Prosody

Prosody is the stress, rhythm, and intonation of speech.*

*According to Wikipedia, today.
Today’s paper:

Comparing the rhythm and melody of speech and music: The case of British English and French.

1. Introduction on language and music
2. Methods for quantifying prosody
3. Experimental design
4. Results and discussion
5. More discussion
Language and Music

“Music is the universal language of mankind.”
—Henry Wadsworth Longfellow (1807–1882)

“Good music is very close to primitive language.”
—Denis Diderot (1713–1784)

“Music expresses feeling and thought, without language; it was below and before speech, and it is above and beyond all words.”
—Robert Ingersoll (1833–1899)

silly quotes from: http://www.quotegarden.com/music.html
Language and Music

Previous debate:

- What is the informational content / communicative power of music?

- If music is a language, what constitutes translation?

- If music is a language, what is analogous to a word, phoneme, and sentence?
Language and Music

Current study:

- Are there similarities between how music and language are patterned rhythmically?

- Are differences between languages reflected in different national styles?
Language and Music

**Previous studies:**

- Are there similarities between how music and language are patterned rhythmically?  **Yes!**

- Are differences between languages reflected in different national styles?  **Yes!**

**New issues:**

- What about melodic patterns?
- Are rhythmic differences due merely to greater variability in vowel duration
Timing in speech

1. Stress-timed
   - English, German, Swedish, Thai

2. Syllable-timed
   - French, Spanish, Finnish, Chinese

3. Mora-timed
   - Japanese, Luganda
Vowel reduction

Stress-timed languages may have a greater degree of vowel reduction:

“Where are you from?”

“Was he lying?”

“Interrogate them.”

“I’m from LA.”

“I was truthful.”

“Interrogate them.”
Vowel length variability

<table>
<thead>
<tr>
<th></th>
<th>shortest</th>
<th>longest</th>
</tr>
</thead>
<tbody>
<tr>
<td>English:</td>
<td>“a”</td>
<td>“strengths”</td>
</tr>
<tr>
<td>French:</td>
<td>“a”</td>
<td>“Schtroumpfs”</td>
</tr>
<tr>
<td>Japanese:</td>
<td>“a”</td>
<td>“kya”</td>
</tr>
</tbody>
</table>
Outline

1. Introduction on language and music
2. Methods for quantifying prosody
3. Experimental design
4. Results and discussion
5. More discussion
Quantifying prosody

Two features: rhythm and pitch

image: http://www.sfu.ca/sonic-studio/handbook/Graphics/Spectrograph2.gif
Quantifying prosody

\[ nPVI = \frac{100}{m - 1} \times \sum_{k=1}^{m-1} \left( \frac{d_k - d_{k+1}}{d_k + d_{k+1}} \right) \]

\[ \%V = \text{percent vowels} \]
Quantifying prosody

VI = Vocalic Interval

vowel duration

CV = σ / μ

pitch height
Quantifying prosody

- vowel duration CV’ed
- pitch height CV’ed
- nPVI (durational contrast)
- pitch interval CV’ed
Outline

1. Introduction on language and music
2. Methods for quantifying prosody
3. Experimental design
4. Results and discussion
5. More discussion
Data sets

20 sentences in each language
Data sets

> 130 themes in each style, from > 5 composers

FIG. 2. Examples of duration and pitch coding of musical themes. (D122: Debussy’s Quartet in G minor for Strings, 1st movement, 2nd theme. E72: Elgar’s Symphony No. 1, in A Flat, Opus 55, 4th movement, 2nd theme.) D122 illustrates duration coding: the relative duration of each note is shown below the musical staff (see text for details). E72 illustrates pitch coding: each note is assigned a pitch value based on its semitone distance from A4 (440 Hz). The nPVI of note durations in D122 is 42.2. The coefficient of variation (CV) of pitch intervals in E72 is 0.79.
FIG. 2. Examples of duration and pitch coding of musical themes. (D122: Debussy’s *Quartet in G minor for Strings*, 1st movement, 2nd theme. E72: Elgar’s *Symphony No. 1, in A Flat, Opus 55*, 4th movement, 2nd theme.) D122 illustrates duration coding: the relative duration of each note is shown below the musical staff (see text for details). E72 illustrates pitch coding: each note is assigned a pitch value based on its semitone distance from A4 (440 Hz). The nPVI of note durations in D122 is 42.2. The coefficient of variation (CV) of pitch intervals in E72 is 0.79.
Rhythm and pitch encoding

FIG. 1. Illustration of the prosogram, using the British English Sentence “Having a big car is not something I would recommend in this city” as uttered by a female speaker. In both graphs, the horizontal axis along the top shows time in seconds, the vertical axis shows semitones re 1 Hz (an arrow is placed at 150 Hz for reference), and the bottom shows IPA symbols for the vowels in this sentence. The onset and offset of each vowel is indicated by vertical dashed lines above the vowels’ IPA symbol. (a) Shows the original F0 contour, while (b) shows the prosogram. In this case, the prosogram has assigned level tones to all vowels save for the vowel in “car,” which was assigned a glide. Note that the pitches of the prosogram do not conform to any musical scale.
1. Introduction on language and music
2. Methods for quantifying prosody
3. Experimental design
4. Results and discussion
5. More discussion
Rhythm features

