A closer look at the different rhythm metrics in L2 English: Inter- and intra-speaker variability and individual difference

L2英語の異なるリズム指標による分析：学習者間、学習者内の変動と個人差要因

Atsushi Miura and Masahiko Komatsu
三 浦 篤 史・小 松 雅 彦

SUMMARY

This report aims to validate the rhythm indices using data from second language learners. Previously, rhythm indices had been well adopted to measure rhythm in both L1 and L2 studies. However, the validity of the indices has still been debated. In this study, 10 L1-Japanese/L2-English read-aloud data were investigated with a focus on inter- and intra-participant differences and proficiency differences. The results of linear mixed-effect models provided the effect of sentence type and learners' target language proficiency on L2 rhythm. Several points were stated for future research.

Key words: L2 rhythm, Individual difference, Rhythm indices, *%V*, *ΔC*, *VarcoV*, *VarcoC*

1. Introduction

Evaluating second language (L2) proficiency is an important debate for not only for second language acquisition (SLA) research, but also for language education. Previous research has provided indices that can measure global evaluation (e.g., fluency, intelligibility, comprehensibility, accentedness) as well as segmental evaluations such as pronunciation. Since Li & Post (2014) pointed out that research on L2 rhythm was a neglected field, the L2 rhythm research has been increasing. One remaining issue in rhythm investigation is that it is not clear how rhythm can be measured. Previous studies provided indices to classify the language rhythm characteristics based on the Rhythm Class Hypothesis (Nespor, 1990). Although the number of studies on rhythm has been growing, the validity of these indices is not yet clear. Dellwo, Leemann, and Kolly (2015) have tried to validate the indices using qualitatively different speeches (e.g., slow vs. fast). A recent study on L2 rhythm provided conflicting results (cf. Ozaki, Yazawa, & Kondo, 2017). Therefore, a validation study using L2 data in a multidimensional perspective is needed to explore the L2 rhythm characteristics and the development of speech rhythm in the L2 acquisition process.

2. Background Literature

2.1 Indices of the Speech Rhythm

Previous studies on the indices of speech rhythm tried to distinguish the language variation, namely, stress-timed (e.g., English), syllable-timed, and/or mora-timed (e.g., Japanese) based on the hypothesis that the rhythm class is related to syllable structure and vowel reduction (Dauer, 1983). Thus, they proposed the indices that refer to the length of the consonants and vowels. Ramus, Nespor, and Mehler (1999) provided that the rhythm type is revealed by the average proportion of vocalic intervals ($\%V$), the standard deviation of vocalic intervals (ΔV), and the average standard deviations of consonantal intervals (ΔC).

Grabe and Low (2002) also attempted to classify the rhythm class, based on the prediction by Low, Grabe, and Nolan (2000). They claimed that the Pairwise Variability Index (PVI), the average of difference in duration between the immediately consecutive intervals of vowels or consonants, can capture the rhythmic differences between languages.

Dellwo (2006) criticized that the indices provided by Ramus et al. (1999) can be affected by the speech rate of the material. He provided the normalized indices, *VarcoV* and *VarcoC* (standard deviations divided by the mean and multiplied by 100), as shown below.

$$VarcoV = 100 \times \Delta V / meanV$$
$$VarcoC = 100 \times \Delta C / meanC$$

So far, these three types of metrics have often been used to investigate speech rhythm, including both L1 and L2. However, even in L1 studies, the debate on reliable indices has been continuing (Asadi et al., 2018; Dellwo et al., 2015; He & Dellwo, 2016; Wiget et al., 2010; see also Ozaki et al., 2017 for L2 research).

