SPEECH RHYTHM METRICS: A TYPOLOGICAL SURVEY

Matthew K. Gordon, Argyro Katsika, Sherry Chien, Jiyoung Jang, Ryan Ka Yau Lai

University of California, Santa Barbara
mgordon@linguistics.ucsb.edu, argyro@ucsb.edu, sherrychien@ucsb.edu, jiyoung@ucsb.edu, kayaulai@ucsb.edu

ABSTRACT

Several hypothesized parameters diagnosing the distinction between syllable-timing and stress-timing were examined for recordings of the *North Wind and Sun* fable for a set of 10 genetically, areally, and structurally diverse languages. Measured properties included the following: duration of interstress intervals, ratio of stressed-to-unstressed vowel and syllable durations, variability in the duration of intervocalic and vocalic intervals, and the standard deviation of consonant and vocalic intervals. Results suggest greater typological diversity than in previous studies and a lack of convergence across diagnostics typically employed to diagnose speech rhythm. This diversity is plausibly due to interlanguage variation in syllable complexity and word length, the latter of which is attributed to differences between languages in their degree of morphological synthesis. Nevertheless, certain parameters suggest rhythmic categories; however, potentially more than just a binary division between syllable- and stress-timing.

Keywords: rhythm, syllable-timed, stress-timed

1. INTRODUCTION

The study of acoustic correlates of rhythm has been a productive area of research since the identification of a perceptual dichotomy between syllable-timed and stress-timed languages [1, 2]. Research has examined a number of potential quantitative metrics for diagnosing the distinction between the two categories but no single property has been demonstrated to reliably distinguish syllable- and stress-timing across a broad sample of languages. More generally, evidence for a typologically robust perceptual dichotomy between syllable- and stress-timed languages is lacking [3]. Rather, languages fall along a continuum capped at either end by canonical prototypes forming the basis for the hypothesized categorical division, e.g., English (stress-timed) vs. Spanish (syllable-timed). Furthermore, the distance between stresses, the "interstress intervals", tend to vary little across rhythmic types with most intervals ranging from 300-700 milliseconds [4].

This paper aims to enrich the research program on rhythm classes in two ways. First, we extend the typological purview of investigation to include a broader sample of areally and genetically diverse languages than have previously been examined. By expanding the breadth of cross-linguistic coverage, we aim to gain more reliable insight into the range of variation in rhythmicity and the factors that contribute to it. Second, we explore the effects of not only syllable structure examined in previous work (e.g., [5, 6, 7]) but also the length of morphological (grammatical) words on commonly adopted acoustic correlates of rhythm. A language's characteristic word length, calculated here in terms of number of syllables per word, potentially exerts a substantial effect on its rhythmic profile due to the one-to-one mapping between morphological words and stresses in many languages, a mapping that creates relatively long interstress intervals for long words.

2. METHODOLOGY

2.1. Languages and Corpus

A set of 10 genetically and areally diverse languages were selected for which there exists a JIPA Illustration containing a recording of the *North Wind and the Sun* fable. This fable has been used to quantity rhythm in previous studies [7] since it provides a comparable sample (in length and style) of phonetically transcribed naturalistic speech across languages. The languages in our sample all are described as having stress-based prosodic systems but differ in two important respects germane to rhythm. They permit differing degrees of syllable complexity and they represent varied morphological subtypes along a continuum ranging from relatively analytic, e.g., Kalasha, to highly agglutinative, e.g., Yine; this morphological variation leads to corresponding variation in word length.

The list of languages, their geographic locations and genetic affiliations, mean syllable complexity (in number of segments per syllable) and mean morphological word length (in number of syllables per word) is provided in Table 1.

2.2. Measurements

The North Wind and the Sun recordings were parsed into segments, syllables, and sentences using text grids in Praat. Surface syllabification conventions (including epenthetic vowels) were followed and words were parsed following the orthographic breaks; words were thus morphological rather than prosodic. Sentences were divided based on punctuation if marked in the Illustration, and otherwise on the basis of major prosodic breaks. Long vowels and diphthongs were treated as unitary segments while geminate consonants were treated as two segments when split by a syllable boundary.

