



# Rhythm of East-Asian Speakers of English in English Conversation

Hyunsong Chung<sup>1</sup>

<sup>1</sup>Korea National University of Education

hchung@knue.ac.kr

## Abstract

This paper investigated the rhythm of East-Asian speakers of English in conversation. A speech corpus of 150 conversations between speakers of English in East Asia with different L1 backgrounds was collected and the rhythm was analysed. It was found that L1 difference of the interlocutors and the speakers' daily use of English influenced %V, while the speakers' daily use of English influenced  $\Delta V$ . A weak correlation between the two speakers' rhythm in each conversation was also found in %V and  $\Delta V$ . No significant effects were found in PVI. The results revealed that the speakers tended to accommodate the rhythm of their utterances to that of the other interlocutors'. Further study on the speaking rate of the speakers, the percentage of function words and the number of pauses in the utterances might be required to overcome some inconsistencies found in the results of the rhythm metrics used in this study.

**Index Terms:** English rhythm, %V,  $\Delta V$ , VarcoV, PVI

## 1. Introduction

This paper investigated the rhythm of East-Asian speakers of English who live in Hong Kong. This study is different from previous studies in that it used the speech data of impromptu and spontaneous utterances in conversations which involve interactions between speakers. Most of previous studies used the speech data of read speech, carrier-phrases or monologues in interviews which lack interactions between speakers.

Different rhythm metrics have been proposed to analyse the rhythm. Most of them were based on interval durations of vowels or consonants in the utterances as in the followings:

- $\Delta V$ : the standard deviation of vocalic interval duration [1]
- $\Delta C$ : the standard deviation of consonantal interval duration [1]
- %V: the sum of vocalic interval duration divided by the total duration of vocalic and consonantal intervals and multiplied by 100 [1]
- VarcoC: the standard deviation of consonantal interval duration divided by the mean consonantal interval duration and multiplied by 100 [2]
- VarcoV: the standard deviation of vocalic interval duration divided by the mean vocalic interval duration and multiplied by 100 [3]
- PVI (Pairwise Variability Index): the difference in duration of vowels between successive syllables [4]
- VI (Variability Index): the normalised duration of consecutive syllables [5]

[1] argued that %V is the measure which offered the best acoustic correlate of rhythm classes, based on their findings that %V was smaller in English than in French. [4] was able to provide an acoustic basis for the existence of rhythm classes across languages using the PVI. [5] analysed the rhythm of Singaporean and British speakers of English using VI and found that the average VIs of the Singaporean speakers were smaller than those of the British speakers. [2] found that VarcoC seemed to differentiate rhythm classes better than  $\Delta C$ , because  $\Delta C$  varied considerably as a function of speech rate. According to [3], rate-normalised metrics of vocalic interval variation, VarcoV and PVI were shown to discriminate between hypothesized "rhythm classes," as did %V. According to them, VarcoV offered the most discriminative analysis in quantifying the influence of first language on second language rhythm. Other than these metrics of interval variation, speech rate, articulation rate, the percentage of functions words in the utterances (%fw) and the number of pauses in the utterances (NumSil) were also used to analyse the rhythm [6].

Many studies tried to analyse the rhythm of Asian speakers of English. [7], [8], [9] found that tense vowels were generally twice as long as lax vowels, but they were not reduced in unstressed syllables in the speech data of speakers of English in ASEAN region. [10] calculated the VIs of read speech by Korean and American speakers of English and compared the values. According to them, Korean speakers had difficulty in implementing the different rhythmic patterns of stressed and unstressed syllables. [11] calculated the VIs of Korean speakers of English in London when they communicated with speakers of English with different L1 backgrounds. It was found that there was no significant variability in the measurement of the syllable-to-syllable duration for the utterances of Korean speakers, regardless of their interlocutor's language background.

[12] investigated the correlation among the perceptual evaluation of rhythm, perceived fluency scores and the values from rhythm metrics, using the speech data of Korean speakers of English. The results showed that %V, VarcoV, PVI which are based on interval durations of vowels and %fw and NumSil had more correlations with the perceptual evaluation and perceived fluency scores than the other metrics which are based on interval durations of consonants.

[13] analysed the rhythmic patterns of spontaneous speech and read speech of native speakers of English using the rhythm metrics. It was found that the vocalic variation was bigger in spontaneous speech than in read speech and that PVI and VarcoV were better acoustic correlates to show this variation.