2.2 Adopting Rhythm Indices in Second Language Research

In research on second language speech performance, speech rhythm has been neglected to investigate (Li & Post, 2014). Some research tried to provide the L2 rhythm characteristics, adopting the measurements shown in the previous section. Guilbault (2002) proposed a variability index (VI) metric and discussed the effect of L2 proficiency and length of exposure to the target language on L2 rhythm, concluding that the categories (i.e., stress- vs. syllable-timed) were not the main effects. Gut (2009) conducted the experiment before and after the interventions by two groups (i.e., 9- and 6-month stay abroad), showing no effect of length of abroad. The research which adopted the metrics shown above (i.e., *%V*, *ΔV*, *ΔC*, and *PVI*) has provided mixed results. Sarmah, Gogio, and Wiltshire (2009) focused on the differences between the two learner groups (i.e., less than 6 months and more than 18 months stay abroad). According to their results, the vocal values of the former group, such as vocalic *nPVI* (normalized PVI) and *%V* were higher than the others' values. Stockmal, Markus, and Bond (2005) also found significant differences between less proficient and more proficient learners in terms of consonantal variability but not vocalic variability. Ozaki (2015) used corpus data to compare the proficiency difference. In the result, as for the vocalic intervals such as *%V*, *ΔV*, vocalic *rPVI* (raw PVI) were significantly different. That is, more proficient learners can utter native-like rhythms, which is also supported by perception experiments in which native speakers judge speech (Ordin & Polyanskaya, 2015a; Polyanskaya, Ordin, & Busa, 2017). However, she stated that the result of consonantal intervals was unclear.

Regarding the rhythm of L2 other than English, Wagner (2014) investigated the L2 Polish rhythm in German and Korean learners, suggesting the effect of L1 transfer on the L2 rhythm system. Similarly, Li and Post (2014) investigated the differences and commonalities of the L2 rhythmic system with a focus on the development of speech rhythm. They compared different L1 groups to shed light on the question of whether the development of L2 rhythms follows a universal path or not. Their results suggested a multivariate path to develop the L2 rhythm and the possibility of L1 transfer. This is in line with some previous studies with different L1 backgrounds (e.g., Ordin & Polyanskaya, 2015b for L1-French and L1-German/L2-English).

A recent study by Lee and Low (2021) investigated Japanese English with particular focus on the value of vocalic *nPVI*, compared with Chinese and Singaporean English. Their results showed that Japanese English is qualitatively different from the others.

To sum up, the L2 rhythm seems to be affected by the L1 background. Especially, vocalic measures typically follow the development of the native one, although the consonantal characteristics are not yet clear. However, we still do not know about the validity of the metrics. In fact, as in the L1 rhythm research, rhythm metrics modestly related to language differences as well as speakers' individual differences between speakers (Arbaniti, 2012). In the case of L2, there exist more variations in individual characteristics with respect to their proficiency, instruction background, length of exposure, and aptitude. Thus, research comparing all different rhythm measures shown above, using L2 speech data and taking a closer look at the individual variation of rhythm is needed to validate the rhythm indices and explore L2 rhythm characteristics. In this report, we focus on the differences between Ramus et al.'s and Dellwo's indices with respect to learners' proficiency and sentence characteristics.

3. The Current Study

Two research questions are addressed.

1) To what extent do the sentence characteristics affect L2 rhythm values?

2) To what extent does proficiency affect L2 rhythm?

First, the characteristics of the sentence (i.e., length, number of consonants/vowels) can affect the L2 rhythm values. That is, the variation in intra-participant data can vary. Because Dellwo's (2006) *Varcos* are normalized by speech rate, it can be predicted that the variations in intra-participant are smaller in *Varcos* index than Ramus' one. Second, it can be predicted that proficiency may affect L2 rhythm similarly to the previous studies (Ordin & Polyanskaya, 2015a; Ozaki, 2015; Polyanskaya et al., 2017; Stockmal et al., 2005).

