

Predicting prosody in poetry and prose

Greg Kochanski, Anastassia Loukina, Elinor Keane, Chilin Shih[†], and Burton Rosner
Oxford University Phonetics Lab
[†]University of Illinois and the Beckman Institute

Rhythm is expressed by recurring, hence predictable, beat patterns. Poetry in many languages is composed with attention to poetic meters while prose is not. Therefore, one way to investigate speech rhythm is to evaluate how prose reading differs from poetry reading via a quantitative method that measures predictability.

Our corpus consisted of 42 paragraphs read in each of Standard Modern Greek, Parisian French, Southern British English, Standard Russian, and Taiwanese Mandarin. Every reader also read 4 poems consisting of 8-12 lines. Where possible, we selected children's poetry with a regular metrical pattern that could easily be read with a strong rhythm.

We built a specialized speech recognition system, based on the HTK toolkit, that produced a sequence of **C** (consonantal), **V** (vowel-like) and **S** (silence/pause) segments. Once the segment boundaries were defined, five acoustic properties were computed for each segment: duration, loudness, friction, the location of the segment's loudness peak, and the rate of spectral change. We then computed 1085 linear regressions to predict these properties in terms of the preceding 1 to 7 segments.

For the regressions over the entire corpus, values of Pearson's r^2 varied widely: some regressions explain a negligible 2% of the total variance, and others up to 43%. Duration was the least predictable property: on average, r^2 was only 8%. This is noteworthy because all the published rhythm measures are based only on duration. The best predictable property was the rate of spectral change.

Overall, poetry was much more predictable than prose (r^2 values are roughly twice as large and our method allowed predicting up to 79% of variance). This is consistent with the intuition that poetry is more 'rhythmical'. We also observed that poetry was more predictable across long ranges than prose. While in prose the mean difference between r^2 for the regressions based on 1 and 7 preceding segments was 6%, in poetry this difference was 25%. Given that all poetry in our corpus had regular metrical pattern, this confirms that the long-range effects we observe are likely to be related to such linguistic units as feet.

The predictability of a language depends on what is being predicted and the context of the target phones, so we anticipate that there will be at least several different ways to characterize the rhythm of each language. We propose that this approach could form a useful method for characterizing the statistical properties of spoken language, especially in reference to prosody and speech rhythm.