Predicting Prosody in Poetry and Prose

Greg Kochanski, Anastassia Loukina, Elinor Keane, Chilin Shih, and Burton Rosner

* We define rhythm as the degree to which acoustic properties can be predicted, given prior context.

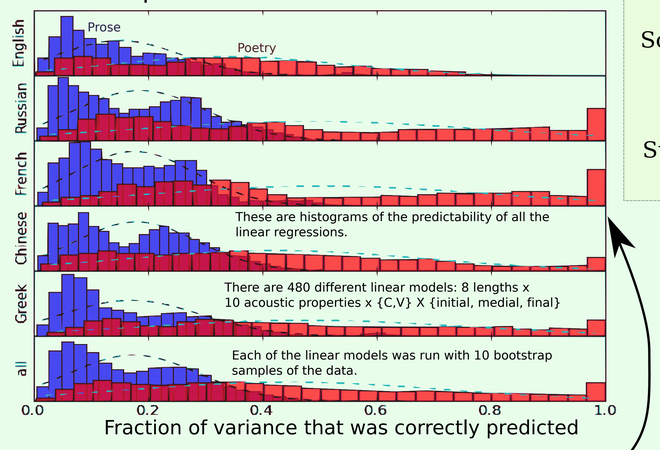
* Rhythm is expressed by recurring patterns, and recurring patterns are predictable from history.

* Much poetry has a meter, a rhythmic beat. We test our idea by comparing the predictability of poetry to that of prose.

* To measure predictability, we use linear regressions and measure the fraction of variance that they correctly predict.

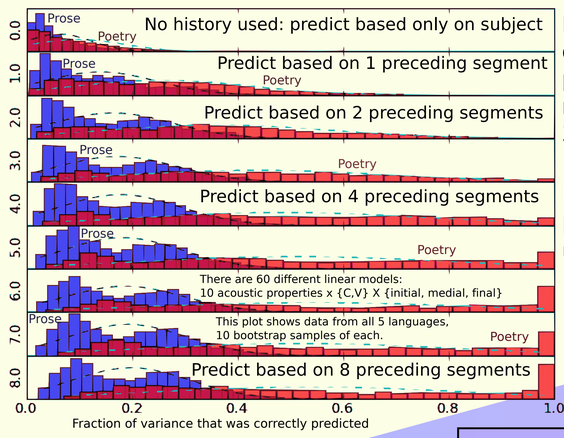
* 5 Languages

In all languages studied, acoustic properties of poetry are much easier to predict than those of prose.



Languages:
Southern British English
Standard Russian
Parisian French
Taiwanese Mandarin
Standard Modern Greek

We thank the ESRC for primary support via RES-062-23-1323; C. Shih thanks the US NSF.



For prose, acoustic properties depend only on the immediate past: predictions do not improve much as the model uses more than 3 segments worth of input.

For prosody, the accuracy of the prediction continues to increase up to 8 or more prior segments.

Some of these linear regressions provide nearly perfect prediction of certain acoustic properties in certain contexts.

Sort segments into type (i.e. **C** or **V**) and context (i.e. phrase-initial, -medial, or -final). Phrases are defined by **S**ilence segments.

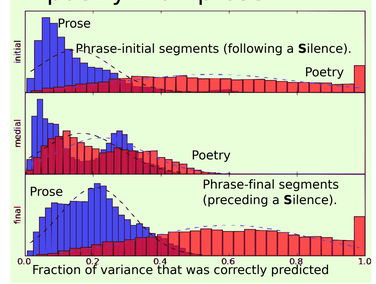
For each segment in the group, use the acoustic properties for the N preceding segments (i.e. 10N independent variables) as predictors. Subject identity is a further predictor.

Build a separate linear model for each of the 10 acoustic properties.

For each linear model, fit it to 10 bootstrap samples of the data.

For each individual fit, compute the ratio of the variance explained by the model to the variance before the fit. Add this to a histogram.

Factoid:
* In poetry, phrase-initial and -final segments are much more predictable than medial.
* Medial segments are only modestly more predictable in poetry than prose.

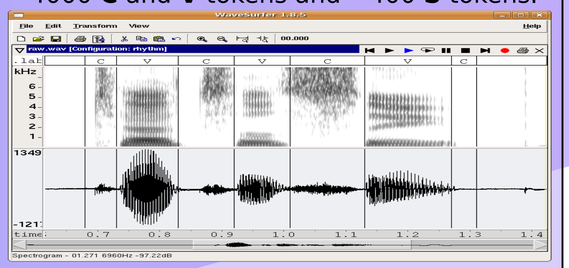


Prose sub-corpus: 41 paragraphs read in each language (69 shorter paragraphs in Mandarin).

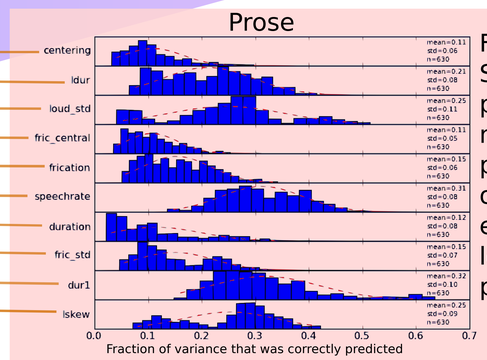
Poetry sub-corpus: 4 poems consisting of 8-12 lines. Typically, there were 8 syllables per line. Most of the poetry was either iambic or trochaic. Where possible, we selected children's poetry with a regular metrical pattern that could easily be read with a strong rhythm.

Ten subjects per language (except 24 for English, 9 for Greek).

Segment the data into **C**onsonantal, **V**ocalic regions and **S**ilences. We built a bespoke language-independent ASR system with the HTK toolkit. The system was trained on manually segmented data: ~4000 **C** and **V** tokens and ~400 **S** tokens.



Compute 10 acoustic properties for each segment:
1) Is the segment louder at the beginning or end?
2) How much does the spectrum change?
3) Standard deviation of partial voicing
4) Log of duration
5) Local estimate of the speech rate.
6) Partial voicing / partial frication - average
7) Partial voicing / partial frication at center
8) Standard deviation of loudness
9) Loudness times duration
10) How close is the segment to the utterance's average speech sound?



Factoid:
Some acoustic properties are much more predictable than others, but everything is at least moderately predictable.

* Poetry is much more predictable than prose. This is consistent with the intuition that poetry is more `rhythmical'.

* Some properties of poetry are almost perfectly predictable.

* Long (i.e. 6-8 segment) predictors improve the prediction of poetry, but for prose, they were little better than shorter (i.e. 3-4 segment) predictors. These long predictors may be related to feet and other supersegmental structures.

* This could be a new approach to characterizing the statistical properties of