Rhythm metrics of spontaneous speech and accent in Japanese and English

Yoichi Mukai *[1] **and Benjamin V. Tucker** †[1]
[1]Dept. of Linguistics, University of Alberta, Alberta, Canada, T6G 2E7

1 Introduction

The rhythm class hypothesis is an attempt to classify the human impression of language rhythm based on the concept of isochrony, which posits that language rhythm underlies the regulation of a particular unit re-occurring at regular intervals [1]. The present study provides an experimental evaluation of rhythm metrics between read and spontaneous speech in Japanese and English.

Researchers, such as Ramus et al. [2] and Grabe & Low [3], have proposed several rhythm metrics by adapting Dauer's [4] notion that phonological and phonetic factors modulating structures of syllables contribute to the perception of different language rhythms. Two types of rhythm metrics have been proposed, 'global' and 'local' metrics. Global metrics reveal the degree of overall durational variability of segments in utterances, e.g., %V—the relative proportion of vocalic intervals, and VarcoV—the standard deviation of vocalic intervals divided by the mean of vocalic intervals and multiplied by 100 [5]. Local metrics, on the other hand, capture durational differences between consecutive vocalic and consonant intervals, such as nPVI_V—the mean of the differences between successive vocalic intervals divided by their sum and multiplied by 100, see Eq. (1), and rPVI_C—the mean of differences between successive consonant intervals, see Eq. (2) [3].

$$\text{nPVI} = \sum_{m-1}^{k=1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1}/2)} \right| / (m-1) \qquad (1)$$

$$\text{rPVI} = \sum_{m-1}^{k=1} |d_k - d_{k+1}| / (m-1) \qquad (2)$$

Researchers have also applied these metrics to measure the influence of first language (L1) rhythm on second language (L2), and suggest that VarcoV and nPVI_V are particularly useful in predicting impressionistic judgements of accentedness [6].

Researchers, however, have identified inconsistencies in the classification of language rhythms across studies [3]. One possible cause is the type of speech collected (e.g., whether speech is spontaneously produced or carefully read). Though a few studies have compared read sentences to spontaneous speech using these metrics [7], Japanese has not yet been investigated in this respect. We predict that these metrics will differentiate read and spontaneous speech, with spontaneous speech falling on the faster side of each metric by showing a smaller proportion of vowels (lower %V) and more variability in the duration of vocalic intervals (higher VarcoV) than read sentences. Also, that spontaneous speech will allow for better discrimination of language groups by showing that L2 English falls between L1 English and L1 Japanese.

∗. mukai@ualberta.ca
†. bvtucker@ualberta.ca

2 Method

Spontaneous speech was collected from two existing data sets, The Wildcat Corpus [8] and unpublished data from Warner [9]. Forty-eight utterances were retrieved from three native speakers of Japanese (one male ; three utterances × three speakers = nine utterances), English (three males ; three utterances × three speakers = nine utterances), and six English-speaking Japanese subjects (one male ; five utterances × six speakers = 30 utterances). All were manually labeled using Praat [10], and segmented following Arvaniti [7]. Praat was then used to calculate %V, VarcoV, rPVI_C, and speech rate. We compare the present research to Grenon & White's [11] data (read speech) for L1 and L2 rhythm in Canadian English and Japanese. The resulting data were modeled, with linear mixed- effects regression using the lme4 package [12] in R to investigate the relationship between the metrics and the language groups across the speech types. The metrics and speech rate were entered as separate dependent variables, and the language groups and speech types were treated as fixed effects, with subjects (speakers) as random effects.

3 Results

Table 1 shows the means and standard errors of the metric scores and speech rates in the L1 English, L1 Japanese, and L2 English for the current study. Figure 1 illustrates the comparison of %V and VarcoV between the current study (spontaneous speech) and Grenon & White (read speech) [11]. A comparison of the present study to the Grenon & White's [11] data shows that the two data sets are significantly different for all the metrics except rPVI_C (%V (t=-2.69), VarcoV(t=2.27), Speech Rate(t=-5.04)). Analysis of %V for the spontaneous speech data indicates that the vowel for L1 English takes a significantly larger portion of the utterance than L1 Japanese (t=2.66). A similar result is also found for rPVI_C (t=2.924)

Groups	L1EN	L1JP	L2EN
Metrics			
%V	39.87 (1.19)	49.47 (2.34)	44.98 (1.57)
VarcoV	52.90 (5.26)	65.84 (5.86)	57.33 (3.22)
rPVI_C	83.18 (12.27)	42.53 (2.57)	68.39 (5.11)
SRate	3.82 (0.20)	4.72 (0.34)	4.11 (0.19)

Table 1: The means and standard errors of the metric scores and speech rates (SRate) in L1 English (L1EN), L1 Japanese (L1JP), and L2 English (L2EN).