3.1 Methods

3.1.1 Participants

10 participants (Male: 3, Female: 7) were recruited from a private university in Japan. The average age is 20.6 (age range: 19–22). All participants have learned English for about 9 years in an English as a Foreign Language (EFL) setting. To ensure their global English proficiency, they were asked to answer their scores of the TOEIC test (Test of English for International Communication) (*M*=505, *Range*=400–625), indicating that their English proficiency was estimated around the intermediate (B1) level of the benchmark from the Common European Framework of Reference (CEFR) (Tannenbaum & Wylie, 2008). Since some participants did not provide their TOEIC scores, a short proficiency test called Minimal English Test (Goto, Maki, & Kasai, 2010) was conducted to ensure their proficiency differences among the participants. Regarding their daily contact with English, all participants used English only in classes. None of them has studied abroad for more than a half year.

3.1.2 Materials and procedure

"The North Wind and the Sun" form The International Phonetic Association (1999) was used. The passage is shown below.

> *The North Wind and the Sun were disputing which was the stronger / when a traveler came along wrapped in a warm cloak / They agreed that the one who first succeeded in making the traveler take his cloak off should be considered stronger than the other / Then the North Wind blew as hard as he could / but the more he blew the more closely did the traveler fold his cloak around him / and at last the North Wind gave up the attempt / Then the Sun shined out warmly / and immediately the traveler took off his cloak / And so the North Wind was obliged to confess that the Sun was the stronger of the two*
> [Slashes are inserted by the authors.]

First, participants were asked to read the passage silently (slashes were not included in the text they read). They were asked to search for the meaning and pronunciation with dictionaries when they encounter words they do not know while reading silently. After that, the recording session was conducted. They read the passage aloud three times in the session. All of them are recorded. Since the first recording is for practice, it was excluded from the analysis. Among the other two recordings, the one which they read with fewer repetitions based on the researchers' judgement was used for the current analysis. Feedback on pronunciation and grammar was not given throughout the whole session. The paper-pencil background questionnaire was completed prior to the recording session. To verify their proficiency at the testing point, the participants also took the Minimal English Test (MET) (Goto et al., 2010).

3.1.3 Data Labeling and analysis

The recorded data were first segmented into each vowel and consonant with the BAS web services (Schiel, 1999; Kisler, Reichel, & Schiel, 2017). Then, using *Praat* (Boersma & Weenink, 2016), a trained coder checked the automatic coding and manually adjusted the boundaries if necessary. The finalized data were calculated and measured with the indices with *%V*, ΔV and ΔC (Ramus et al., 1999) and *VarcoV* and *VarcoC* (Dellwo, 2006). To compare the intra-participant variability, the passages were divided into nine sentences or phrases (see the passage in the previous section: / shows the separation for the analysis). To investigate the effects of proficiency and sentence characteristics, each measurement was statistically analyzed through a series of linear mixed effect models with each value as a dependent variable, proficiency and sentence type as a fixed effect, and participants as a random effect[1], using free statistical software JASP (version 0.13.1; JASP Team, 2020). To validate the effect of individual learners' proficiency, correlation analyses for each value is conducted. To look closer at each rhythm metrics, the individual visualizations are conducted based on the previous studies.

4. Result

The results from several linear mixed-effects models were summarized in Table 1, which shows the effect of proficiency and sentence type.

Table 1. Results of the linear mixed-effects model.

Value		*df*	*F*	*p*
%V	Proficiency	8, 1	0.394	.850
	Sentence	8, 8	2.84	**.081+**
	Proficiency * Sentence	64, 8	0.612	.867
*Δ*V	Proficiency	8, 1	0.837	.694
	Sentence	8, 8	0.892	.562
	Proficiency * Sentence	64, 8	0.794	.720
*Δ*C	Proficiency	8, 1	0.839	.693
	Sentence	8, 8	3.021	**.069+**
	Proficiency * Sentence	64, 8	0.703	.795
Varco V	Proficiency	8, 9	4.057	**.026***
	Sentence	8, 9	2.433	.104
	Proficiency * Sentence	64, 9	1.598	.229
Varco C	Proficiency	8, 9	7.419	**.003***
	Sentence	8, 9	4.443	**.020***
	Proficiency * Sentence	64, 9	0.722	.786

