contest. Student 2 (S2) is a Japanese female student who was also 22 years old and in her 3rd year in university at the time of the study, and spent close to 200 hours practicing. The contest’s panel of judges gave a higher score to the male student’s imitation.

2.3. Materials

We examined the original speech sample by native speakers of English (hereafter, called “English L1”), and the first imitation attempt from each student, which was over a month before the contest (hereafter, called “English L2 First Imitation”), and the last imitation attempt, which was the actual performance delivered at the contest (hereafter, called “English L2 Last Imitation”). Finally, we examined the Japanese version of the same speech sample (hereafter, called “Japanese L1”). In total, the data examined amounts to approximately 17 minutes of speech. Each S1 sample contained 42 breath groups, and each S2 sample 45. S1 used a TED talk, Keep your goals to yourself by Derek Sivers, and S2 used a speech by Michael Jackson, Sadly, we live in a state of fear. From all these samples, the subset of 12 sentences per participant (see Appendix) was chosen and used for a second set of measurements, as explained in the next section.

In addition to the analysis of the speech samples, we conducted an interview with each participant individually after the contest to see how they perceived their improvements and how much they were aware of their interlanguage. They were asked to listen to their oldest performance, which was six weeks before the contest, and then they were asked questions such as, “Comparing the first attempt and the actual performance at the contest, what sort of things were difficult for you at first?” , “What sort of things were you able to improve?” , and “How did you prepare for the contest?” Their responses were transcribed and translated into English.

2.4. Measurements

Because the native speakers of English and the Japanese participants had different speech rates, we used normalized measures. To measure micro-rhythm, we labeled all the consonants and vowels in English L1, English L2 First Imitation, English L2 Last Imitation, and Japanese L1. We input vowel and consonant duration intervals into Correlatore and obtained the following measures for each one of the intonational phrases of the samples: Vowel%, VarcoV, VarcoC, and normalized PVI for vowels and consonants.

In order to relate the variation of the vowel intervals to three potential sources of variation—namely stress, content vs. function words, and final IP lengthening, we labeled the 12 sentences (see Appendix) from each sample for function and content words as well as for stress. We computed the ratios for vowel durations in adjacent content vs. function words (i.e., the stressed vowel of the content word vs. the unstressed vowel of the function word as in ‘image of’ in sentence 3, S1), stressed vs. unstressed vowels in word-internal trochaic feet (e.g., underlined a/i in ‘imagine’, underlined e/i in ‘telling’), the last vowel of an intonational phrase with the preceding stressed vowel (e.g., ‘image of yng’), and target words to measure stress and content vs. function words that did not align on their right edge with any IP or IP boundary.

To measure macro rhythm, we labeled duration intervals between pitch turning points on the 12 sentences (see Appendix): the distance between tonal events, or T-to-T, the distance between peaks, or H-to-H, and the distance between valleys, or L-to-L. With these duration intervals, Varco scores for H, L, and T were calculated for each sentence.

3. Results

First, the duration-based micro-rhythm measures of V%, Varco and normalized PVI for vowel and consonant intervals were calculated on the whole set of speech samples per participant—namely English L1, English L2 First Imitation, English L2 Last Imitation, and Japanese L1. As illustrated in Figure 1, the native speech samples displayed the expected results for morae-timed and stress-timed languages showing higher %V values and less variation of vowel intervals (in both Varco V and normalized PVI-V) in Japanese L1 than English L1. With regards to English L2, both students started with V% values that were close to the English target from the first imitation attempt. S2 approximated English values on VarcoV and nPVI-V from the first imitation attempt. Only S1 showed improvement. He started with values close to those of Japanese L1 and the final imitation attempt ended with values close to the American English native speaker. As for consonant intervals (Figure 2), only S1 showed improvement from less to more variation. In contrast, S2 overshot consonant variation from the first imitation attempt.
Second, since the Varco-V and nPVI-V showed that both students produced native-like values in the variation of vowel intervals, this variation was further examined in relation to three specific contexts, i.e., content vs. function words, IP lengthening, and stress. As illustrated in Figures 3 and 4, both participants produced target-like values for the content-function word ratios and at the end of intonational phrases. Paired T-tests showed no statistically significant differences at an alpha level 0.05. However, with regards to stress (Figure 5), only S1 showed target-like values. S2 failed to reduce the duration of unstressed vowels yielding ratios closer to 1 and larger than those of the native speaker (M .59 ms. SD .50) in both her first imitation attempt (M 1.10, SD .50, t(16)=5.120, p <0.001) and her last imitation attempt (M .81, SD .33, t(7)= -3.304, p=0.013).

Third, results of the macro-rhythm measures T-to-T, H-to-H, and L-to-L are displayed in Figures 6, 7, 8 respectively. S1 approximated native-like patterns for the three measurements. In contrast, S2 only approximated native-like patterns for the H tone, failing to approximate English patterns for T (M .67, SD .01) in the first imitation attempt (M .4, SD .04, t(16)=3.947, p=0.001) and the last one (M .5, SD .07, t(14)= 1.941, p= 0.05), and for L (M .807, SD .06) especially in the first imitation attempt (English L2 First Imitation: M .46, SD .04, t(14)=3.018, p= 0.009; English L2 Last Imitation: M.58 SD .057, t(14)=1.8484, p=0.08). These results indicate that S1 consistently decreased the high macro-rhythm regularity of Japanese to approximate the more irregular macro-rhythm of English. In contrast, S2 approximated the F0 peaks to target F0 peaks while failing to reduce regularity for T and L.
Finally, the debriefing interview investigated the participants’ awareness of these micro- and macro-rhythm patterns. As summarized in Table 1, both S1 and S2 showed awareness of speech rate, English phonemes that Japanese does not have, and duration contrast between content vs. function words. However, they did not show much understanding of how intonation and stress work in English.