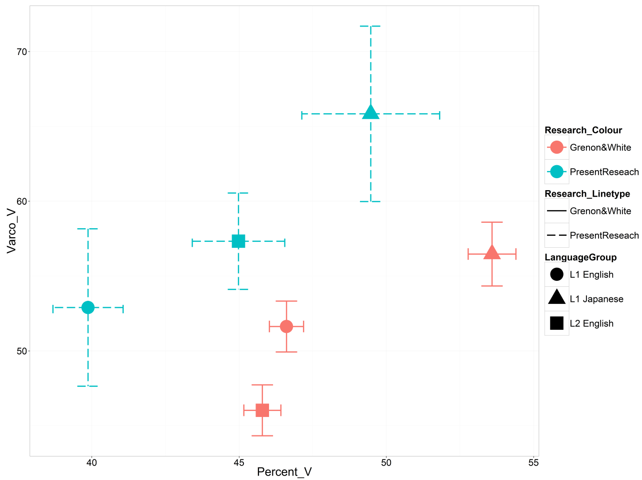

Figure 1: The plot shows the means and standard errors of the %V and VarcoV scores in L1 English, L1 Japanese, and L2 English speakers both in the present study (spontaneous speech) and Grenon & White (read speech) [11].

which indicates that there is more variability in the successive consonant intervals for L1 English as compared to L1 Japanese. Additionally, we find that L2 English significantly differs from L1 Japanese for rPVI_C (t=2.356). Analysis of VarcoV and Speech Rate do not show any significant differences between language groups (though some comparisons are trending toward significance).

4 Discussion

The variability and lack of some significant effects found in the present study may be due to the limited data size (we have only 48 utterances) rather than speech type differences. However, statistical analysis reveals that there are some significant differences between the metrics of the two speech types. As predicted, there are significant differences between spontaneous speech and read speech. Further, we find that spontaneous speech better discriminates the language groups by showing L2 English between L1 English and L1 Japanese, as predicted. It is possible that the read speech causes L2 English speakers to overemphasize English rhythm causing the L2 English not to fall between the two L1s. The difference may also be due to the differing L2 fluency levels. We predicted that spontaneous speech would fall on the faster side of each metric. However, the spontaneous data is produced at a slower speech rate than the read sentences, and it is produced with a smaller proportion of vowels and more variability in the duration of vocalic intervals. This is likely because subjects in Grenon & White [11] practiced reading their sentences before recording, and the familiarity with the sentences encouraged them to read at a high rate, but maintained the proportion of vowels and variability in the duration of vocalic intervals. It is noteworthy that L2 English significantly differs from L1 Japanese for rPVI_C, suggesting that spontaneous L2 English is similar to L1 English. In other words, their L2 speech is rhythmically less accented in terms of rPVI_C.

5 Conclusions

This study provides an experimental evaluation of rhythm metrics between read and spontaneous speech in Japanese and English. Statistical analysis indicates that these metrics are sufficient for distinguishing between the two speech types. Additionally, our findings indicate that L2 speakers do fall between L1s in spontaneous speech. Further examination is required with a larger sample of data to further investigate the details of linguistic isochrony and its relationship to impressionistic observations of language rhythm.

Acknowledgments

We would like to thank Laurence White and Izabelle Grenon for generously sharing their data with us. We would also like thank Natasha Warner for sharing her spontaneous data, and Volker Dellwo for sharing his Praat scripts.

References

[1] David Abercrombie et al. *Elements of general phonetics*, volume 203. Edinburgh University Press Edinburgh, 1967.

[2] Franck Ramus, Marina Nespor, and Jacques Mehler. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3) :265–292, 1999.

[3] Esther Grabe and Ee Ling Low. Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, 7(515-546), 2002.

[4] Rebecca M Dauer. Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*, 1983.

[5] Volker Dellwo. Rhythm and speech rate : A variation coefficient for deltac. *Language and language-processing*, pages 231–241, 2006.

[6] Laurence White and Sven L Mattys. Calibrating rhythm : First language and second language studies. *Journal of Phonetics*, 35(4) :501–522, 2007.

[7] Amalia Arvaniti. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3) :351–373, 2012.

[8] Kristin J Van Engen, Melissa Baese-Berk, Rachel E Baker, Arim Choi, Midam Kim, and Ann R Bradlow. The wildcat corpus of native-and foreign-accented english : Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and speech*, 53(4) :510–540, 2010.

[9] Natasha Warner. Speech reduction across languages and dialects, 2015.

[10] Paul Boersma and David Weenink. Praat, a system for doing phonetics by computer. 2001.

[11] Izabelle Grenon and Laurence White. Acquiring rhythm : A comparison of l1 and l2 speakers of canadian english and japanese. In *Proceedings of the 32nd Boston University conference on language development*, pages 155–166, 2008.

[12] Douglas Bates, Martin Maechler, and Ben Bolker. lme4 : Linear mixed-effects models using s4 classes. 2012.