Note. +p<.10. *p<.05

The linear mixed effect models were carried out for each rhythm metrics. As for *%V* and *ΔC*, sentence characteristics were a marginally significant effect ($p = .081$ and $p = .069$ respectively), but proficiency and their interaction did not detect significance ($p = .850$ and $p = .867$ for *%V*, and $p = .693$ and $p = .795$

for *ΔC*). There is no significant effect on the *ΔV* value ($p = .694$ for proficiency, $p = .562$ for sentences, and $p = .720$ for interaction). Regarding the Varco values, proficiency was a significant effect in both *VarcoV* and *VarcoC* ($p = .026$ and $p = .003$, respectively). Sentence characteristics were a significant effect only for *VarcoC* ($p = .020$) but not for *VarcoV* ($p = .104$). The interactions had no significant effect ($p = .229$ for *VarcoV* and $p = .786$ for *VarcoC*).

Table 2 shows the results of correlation analysis between proficiency and each value. There is no significant correlation between each value and the proficiency score.

Table 2. Results of the correlation analysis.

Value	Pearson's *r*	*p*	*95% CI*	
			Lower	*Upper*
%V	0.158	.137	–0.051	0.353
ΔV	–0.078	.465	–0.281	0.131
ΔC	–0.110	.303	–0.310	0.100
Varco V	0.017	.875	–0.191	0.223
Varco C	0.182	.086	–0.026	0.375

In addition to statistical analyses, each rhythm value was plotted based on the previous study (see Figures 1 and 2).

5. Discussion

5.1 Comparing L2 rhythms with focus on sentence characteristics.

The first research question is to what extent L2 rhythm is affected by sentence characteristics. To investigate this effect, the recorded passage was segmented into nine (see 3.1.2). According to the results, *%V*, *ΔC*, and *VarcoC* in this study were affected by the characteristics of the sentences. These results suggested that some passage was outstanding for the outcome values. It is clear that *%V* is affected by the characteristics of the sentence. Since it is calculated through the length of vowels, sentences that have many vowels and those which has less vowels will be different. Similar to that reason, *ΔC* can be different. The interesting point of these results is *VarcoC*. Although

Dellwo (2006) noted that *Varcos* are not influenced by speech rate, our results suggested the possibility that *Varcos* values can be affected by sentence characteristics, including speech rate and/or speed of utterance. However, this point is consistent with his recent research (Dellwo et al., 2015). While it is clear that sentence characteristics (e.g., length) are related to the rhythm, further investigation is needed to ensure the variation within a speaker. It is not clear what type of sentence characteristics affect their value. To answer these, it is necessary to use different materials such as Wolf passage, which is considered more reliable to investigate English pronunciation (Deterding, 2006), comparing L1 and L2 in the same speaker, or comparing the same passages they speak several times, which remains for future research.

5.2 Proficiency may affect Varcos but not others.

Based on the proficiency results, only *Varcos* can show the significant effect of the proficiency (see Table 1). However, further correlation analysis (see Table 2) revealed that there is no correlation between the proficiency score and the values. Therefore, the significant effect suggested that some participants were significantly different from other participants. Taking into account the individual results of each value (Figures 1 and 2), there seems to be not proficiency effect. That is, highly proficient learners do not speak in a steady rhythm, nor do they speak in a rhythm similar to that of native speakers. It can be predicted that more proficient learners can speak in a steady manner but lower proficiency learners cannot. However, the result of this study suggests that at least proficiency is not the key to the stability of the L2 rhythm. However, the data were not from advanced learners, and more advanced learners may utter in a more stable rhythm.

