

CLASSIFYING TEXT BY RHYTHM AND SOUND

Nina Wang

MOTIVATION

- Some literary works have a defined rhythmic pattern:
 - Shakespeare's iambic pentameter
 - Homer's use of hexameter
- Do other types of writing also have rhythm and melody?
 - Poetry
 - News articles
 - Academic writing

˘ / ˘ / ˘ / ˘ ˘ / ˘ / ˘ / ˘ / ˘ /
To be | or not | to be | that is | the question

MELODY IN WRITING

“This sentence has five words. Here are five more words. Five-word sentences are fine. But several together become monotonous...

Now listen. I vary the sentence length, and I create music. *Music. The writing sings. It has a pleasant rhythm, a lilt, a harmony.* I use short sentences. And I use sentences of medium length. And sometimes when I am certain the reader is rested, I will engage him with a sentence of considerable length, a sentence that burns with energy and builds with all the impetus of a crescendo, the roll of the drums, the crash of the cymbals—sounds that say listen to this, it is important.

... Don't just write words. Write Music.”

-Gary Provost

RESEARCH QUESTIONS

1. Do different genres of writing “sound” different? Can we quantify that?
2. Can we use the “sound” of a text to identify its genre?

QUANTIFYING “SOUND”

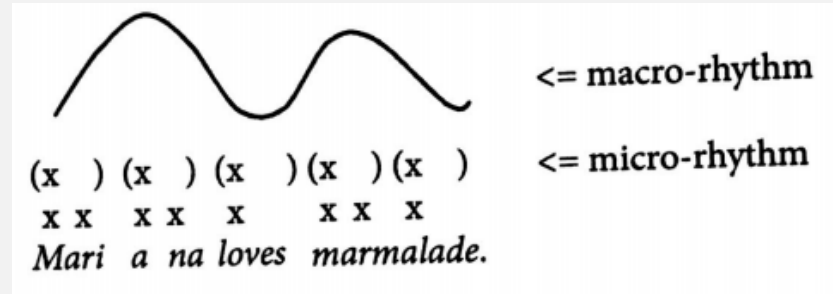
1. **Micro-rhythm:** Regular alternation of strong & weak beats

We're / áll in / túitively fa / míliar with the i / déa of a de / rívative.

- Split sentences by stresses → evenly spaced feet

2. **Macro-rhythm:** Regular alternation of high & low tones (pitch)

- Proposed by Sun-Ah Jun
- Split by H/L peaks → evenly-spaced hills



3. **Melody:** Variation in the structure

PROCESS

1. Transform texts into audio streams using Amazon Polly, a text-to-speech conversion software
 - Dataset: novels, poetry, and academic papers
2. Process audio files with Praat, to view information about features like pitch and intensity
3. Analyze the three “sound” metrics in the audio files.
4. Use established machine learning techniques (SVM) to classify writing based on its rhythmic/melodic features.

RESULTS

- N/A

CONCLUSIONS

- N/A