Language	*Country*	*Family*	*Seg/σ*	*σ/Wd*
Abkhaz (Cwyzhy)	Turkey	N. Caucasian	2.32	2.80
Arrernte (Central)	Australia	Australian	2.08	3.00
Estonian	Estonia	Uralic	2.65	2.00
Gitksan	British Columbia	Tsimshianic	2.74	1.58
Kalasha (Bumburet)	Pakistan	Indo-European	2.19	1.93
Kedayan	Malaysia	Austronesian	2.36	2.14
Nen	New Guinea	Yam	2.33	2.22
Quichua (Salasaca)	Ecuador	Quechuan	2.19	3.07
Seri	Mexico	isolate	2.56	1.53
Yine	Peru	Maipurean	2.41	3.83

Table 1: Languages in the study

Several measurements commonly used to quantify rhythm were extracted. Interstress Intervals (ISI) were measured from the onset of each stressed syllable up to but not including the onset of the next stressed syllable. Primary stress, but not secondary stress, was indicated in the transcriptions for all languages. Thus, only primary stresses were included in the calculation of interstress intervals. Certain texts did not mark stresses for monosyllabic words; these were manually added for content words. In addition, the following measures were calculated, following [7]: the raw Pairwise Variability (rPVI) across Interstress Intervals, the duration ratio of stressed-to-unstressed syllables and vowels, and the standard deviations of the vocalic (ΔV) and consonantal (ΔC) phases of the sentences.

3. RESULTS

3.1. Interstress intervals, word and syllable length

The densest concentration of interstress intervals is similar across languages (Fig. 1), hovering between 300ms and 700ms, with the overall distribution of intervals varying greatly as a function of language. Gitksan, Seri, Abkhaz and Yine tend toward longer interstress intervals, where Yine shows particularly wide variation with several very long intervals. This variability is likely attributed in large part to the relatively large proportion of long words in Yine, which is observed in the interstress intervals by word length in syllables (Fig. 2). Conversely, Arrernte and Quichua show relatively little variation in interstress intervals despite having relatively long words, likely due to the counterbalancing effect of simple syllable structure (Fig. 3).