Because many of previous studies showed that rhythm metrics based on vocalic intervals were better acoustic correlates to investigate the rhythm, this study used %V,  $\Delta V$ , VarcoV and PVI to analyse the rhythm of EAST-Asian speakers of English in Hong Kong. After obtaining average

values from these metrics, we investigated how gender, L1 difference, duration of English language learning and the daily use of English affected the values. In order to identify whether there was rhythmic accommodation between speakers, the values of each speaker were compared with their partners' in the conversation.

## 2. Data

The recordings were conducted at a university in Hong Kong, because it was believed that it was much easier to recruit speakers of English with different L1 backgrounds there. We collected the speech data of 750 minutes conversations by 100 speakers who had a total of 21 different L1 backgrounds. 75 female speakers and 25 male speakers participated in the recordings. Table 1 shows the demographics of the participants.

Table 1: *Demographics of the participants.*

L1	Number
Cantonese	30
Mandarin	26
Korean	20
Others (Number/L1)	24/18

Other L1s included Urdu, Tagalog, Putonghua, Japanese, Hokkien, Cantonese/Mandarin (Bilingual), Tamil, Spanish/Cantonese, Sinhala, Punjabi, Malay, Hakka, Jiang Huai Guan Huam, Indonesian, Hindi, English, Cantonese/Hainanese (Bilingual), and Bengali. The participants' exposure to English was as follows in Table 2.

Table 2: *Participants' exposure to English.*

	Length of Learning (Year)	Daily Use of English (%)
Number		100
Mean	14.6	45.3
Minimum	7	5
Maximum	27	100
S.D.	3.2	25.3

We made a group of four in 25 different groups. Each speaker had a conversation with one of three other interlocutors in the same group in three different five-minute sessions, which produced a total of six conversations for each group. The recordings were conducted in two soundproof rooms using Tascam DR-100 and Tascam HD. Two channels were used to collect each speaker's utterances. The topics for each session included stresses in life, future dream, movie, an ideal spouse, teacher qualifications and sports.

Orthographic transcriptions for each conversation were conducted manually and then each conversation was forced aligned by using FAVE-align programme [14]. Using Praat scripts and Python, the duration was automatically calculated and it was put into the calculation for rhythm metrics.

For the statistical analysis, we grouped the individual utterances into 'Same' and 'Different' according to the interlocutors' L1 background difference. Based on duration of English language learning, we made three groups (S: less than 13 years; M: between 14 and 15 years; L: more than 16 years.). We also made three groups according to the daily use of English (L: less than 25%; M: 26-50%; H: 60-100%). Values of %V,  $\Delta V$ , VarcoV and PVI were subjected to two-way

ANOVAs respectively with the dependent variable (each of %V,  $\Delta V$ , VarcoV, PVI values) and the independent variables speaker's gender (M, F), L1 difference between speakers (Same, different), duration of English language learning (H, M, L) and the daily use of English (H, M, L).

## 3. Results

### 3.1. %V

Because %V is the sum of vocalic interval duration divided by the total duration of vocalic and consonantal intervals, the bigger the value, the more syllable-timed the utterance. The two-way ANOVA results showed that L1 difference [ $F(1,282)=10.3$ ,  $p<0.01$ ] and the daily use of English [ $F(2,282)=3.3$ ,  $p<0.05$ ] had significant effects on rhythm. Interactions among independent variables were not found. Difference of %V values according to L1 difference can be found in Table 3 and Figure 1.

Table 3: *%V according to L1 difference.*

L1	N	%V	SD	Diff.	p
Same	29	40.00	4.64		
Different	271	43.60	5.77	3.54	0.01

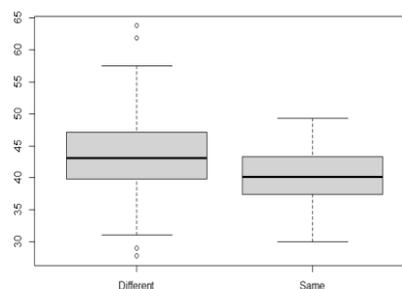


Figure 1: *A boxplot of %V according to L1 difference.*

%V was smaller when the speaker talked to the interlocutor who had the same L1 background, while it was bigger when they talked to the interlocutor who had different L1 background. It revealed that the speakers tended to be more syllable-timed when they had a conversation with the interlocutors who had different L1 backgrounds.

