

What Rhythm Metrics Don't Capture: A First and Second Language Study with Korean and English

Hae-Sung Jeon, Department of Linguistics, University of Cambridge (hsj24@cam.ac.uk)

AIM: Evaluation of the rhythm metrics by their application to Korean, English, and Korean-accented English speech

Rhythm of Speech and Language

- Speech rhythm: a perceptual consequence of the arrangement of strong and weak units
- Morse-code rhythm (stress-timed languages, e.g., English, German) vs. machinegun rhythm (syllable-timed languages, e.g., Spanish, French) [5] [10]
- Reliable rhythmic differences between languages: French newborns discriminated English from Japanese, English/Dutch from Spanish/Italian with only prosodic cues, but did not discriminate English from Dutch [8]

Standard British English

- Typically stress-timed
- Word stress; vowel reduction in unstressed syllables
- Complicated syllable structures

Seoul Korean

- Possibly more syllable-timed than French or Spanish
- No word level stress
- Phrase-level accent: Accentual Phrases with underlying (H/L)HLH patterns [4]
- Relatively simple syllable structure (i.e., (C)(S)V(C)); no phonemic vowel length distinction
- In a perception experiment with *flat-sasasa* type of stimuli, native Italian speakers and Korean speakers could discriminate Korean from Italian and English, but not from Japanese [2]

Quantification of Speech Rhythm

- Acoustic approach: speech divided into consonantal and vocalic intervals e.g., at coach stations [ət kəʊtʃ steɪʃ nɪz] → /ə/, /k/, /t/, /ə/, /tʃ/, /eɪ/, /ʃ nɪz/

Interval Measures [12] and their normalisation [3]

- %V: the percentage of the vocalic interval
- ΔC, ΔV: the standard deviation of consonantal/vocalic interval
- VarcoΔC: (ΔC / mean consonantal duration) × 100
- VarcoΔV: (ΔV / mean vocalic duration) × 100

PVIs (Pairwise Variability Indices) [6]

$$rPVI-C = \left(\sum_{k=1}^{m-1} |d_k - d_{k+1}| \right) / (m-1)$$

$$nPVI-V = 100 \times \left(\sum_{k=1}^{m-1} |d_k - d_{k+1}| / (d_k + d_{k+1}) / 2 \right) / (m-1)$$

m is the number of intervals, d is the duration of the kth interval

Prediction: Stress-timed languages would show higher ΔV, ΔC, VarcoΔV, VarcoΔC, nPVI-V, rPVI-C and lower %V scores than syllable-timed languages.

Rhythm of Second Language Speech

- L2 learners go through transitional stages when their L1 and L2 are rhythmically distinct [11]
- Hypothesis: L2 speakers' rhythm scores would lie between those of the respective L1 groups; advanced group of speakers would show closer scores to native speakers of L2

Experiment

Participants

- L1 groups: Southern British English speakers (Eng), Seoul Korean speakers (Kor), 5 speakers for each group
- L2 groups: Korean learners of English at intermediate level (K-E:int, 5 speakers) and at advanced level (K-E:adv, 7 speakers)
- Five sentences were used from 15 sentences read by each speaker.

Table 1. L2 speakers' profile (the means and standard deviations)

	K-E:int	K-E:adv
Age of first exposure to L2		
	11 years (3.13)	12 years (7.87)
Age of arrival to L2-speaking country		
	31 years (5.96)	18 years (11.18)
Length of residence in L2-speaking country		
	12 months (10.59)	54 months (26.01)
L2 proficiency test [9]		
	17.00/25 (3.39)	22.14/25 (1.46)

Segmentation

- The vocalic and intervocalic intervals were segmented by auditory and visual inspection using Praat [1]
- The first consonantal intervals, the final vocalic intervals and the following consonantal intervals were excluded, due to the unclear boundary of the utterance-initial consonants and the utterance-final lengthening effects.
- Glottalisation periods between successive vowels were excluded, and the durations of the two adjacent vocalic intervals were summed up and counted as one interval.
- Vowel insertions by L2 speakers in English utterances were marked as vocalic intervals.

References

- [1] Boersma, P. & Weenink, D. Praat: doing phonetics by computer. [Computer program], from <http://www.praat.org/>
- [2] Cho, M.-H. 2004. Rhythm typology of Korean speech. *Cognitive Processing*, 5, 249-253.
- [3] Dellwo, V. 2006. Rhythm and speech rate: A variation coefficient for deltaC. In P. Kamrowski, & I. Szigetl (Eds.), *Language and language processing: Proceedings of the 38th linguistic colloquium* (pp. 231-241). Pilszcsaba 2003. Frankfurt: Peter Lang.
- [4] Jun, S.-A. 1996. *The Phonetics and Phonology of Korean Prosody: intonational phonology and prosodic structure*. Garland Publishing Inc., New York: NY.
- [5] Lloyd James, A. (1940). *Speech Signals in Telegraphy*. London: Pitman & Sons.
- [6] Low, E. L., Grabe, E. & Nolan, F. 2009. Quantitative characterisations of speech rhythm: syllable-timing in Singapore English. *Language and Speech*, 43, 377-401.
- [7] Mok, P., & Dellwo, V. to appear. Comparing native and non-native speech rhythm using acoustic rhythm measures: Cantonese, Beijing Mandarin and English. In Proceedings of the 4th Speech Prosody (2009), Campinas, Brazil.
- [8] Nazzi, T., Bortolotti, J. & Mehler, J. 1998. Language discrimination by newborns: toward the understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and Performance*, 24, 756-766.
- [9] Oxford English Placement Test. from http://www.lang.ox.ac.uk/tests/english_placement.html, accessed on 02 May 2006