Or, it can be predicted that there are some sentences that are easier to utter for some speakers but not for others. This might not be explained by their target language proficiency but by the individual learner characteristics such as familiarity or attention to speech sounds, as well as their articulation rate. A recent study shows that even native speakers differ in the way they perceive (Kogan & Mora, 2022). Investigation of production difference might be the key for L2 acquisition as well. Therefore, more studies will be needed to investigate the effect of learner characteristics.

6. Conclusion and future directions

In this report, we tried to compare the different indices (Ramus' and Dellow's). Based on our results, Ramus' indices may be affected by sentence characteristics. Especially for *VarcoC*, our results cast doubt on the discussion by Dellwo that *Varcos* are not influenced by speech rate. Our further investigation using the different metrics that was not used in this report (e.g., Ozaki et al., 2017) is necessary. Furthermore, data from more advanced learners should be analyzed to investigate the difference by proficiency levels. What we can conclude here is that the analysis of L2 data can presumably make the validation study of rhythm indices take one step further.

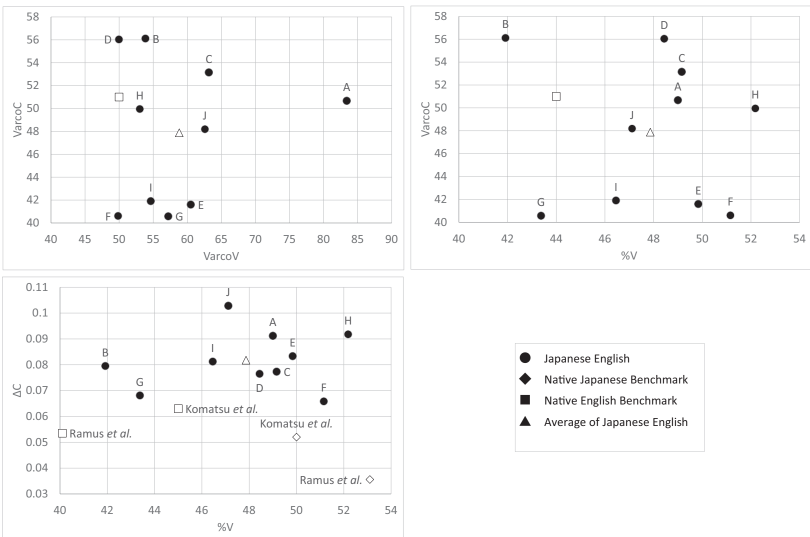

Figure 1 Plot summary for different metrics.

Note. The letters in the figures represent participants' ID. "A" is the highest proficient learner and "J" is the lowest.