Difference of %V values according to the daily use of English is shown in Table 4 and Figure 2.

Table 4: *%V according to the daily use of English.*

Daily Use	N	%V	SD
H	102	42.19	5.98
M	96	43.50	5.38
L	102	44.08	5.77

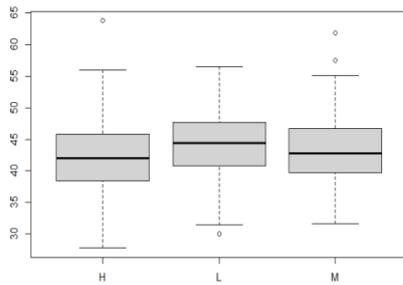


Figure 2: A boxplot of %V according to the daily use of English.

A post-hoc test revealed that the difference between Group H and Group L was significant. Because Group H, who used English more frequently, showed smaller %V value than Group L, who used English less frequently, the utterances by the speakers who used English more frequently tended to be more stressed-timed.

### 3.2. $\Delta V$

Because  $\Delta V$  is the standard deviation of vocalic interval duration, the bigger values indicate stress-timing of the utterances. The two-way ANOVA results showed that only the daily use of English affected  $\Delta V$  values ( $F(2,282)=12.4$ ,  $p<0.001$ ). Other main effects or interactions were not found. Difference of  $\Delta V$  values according to the daily use of English can be found in Table 5 and Figure 3.

Table 5:  $\Delta V$  according to the daily use of English.

Daily Use	N	$\Delta V$	SD
H	102	101.12	32.12
M	96	111.86	38.02
L	102	127.78	46.33

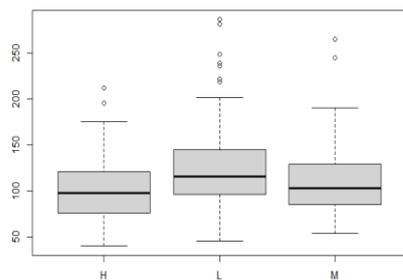


Figure 3: A boxplot of  $\Delta V$  according to the daily use of English.

A post-hoc test showed that there was a significant difference between Group H and Group L. The result was quite surprising, because it means that the speakers who use English less frequently tended to have more stress-timed rhythm. It was also conflicting with the results of %V.

### 3.3. VarcoV

Because VarcoV is the standard deviation of vocalic interval duration divided by the mean vocalic interval duration, the bigger values indicate stress-timing of the utterances. A significant main effect of speaker's gender emerged [ $F(1,282)=9.8$ ,  $p<0.01$ ]. VarcoV according to the speaker's gender can be found in Table 6 and Figure 4.

Table 6: VarcoV according to the speaker's gender.

Gender	N	VarcoV	SD	Diff.	$p$
F	224	82.78	17.31	7.68	0.01
M	76	90.47	22.98		

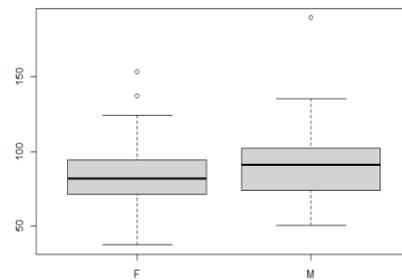


Figure 4: A boxplot of VarcoV according to the speaker's gender.

The result indicates that male speakers' utterances were more stress-timed than female speakers'.

### 3.4. PVI

PVI is the difference in duration of vowels between successive syllables. So, if PVI duration value is bigger, the utterance is more stress-timed. No significant main effects emerged.

### 3.5. Rhythm accommodation

In order to investigate whether the speakers accommodate their rhythm to their partners', the correlation of the rhythm of the speakers was analysed. There was a weak but significant correlation in %V values of the speakers ( $r=0.35$ ,  $p<0.001$ ). A very weak and significant correlation was also found in  $\Delta V$  values of the speakers ( $r=0.12$ ,  $p<0.05$ ). No correlation was found in VarcoV and PVI values. Depending on the rhythm metrics, some degree of rhythm accommodation between speakers was found.