The findings of the interview are summarized in the following table.

Table 1: Self-Perception of their interlanguage

<table>
<thead>
<tr>
<th>Speech rate</th>
<th>Phonemes</th>
<th>Content vs. function words</th>
<th>Macro-rhythm</th>
<th>Micro-rhythm</th>
<th>V and C duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In summary, the participants’ productions showed the following patterns:

**Micro-rhythm:**
(1). For V%, both S1 and S2 approximated target-like values from the first imitation attempt.
(2). For VarcoV and Vnpi, S1 improved, while S2 had target-like values from the first imitation attempt.
(3). For VarcoC and Cnpi, S1 improved, while S2 overshot from the first imitation attempt.

Sources of variation for the duration of vowel intervals:
(4). Both S1 and S2 produced target-like duration variability in content vs. function words, and at the end of IP.
(5). Only S1 imitated duration variations in word stress.

**Macro-rhythm:**
(6). Only S1 approximated target-like macro-rhythm in his imitation attempts in all measures from the first imitation attempt for T-to-T, and H-to-H. S2 failed to approximate macro-rhythm for the T-to-T in both imitation attempts, showed improvement in the L-to-L measures, but achieved target-like imitations only for H-to-H.

**Interview**
(7) Both students described their improvements in speech rate, pronunciation of segments, changes in intensity, and the reduction in time and quality of the vowels in function words in comparison to vowels in content words.
(8) Students were not aware of the variations in duration in relation to stress or at the end of IP, or pitch variations in relation to macro-rhythm.

4. Discussion

Results from the present case study have theoretical and pedagogical implications for L2 rhythm. First, although both S1 and S2 obtained native-like micro-rhythm scores, S1 was judged to sound more native-like, especially for vowel intervals. A detailed analysis of vowel duration showed important differences between participants, namely, only S1 reduced the duration of unstressed vowels. Because stress is the key feature of English micro-rhythm, it is likely that S1 sounded more native-like in part because he reproduced the duration patterns of lexical stress, suggesting that micro-rhythm measures such as VarcoV are too general to capture the L2 rhythm differences of these two participants.

Second, in general, S1 imitated English macro-rhythm better than S2. Moreover, only S1 elongated stressed syllables. This pattern of results may suggest that for Japanese speakers, learning to decrease macro-rhythm entails the production of stress. From a theoretical perspective, this entailment makes sense since the irregular English macro-rhythm relates to the fact that pitch accents are anchored directly into the stressed syllables of the words within a sentence without the mediation of AP. Altogether, these considerations indicate that to account for L2 rhythm, the widely used micro-rhythm measures of VarcoV and nPVI need to be complemented by macro-rhythm measures, and that variation of vowel duration needs to be examined in relation to the phonological patterns that condition this variation in the target language.

As for pedagogical implications, results showed that the two participants were aware of speech rate and vowel reduction in function words but not of stress and macro-rhythm. These differences in awareness suggest that some prosodic features such as speech rate can be taught explicitly, while some features such as stress and macro-rhythm can still be acquired through 200 hours of imitation practice without the participants’ conscious awareness.

5. Conclusion

The findings of the present case study suggest that the acquisition of English macro-rhythm by native Japanese speakers is contingent on the acquisition of stress at the micro-rhythm level. Moreover, a precise description of L2 rhythm requires complementing micro-rhythm measures with macro-rhythm measures.

This case study is based on the speech of two participants. Further studies are needed to investigate the effective acquisition order of micro- and macro-rhythm of English by examining individual differences within a larger set of participants. It will also be interesting to find out whether stress can be taught explicitly, or whether it needs to be taught implicitly first via perception training.

6. References


**Appendix**

The subset of 12 sentences chosen from all the samples per participant. | signals a pause within a breath group, and || signals the end of a breath group.

**Student 1**

01. Imagine deciding right now that you are going to do it. ||
02. Imagine telling someone that you meet today what you are going to do. |
03. Imagine their congratulations and their high praise of you. |
04. Doesn’t it feel good to say it out loud? | Ah, well! |
05. You should have kept your mouth shut. ||
06. Cause that good feeling now will make you less likely to do it. ||
07. Ideally you would not be satisfied until you’d actually done the work. ||
08. So, if this is true, what can we do? |
09. You can delay the gratification of the social acknowledgment brings. ||
10. and I can understand that your mind mistakes the talking for the doing. ||
11. But if you need to talk about something, you can say it in a way that gives you no satisfaction |
12. So audience, next time you are tempted to tell someone your goal, what will you say? ||

**Student 2**

01. Sadly, sadly, we live in a state of fear. ||
02. Everyday we hear of war. ||
03. We hear of nations hurting each other, of neighbors hurting each other, of families hurting each other, and the children killing each other. |
04. We must learn to live, and love each other before it’s too late. |
05. We have to stop! |
06. We have to stop the hurting, we have to stop living in fear of our own neighbors. |
07. We would like all of you now to take the hand of the person to the left and to the right of you. |
08. Now, tell the person next to you that you care for them. |
09. This is what makes the difference. |
10. Together we can make a change of the world. |
11. Together we can help to stop prejudice. |
12. It’s our only hope. |