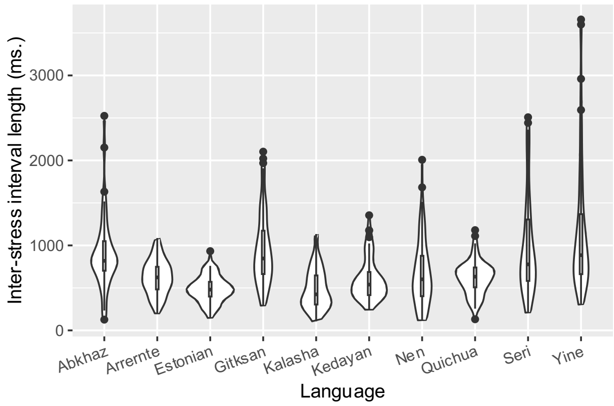

Figure 1: Interstress intervals for examined languages

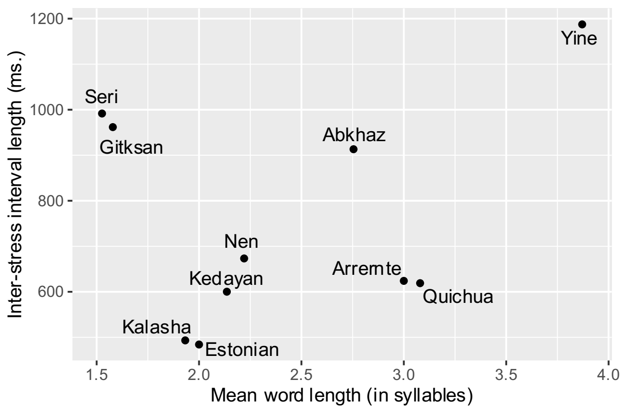

Figure 2: Mean interstress intervals by mean word length

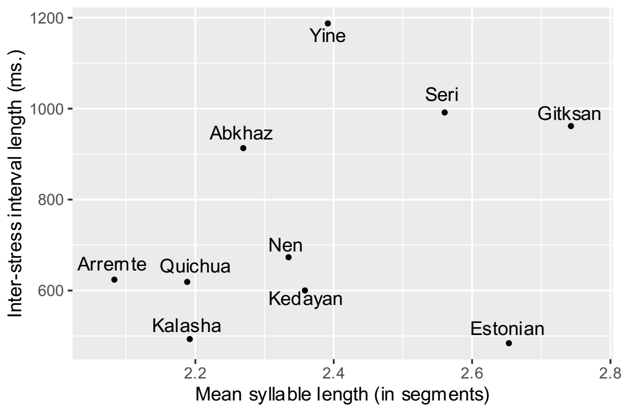

Figure 3: Mean interstress intervals by mean syllable length

The three languages with the characteristically shortest interstress intervals, Estonian, Kalasha and Kedayan, also have relatively short mean word length in terms of syllable count. Estonian is an interesting case since it has relatively complex syllables but short words (despite being a language with moderately synthetic morphology); the short interstress intervals in Estonian thus suggest a greater weight for syllable count in predicting interstress intervals. Gitksan, though, has a similar level of syllable complexity and word length to Estonian but has much longer interstress intervals.

3.2. Relative duration and PVI measures

The duration ratio between the stressed and unstressed syllable intervals within an ISI is shown in Fig. 4 and between stressed and unstressed vowel intervals in Fig. 5. In both figures the average number of syllables per ISI in each language is at the top.

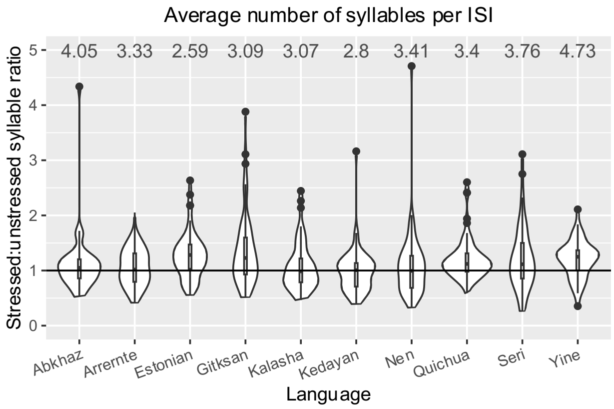

Figure 4: Stressed-to-unstressed syllable duration ratio within interstress intervals

In Fig. 4, most languages have distributions centered just above a one-to-one ratio, indicating that stressed portions of interstress intervals generally comprise more than half of the total duration of the interval. Yine has the longest mean ISI length in number of syllables and also relatively large stressed-to-unstressed interval ratios, and conversely, Kedayan has relatively few syllables per ISI in keeping with also having characteristically short stressed-to-unstressed interval ratios. There is not a consistent relationship, however, between number of syllables in an ISI and the proportion of the ISI that is stressed. Estonian thus has the fewest syllables per ISI but stressed-to-unstressed syllable interval ratios that are among the larger ones for the examined languages. Fig. 5 suggests that the duration ratio observed between stressed and unstressed syllables in Fig. 4 is generally attributed to vowels rather than consonants.

All surveyed languages have many data points reflecting *shorter* stressed than unstressed syllable and vowel intervals within an ISI. Although this finding may seem counterintuitive, it has at least two plausible sources. One is that unstressed syllables are longer due to segmental complexity, i.e., longer consonant clusters in unstressed syllables. Another factor could be final lengthening targeting unstressed domain-final syllables, a possibility given that all of the examined languages have non-final stress either predominantly or exclusively, and final syllables are longer than non-final ones in all languages examined except Quichua.

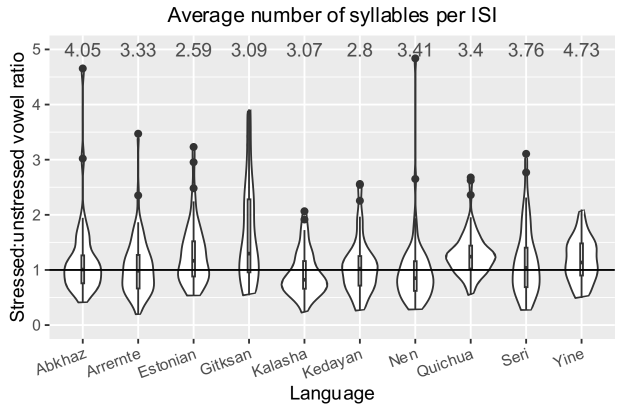

Figure 5: Stressed-to-unstressed vowel duration ratio within interstress intervals

Fig. 6 plots the duration ratio of stressed-to-unstressed syllables for interstress intervals with differing numbers of syllables in order to assess polysyllabic shortening, the tendency for individual segments to shorten in proportion to the number of syllables in an interstress interval [8, 9, 10]. It is difficult to compare languages differing in their word lengths but the clearest tendency toward polysyllabic shortening is observed in Gitksan.