## 4. Discussion and conclusion

The results of this study showed that the listener's L1 background, the daily use of English of the speaker and the speaker's gender affected the rhythm of the speaker's utterances. The speakers also tended to accommodate their rhythm to their partners'.

Through this study, we can see that different rhythm metrics brought out different results. For example, the analysis of %V revealed that the speakers who use English more frequently showed more stressed-timed rhythm than those who use English less frequently while the analysis of  $\Delta V$  came up with the opposite result. According to [6], there is a limitation

in explaining the tendency of rhythm with  $\Delta V$ , because the slow speech rate might affect the results of the analysis. In his study, the  $\Delta V$  values of advanced Korean speakers of English and native speakers of English were smaller than those of low-intermediate Korean speakers of English. [6] proposed that speech rate might be used to complement the limitation of some rhythm metrics.

It was revealed from this study that there were some inconsistencies among the results from different rhythm metrics. In order to complement these inconsistencies, speech rate, articulation rate, the percentage of function words in the utterances (%fw) and the number of pauses in the utterances might be used to investigate the rhythm.

This study is meaningful in that it analysed the rhythm of impromptu and spontaneous utterances with actual interaction in conversations between speakers of English with a relatively diverse L1 background in the East Asia region.

## 5. Acknowledgements

The author would like to thank Dr Peggy Mok of the Chinese University of Hong Kong for providing the facilities and recruiting subjects for recording. Thanks also goes to Dr Tae-Jin Yoon of Sungshin University for his help in making scripts to annotate the speech data.

## 6. References

- [1] F. Ramus, M. Nespors, and J. Mehler, J, "Correlates of linguistic rhythm in the speech signal," *Cognition*, vol. 73, pp. 265-292, 1999.
- [2] V. Dellwo, "Rhythm and speech rate: A variation coefficient for delta C," in *Proceedings of the 38<sup>th</sup> Linguistics Colloquium*, pp. 231-241, 2006.
- [3] L. White and S. Mattys, "Calibrating rhythm: first language and second language studies," *Journal of Phonetics*, vol. 35, pp. 501-522, 2007.
- [4] E. L. Low, E. Grabe, and F. Nolan, F, "Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English," *Language and Speech*, vol. 43, pp. 377-401, 2000.
- [5] D. Deterding, "Letter to the editor: The measurement of rhythm: a comparison of Singapore and British English," *Journal of Phonetics*, vol. 29, pp. 217-230, 2001.
- [6] H. Chung, T.-Y. Jang, W. Yun, I. Yun, and J. Sa, "A study on automatic measurement of pronunciation accuracy of English speech produced by Korean learners of English," *Language and Linguistics*, vol. 42, pp. 165-196, 2008.
- [7] Y. Kachru and C. L. Nelson, C. L., *World Englishes in Asian Contexts*. Hong Kong: Hong Kong University Press, 2006.
- [8] D. Deterding and A. Kirkpatrick, "Emerging South-East Asian Englishes and intelligibility," *World Englishes*, vol. 25 no. 3/4, pp. 391-409, 2006.
- [9] A. Kirkpatrick, *English as a Lingua Franca in ASEAN: A multilingual model*. Hong Kong: Hong Kong University Press, 2010.
- [10] J.-M. Kim, S. Flynn, and M. Oh, "Non-native speech rhythm: A large-scale study of English pronunciation by Korean learners," *Studies in Phonetics, Phonology and Morphology*, vol. 13, no. 2, pp. 245-275, 2007.
- [11] H. Chung, "A study on the rhythm of Korean English learners' interlanguage talk," *Phonetics and Speech Sciences*, vol. 5, no. 3, pp. 3-10, 2013.
- [12] T.-Y. Jang, "Speech rhythm metrics for automatic scoring of English speech by Korean EFL learners," *Malsori*, vol. 66, pp. 41-59, 2008.
- [13] S. Kim and T.-Y. Jang, "Rhythmic differences between spontaneous and read speech of English," *Phonetics and Speech Sciences*, vol. 1, no. 3, pp. 49-55, 2009.
- [14] I. Rosenfelder, J. Fruehwald, K. Evanini, and Y. Jiahong, FAVE (Forced Alignment and Vowel Extraction) Program Suite. Retrieved from <http://fave.ling.upenn.edu/> [Philadelphia, PA: University of Pennsylvania Linguistics Lab (Distributor)] on April 20, 2016, 2011.