Results

Rhythm metrics and speaking rate

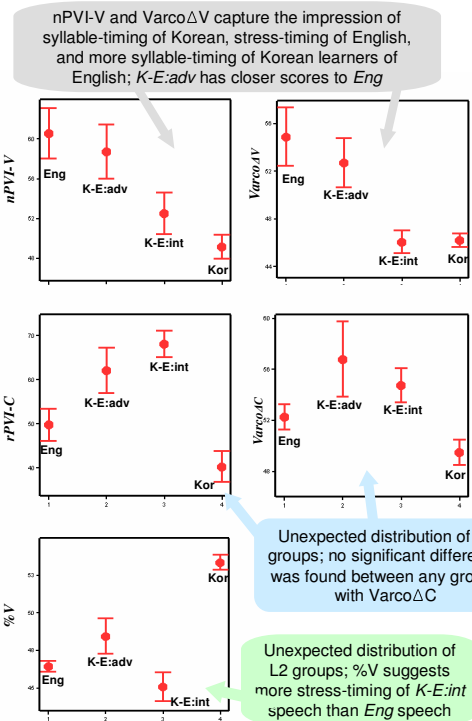
Table 2. Correlations between rhythm metrics and speech rate (spearman's ρ), *p < .05; **p < .01

	%V	ΔV	ΔC	VarcoV	VarcoC	nPVI-V	rPVI-C
E-E	.008	-.589(**)	-.411(*)	-.172	.158	-.127	-.616(**)
K-K	.077	-.044	-.687(**)	.255	-.279	.055	-.655(**)

- Necessity of speaking rate normalisation: non-normalised metrics would suggest more 'syllable-timing' when the speech rate increases

Distribution of the rhythm scores

Figure 1-5. Mean score of each metric. Error bars show mean ± SE.



Summary

Only nPVI-V and VarcoΔV showed the expected distribution of all participant groups, although %V, VarcoΔV, VarcoΔC have been reported as reliable metrics [7] [13].

References (continued)

- [10] Pike, K. L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.
- [11] Wark, B. 1985. Speech Rhythms in Second Language Acquisition. *Language and Speech*, 28, 157-175.
- [12] Ramus, F., Nespor, M., & Mehler, J. 1999. Correlates of linguistic rhythm. *Cognition*, 73, 265-292.
- [13] White, L., & Mattys, S. L. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501-522.

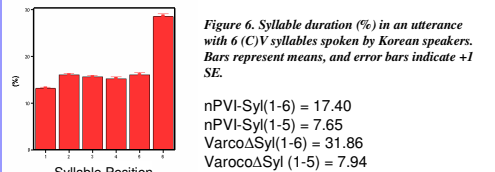
Discussion

What the rhythm metrics do/don't capture

- Rhythm metrics show the final outcome of all segmental and supra-segmental durational effects
- However, prominence in speech comes from the interaction between various acoustic properties (e.g., pitch, loudness).
- Not all long syllables may be perceived as prominent: in Korean, the longest syllable in an utterance is the utterance-final syllable.
- Rhythm metrics do not inform about the prominence distribution and the contrast between strong and weak units that create perceivable speech rhythm.
- The interpretation of L2 data: the rhythm scores of L2 speakers are closer to those of L1 speakers do not necessarily imply native-like speech rhythm of L2 speaker (e.g., the misplacement and the realisation of stress, the larger number of prosodic phrases within an utterance and consequential lengthening)

Methodological Problems

- The treatment of lengthening effects at prosodic boundaries



- e.g., nPVI and VarcoΔSyl calculated with syllable duration differ in a great degree depending on whether the phrase-final syllable is included, particularly in a language where the prominence is not shown by duration
- Finding the best candidates for rhythm units: consonantal metrics with speaking rate normalisation are not sensitive and consequently, the results are not meaningful
- Problems of L2 speakers: the necessity of natural speech elicitation techniques due to the relatively lower competence in reading difficult words

Conclusions

- The use of nPVI-V and VarcoΔV is supported in comparison to other metrics.
- The currently used rhythm metrics may capture durational variability within utterances; however, the prominence distribution is not reflected.
- It is difficult to apply the rhythm metrics with L2 speakers with low proficiency, due to 1) difficulties of finding the source of durational variability and 2) disfluency in reading.

Acknowledgements

The author would like to thank Bob Ladd, Laurence White, Antje Heinrich, and Francis Nolan for discussions.