Another diagnostic for the stress-timed vs. syllable-timed distinction is the Pairwise Variability Index, which is greater for stress-timed languages due to their tendency toward unstressed vowel reduction and their greater tolerance of syllable complexity. Fig. 7 plots raw PVI values for intervals of vocalic (i.e., just vowels) and intervocalic (i.e., just consonants) sounds in each language. Gitksan stands out as an outlier in terms of possessing unusually large rPVI values for both vocalic and especially intervocalic interstress intervals. The languages are otherwise well spread across a range of rPVI values for vocalic and intervocalic intervals.

Fig. 8 plots the standard deviation of the duration of consonant and vocalic intervals within each sentence. Gitksan is again a conspicuous outlier in possessing very large ΔC and ΔV values, while other languages again appear to form a continuum, with positive correlation between the two values. Higher vocalic rPVIs tend to be associated with higher ΔV values, and higher intervocalic rPVIs with higher ΔC. For example, Kedayan has high vocalic rPVI and ΔV

and moderate intervocalic rPVI and ΔC, while Yine has relatively high intervocalic rPVI and ΔC and low vocalic rPVI and ΔV.

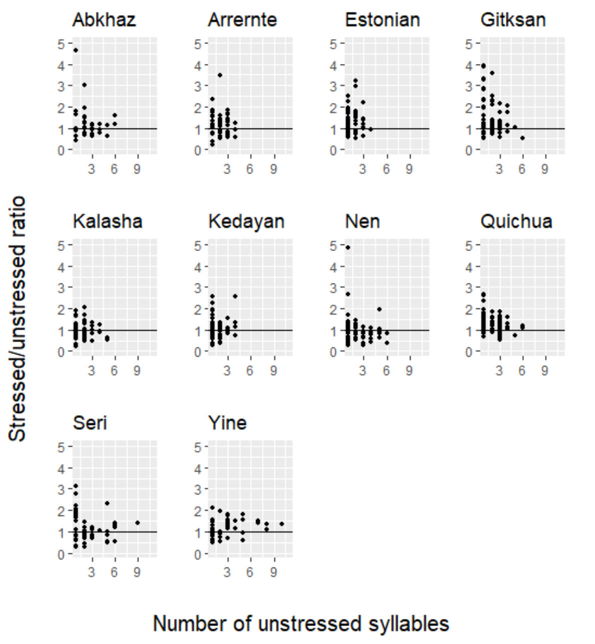

Figure 6: Duration ratio of stressed-to-unstressed syllables by number of syllables in interstress interval

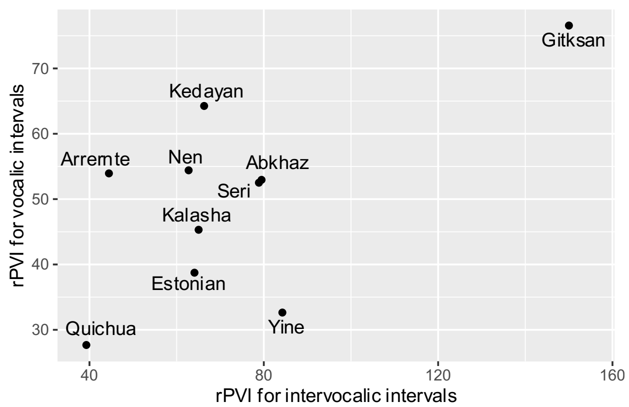

Figure 7: Raw pairwise variability index for intervocalic and vocalic intervals

4. DISCUSSION

Results overall suggest greater divergence between languages in their acoustic diagnostics of rhythm than found in previous research. Interstress interval durations are generally consistent with earlier work [4], although the current study suggests greater variation across and especially within languages, a difference that is likely due to differences in syllable complexity and length of words, the latter of which serves as a proxy for morphological complexity.