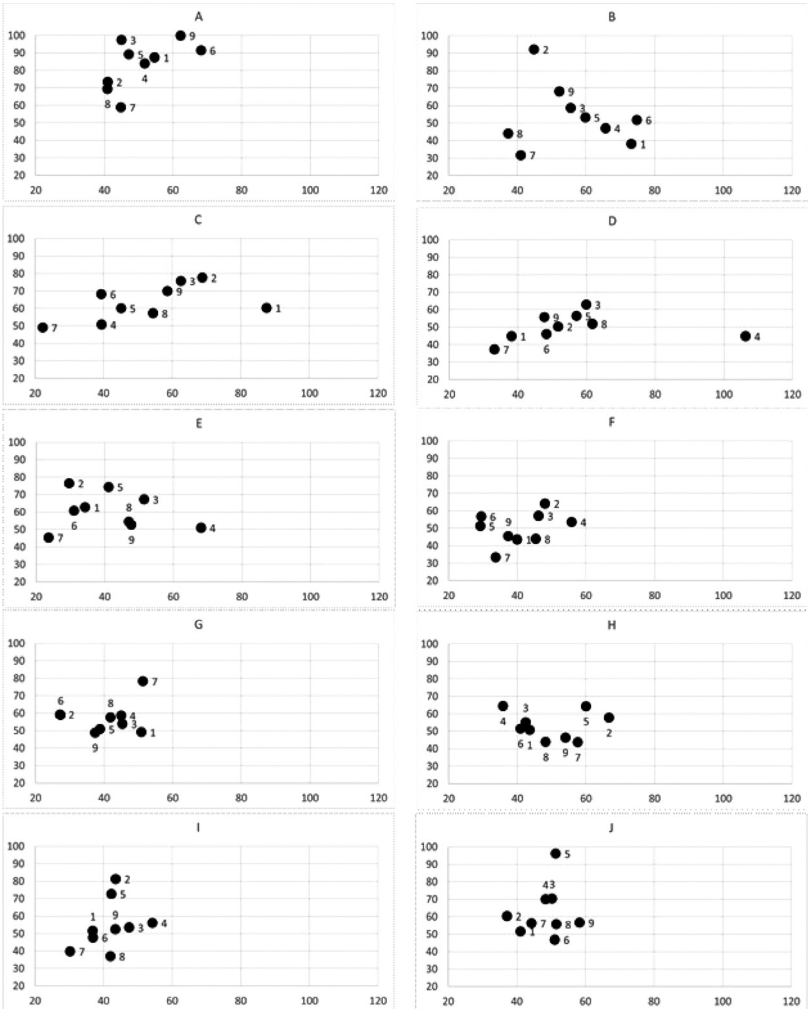

Figure 2 Results of Varcos.

Note. The X axis is the vocalic interval (*VarcoV*) and Y axis is the consonantal interval (*VarcoC*). The proficiency level was ordered alphabetically (i.e., A is the highest; J is the lowest). The numbers in each figure represent sentence/phrase number (see 3.1.2 Materials and procedure)

Acknowledgment

This research started when the first author was in a master's course. We would like to thank all the participants of this study, and Rika Ishizaki who helped us with our data collection. All errors and omissions are, of course, our own.

Note

1 Linear mixed effect model can include some factor as random effect (e.g., participants). However, due to the limited sample size of this study, this modeling here seems less robust.

References

Arbaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*, 351-373. https://doi.org/10.1016/j.wocn.2012.02.003

Asadi, H., Nourbakhsh, M., He, L., Pellegrino, E., & Dellwo, V. (2018). Between-speaker rhythmic variability is not dependent on language rhythm, as evidence from Persian reveals. *International Journal of Speech Language and the Low, 25*, 151-174. https://doi.org/10.1558/ijsll.v23i2.30345

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer*: Version 6.0.14 [computer program]. Retrieved from http://www.praat.org/

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51-62. https://doi.org/10.1016/S0095-4470(19)30776-4

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and Language-Processing: Proceedings of the 38th Linguistics Colloquium, Piliscsada, 2003* (pp. 231-241). Peter Lang.

Dellwo, V., Leemann, A., & Kolly, M. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *The Journal of the Acoustic Society of America, 137,* 1513. https://doi.org/10.1121/1.4906837

Deterding, D. (2006). The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association, 36,* 187-196. https://doi.org/10.1017/S0025100306002544

Goto, K., Maki, H., & Kasai, C. (2010). The minimal English test: A new method to measure English as a second language proficiency. *Evaluation & Research in Education, 23,* 91-104. https://doi.org/10.1080/09500791003734670

Grabe, K., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 525-546). Mouton de Gruyter.

Guilbault, C. P. G. (2002). *The acquisition of French rhythm by English second language* *learners.* (Unpublished doctoral dissertation). University of Alberta, Edmonton.

Gut, U. (2009). *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German.* Peter Lang.