TABLE II. nPVI and CV for speech and music (mean and s.e.). The rightmost column gives the probability that the observed nPVI difference is due to the difference in CV.

|          | English nPVI | French nPVI | \( p \) | English CV | French CV | \( p \) | \( p(\Delta nPVI|\Delta CV) \) |
|----------|--------------|-------------|------|------------|-----------|------|-----------------|
| Speech   | 55.0 (3.0)   | 35.9 (1.8)  | <0.001 | 0.55 (0.03) | 0.36 (0.02) | <0.001 | 0.01            |
| (vowels) |              |             |       |            |           |      |                 |
| Music    | 47.1 (1.8)   | 40.2 (1.9)  | <0.01  | 0.61 (0.02) | 0.58 (0.02) | 0.34 | 0.001           |
| (notes)  |              |             |       |            |           |      |                 |
FIG. 3. The relationship between CV and nPVI for speech (a, b) and music (c, d). For speech each dot represents one sentence; for music each dot represents one theme. The best fitting regression line for each panel is also shown. English speech: nPVI = 23.5 + 57.3 × CV, \( r^2 = 0.34 \), df = 18, \( p = 0.03 \); French speech: nPVI = 11.7 + 66.5 × CV, \( r^2 = 0.43 \), df = 18, \( p < 0.01 \); English music: nPVI = 26.6 + 33.9 × CV, \( r^2 = 0.14 \), df = 134, \( p < 0.001 \); French music: nPVI = 6.2 + 59.1 × CV, \( r^2 = 0.36 \), df = 178, \( p < 0.001 \). For the musical data, hatched lines show the lower possible limit of the nPVI and CV at 0 on each axis: the axes range into negative numbers for display purposes only, so that the points at (0,0) can be clearly seen. Themes with a score of 0 for nPVI and CV have notes of a single duration. There were two such English themes and eight such French themes.
nPVI difference due to variability difference?

FIG. 4. Result of Monte Carlo analysis for English vs French speech. The actual nPVI difference between the two languages in this study (19.1 points) is shown by an arrow. See text for details.
Melodic features

<table>
<thead>
<tr>
<th></th>
<th>English Pitch CV</th>
<th>French Pitch CV</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch</td>
<td>0.71 (0.04)</td>
<td>0.75 (0.04)</td>
<td>0.32</td>
</tr>
<tr>
<td>height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch</td>
<td>0.88 (0.05)</td>
<td>0.68 (0.03)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>intervals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Music</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch</td>
<td>0.69 (0.01)</td>
<td>0.71 (0.01)</td>
<td>0.14</td>
</tr>
<tr>
<td>height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch</td>
<td>0.76 (0.02)</td>
<td>0.71 (0.02)</td>
<td>0.03</td>
</tr>
<tr>
<td>intervals</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE III. Pitch variability in speech and music, measured in terms of pitch height (the CV of pitch distances from the mean pitch of a sequence) or pitch intervals (the CV of pitch interval size within a sequence). Mean and s.e. are shown.
Melodic features

FIG. 5. Pitch interval variability in English and French speech and music. Pitch interval variability is defined as the CV of absolute interval size between pitches in a sequence. Error bars show standard errors.
Rhythm and Melody together

**FIG. 6.** Rhythm-melody (RM) space for speech and music. Axes are nPVI and MIV. Error bars show standard errors.
Rhythm and Melody together

FIG. 7. nPVI and MIV values for individual composers. Error bars show standard errors. Note the almost complete separation of English and French composers in RM space, despite large overlap between the nationalities along either single dimension.
Findings:

• Difference in vowel-length variability cannot account for difference in nPVI, either in speech or in music.

• Little note-length variability difference in music.

▶ Therefore, relationship between music and language based on nPVI, not CV.
Findings:

• Language and musical styles differ on the basis of interval size variability.

• Despite this, average interval size the same in each.

• Composers well separated in RM space.
Outline

1. Introduction on language and music
2. Methods for quantifying prosody
3. Experimental design
4. Results and discussion
5. More discussion
Future work

• Methodology very general

‣ Apply to more languages!

‣ Thai and Chinese: both two tonal languages with very different nPVIs.
Music: Y U No the same?
Music: Y U No the same?

• Indirect route:
  • vocal music → folk music → Nationalist art music

• Direct route:
  • statistical learning of prosody patterns → unconscious influence
RM space

FIG. 6. Rhythm-melody (RM) space for speech and music. Axes are nPVI and MIV. Error bars show standard errors.
RM space

- Look for more interactions between rhythm and melody
Quantifying non-native prosody

• Potential application in language learning:
  • Automatic estimate of degree of non-native prosody
  • Requires correlating non-native prosody and rhythmic patterns
    • (hint: vowel duration variance)
More dimensions

• Use loudness as a third feature

• (hint: intensity variability separates English and French speech)

FIG. 6. Rhythm-melody (RM) space for speech and music. Axes are nPVI and MIV. Error bars show standard errors.
Automating the gruntwork
Thank you!