Metrics commonly used to quantify the impressionistic dichotomy between stress-timed and syllable-timed languages revealed some distinctions between languages but not consistently convergent ones. Quichua and Gitksan have relatively large durational differences between stressed and unstressed vocalic intervals, suggesting stress-timed rhythm. However, Quichua falls at the opposite end of the continuum from Gitksan in terms of vocalic and intervocalic rPVI, ΔC and ΔV with values that are more consistent with syllable-timing. This divergence may be explained in part by a difference between the two languages in the locus of complexity: Quichua has long words but relatively simple syllable structure whereas the opposite is true of Gitksan. Similar word length, however, does not ensure parallel rhythmic metrics since Yine, which has the longest words in the sample, diverges from Quichua in intervocalic rPVI.

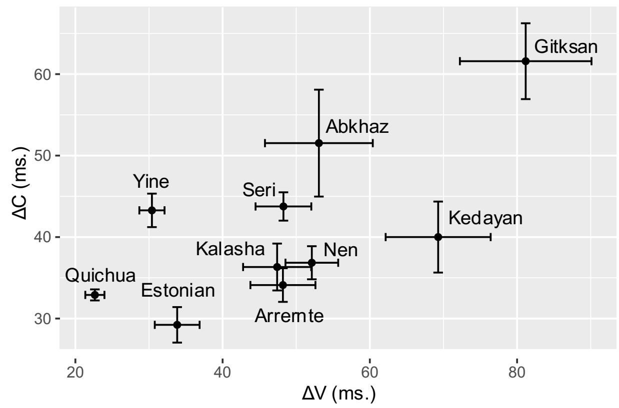

Figure 8: Mean ΔC vs. ΔV values with Student's *t*-test std. errors

The vocalic and intervocalic rPVI, ΔC and ΔV parameters do not show any clear clustering of languages, but a continuum, although there is a correlation between ΔC and ΔV observed. Some languages have low values for all four measures, a single language has extremely high indices for all four, and other languages have intermediate values. While the values for languages with low measures on all four are consistent with those characteristic of syllable-timed languages [7], values associated with the other languages are consistent with a stress-timed designation, suggesting that stress-timing might occupy a larger perceptual space, presumably attributed to the relative homogeneity (compared to the current sample) of languages forming the basis for the original impressionistic classification.

5. SUMMARY

An acoustic study of rhythm metrics in a structurally, genetically, and areally diverse set of languages indicates commensurate heterogeneity in the acoustic diagnostics of rhythm. Two categories, or even two ends of a unidimensional continuum, may be insufficient to capture the rhythmic typology.

6. ACKNOWLEDGMENTS

We thank the students (Gia Ahn, Gabriel Bernal, Jason Galaviz, Yan Lashchev, Katie Lee, Yujie Li, Jade Morton, Ella Peter, Kirra Sebastian, Lily Yin, Alice Zhang, Boyi Zheng, Dingyan Zhou) who assisted with the data analysis as well as the anonymous reviewers who provided feedback on the earlier version of the paper.

7. REFERENCES

[1] Pike, K. L. 1945. *The Intonation of American English*. University of Michigan Press.
[2] Abercrombie, D. 1967. Elements of General Phonetics. Edinburgh University Press.
[3] White, L., Malisz, Z. 2020. Speech rhythm and timing. In: Gussenhoven, C., Chen, A. (eds.), *Oxford Handbook of Language Prosody*. Oxford University Press, 167-182.
[4] Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.
[5] Dasher, R., Bolinger D. 1982. On pre-accentual lengthening. *Journal of the International Phonetic Association* 12, 58-71.
[6] Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3), 265-292.
[7] Grabe, E., Low, E. L. 2002. Durational variability in speech and the Rhythm Class Hypothesis. In: Gussenhoven, C., Warner, N. (eds), *Laboratory Phonology* 7. De Gruyter Mouton, 514-546.
[8] Lindblom, B., Rapp, K. 1973. *Some temporal regularities of spoken Swedish*. Papers from the Institute of Linguistics, University of Stockholm 21.
[9] Meyer, E.A., Gombocz, Z. 1909. *Zur Phonetik der ungarischen Sprache*. Berling.
[10] Kim, H., Cole, J. 2005. The stress foot as a unit of planned timing: evidence from shortening in the prosodic phrase. *Interspeech* 2005, 2365-2368.