He, L., & Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech Language and the Low, 23,* 243-273 https://doi.org/10.1558/ijsll.v23i2.30345

Kisler, T., Reichel U. D., Schiel, F. (2017): Multilingual processing of speech via web services. *Computer Speech & Language, 45,* 326-347. https://doi.org/10.1016/j.csl.2017.01.005

Kogan, V. V., & Mora, J. C. (2022). The effects of individual differences in native perception on discrimination of a novel non-native contrast. *Laboratory Phonology,* 13. https://doi.org/10.16995/labphon.6431

Komatsu, M., Aoyagi, M., &, Sugawara, T. (2005). OGI_TS onsei raberu ni mirareru Nihongo jihatsu onsei no rizumu: Onsosū no bunseki kara onseichō no bunseki e [Rhythm of Japanese spontaneous speech observed in phonetic labels of OGI_TS: From the analysis of the number of phonemes to the analysis of the length of speech segments]. In The Phonetic Society of Japan (Ed.), *Proceedings of the 19th General Meeting of the Phonetic Society of Japan* (pp. 107-112).

Lee, D. D., & Low, E. (2021). The sounds of Japanese English: Monophthong vowels and rhythmic patterning. *Asian Englishes, 23,* 30-50. https://doi.org/10.1080/13488678.2020.1868815

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition, 36,* 223-255. https://doi.org/10.1017/S0272263113000752

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: 'Syllable-timing' in Singapore English. *Language and Speech, 43,* 377-401. https://doi.org/10.1177/00238309000430040301

Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), *Logical issues in language acquisition* (pp. 157-175). Foris.

Ordin, M., & Polyanskaya, L. (2015a). Perception of speech rhythm in second language: The case of rhythmically similar L1 and L2. *Frontiers in Psychology, 6,* 316. https://doi.org/10.3389/fpsyg.2015.00316

Ordin, M., & Polyanskaya, L. (2015b). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America, 138,* 533-544. https://doi.org/10.1121/1.4923359

Ozaki, Y. (2015). Nihongo bogo washa niyoru eigo hatsuwa rizumu: J-AESOP ko-pasu wo motoni [English utterance rhythm by native Japanese speakers: Based on the J-AESOP corpus]. *JATLaC Journal, 10,* 58-61.

Ozaki, Y., Yazawa, K., & Kondo, M. (2017). L2 English speech rhythm of Japanese speakers: An alternative implementation of the Varco metrics. In Phonetics Teaching and Learning Conference (Ed.), *Proceedings of the Phonetics Teaching and Learning*

Conference 2017 (pp. 84-88). Chandler House.

Polyanskaya, L., Ordin, M., & Busa, M. G. (2017). Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language. *Language and Speech, 60,* 333-355. https://doi.org/10.1177/0023830916648720

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73,* 265-292. https://doi.org/10.1016/S0010-0277(00)00101-3

Sarmah, P., Gogio, D. V., & Wiltshire, C. (2009). Thai English: Rhythm and vowels. *English World-Wide, 30,* 196-217. https://doi.org/10.1075/eww.30.2.05sar

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In J. J. Ohara, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.). *Proceedings of the XIVth International Congress of Phonetic Sciences: ICPhS 99* (pp. 607-610). The American Institute of Physics.

Selinker, L. (1972). Interlanguage. *IRAL: International Review of Applied Linguistics in Language Teaching, 10,* 209-231. https://doi.org/10.1515/iral.1972.10.1-4.209

Stockmal, V., Markus, D., & Bond, D. (2005) Measures of native and non-native rhythm in a quantity language. *Language and Speech, 48,* 55-63. https://doi.org/10.1177/00238309050480010301

Tannenbaum, R. J., & Wylie, E. C. (2008). Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology. *ETS Research Report Series, 2008,* i-75.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association.* Cambridge University Press.

Wagner, A. (2014) Rhythmic structure of utterances in native and non-native Polish. In N. Campbell, D. Gibbon, & D. Hirst (Eds.). *Proceedings of Speech Prosody 2014* (pp. 337-341).

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America, 127,* 1559. https://doi.org/10.1121/1.3293004

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35,* 501-522. https://doi.org/10.1016/j.wocn.2007.02.003