Prosodic Characterization of Reading Styles using Audiobooks Corpus

Michael Proctor, Athanasios Katsamanis

http://sail.usc.edu

Method: Quantifying Prosody

Audio samples prepared for further analysis by forced-alignment phonetic transcription of each complete recording sampled in the listener survey.

- companions' texts sourced from LibriVox, Project Gutenberg [22, 23]
- forced alignment using Salabign: adaptive, iterative speech recognition & text alignment facilitating processing of audiobook-length speech recordings, and robust to transcription errors [14]
- transcriptions and interval timings generated at sentence, word, and phone-based levels of analysis

To compare the prosodic characteristics of each reader's speaking style, metrics were calculated for each test and reader including:

- percentage of vowels or vocatric intervals (SVL)
- coefficient of variation of vocatric intervals (VarCoV)
- coefficient of variation of intervocalic intervals (VarCoC)
- normalized pair-wise variability index (PVI)

Results: Reader Prosody

Individual listener's preferences differ, but overall, clear preferences emerge:

<table>
<thead>
<tr>
<th>Reader</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
<th>M8</th>
<th>M9</th>
<th>M10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader A</td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
<td>F4</td>
<td>F5</td>
<td>F6</td>
<td>F7</td>
<td>F8</td>
<td>F9</td>
<td>F10</td>
</tr>
<tr>
<td>Reader B</td>
<td>F1</td>
<td>F2</td>
<td>F3</td>
<td>F4</td>
<td>F5</td>
<td>F6</td>
<td>F7</td>
<td>F8</td>
<td>F9</td>
<td>F10</td>
</tr>
</tbody>
</table>

Method: Perceptual Characterization

Preferences for reading styles evaluated by asking listeners to evaluate speech samples from different readers, using a head-to-head comparison paradigm:

- list x 10-second speech samples extracted at random intervals from audio recordings of each reader to be evaluated
- recordings taken from two works of a single author (Jack London) of standard 20th Century American English [12, 13]
- auditors: 13 native speakers of General American English

Listeners compared all readers by auditioning 3 random samples of each reader, juxtaposed against 3 samples of each other reader

- forced choicelol preference decision task
- hierarchy of readers constructed from cumulative rankings of listener preferences

Characterizing Read Speech

- differences in the realization of read speech (cf. spontaneous): [1-17]
- higher F0, more F0 variation, more F0 declination
- lower speech rate + longer pauses
- longer major tone units
- less shimmer, less vowel reduction

Results: Listener Preferences

<table>
<thead>
<tr>
<th>Reader</th>
<th>Male Reader Rankings</th>
<th>Female Reader Rankings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader A</td>
<td>F1</td>
<td>F2</td>
</tr>
<tr>
<td>Reader B</td>
<td>M1</td>
<td>M2</td>
</tr>
</tbody>
</table>

Future Directions

- broader survey of reading styles:
- more listeners
- more samples within and across literary genres
- control for specific prosodic and extra-prosodic factors through selection or manipulation of reading voices
- cross-language listener comparisons: native speakers of syllable-timed vs. foot-timed languages
- more sophisticated metrics capable of capturing subtle and segmental features of speech in multiple dimensions

Acknowledgements

Research supported by NH Grant ROI DCO6T124-01