Given the multiplicity of acoustic cues to prosody,

Q1) What is the variation and uniformity across speakers in the phonetic implementation of prosody?

Q2) What are the underlying production mechanisms of prosody?

Q3) How are listeners affected by speaker variability in their interpretation of prosody?

Background

• Prior studies employing controlled ‘laboratory’ speech (e.g. simple sentences, read speech) show acoustic effects of prosody in many languages, and also show that native listeners respond to acoustic prosodic cues in interpreting utterance meaning. [Ref: 2005. Shattuck-Hufnagel, S., 2005; and Hufnagel, S.]

• Few studies attempt (1) to approximate prosody production and perception in everyday speech communication (Shattuck-Hufnagel, S., 2005; and 2) to directly examine speaker-dependent variability in the acoustic encoding of prosody. [Ref: Shattuck-Hufnagel, S., 2005.]

Rapid Prosody Transcription (RPT)

1. 54 short excerpts (~11 – 58 sec) were selected from the Buckeye corpus of American English spontaneous speech [Ref: 2007.]

2. Orthographic transcriptions were produced for each sound file, with no punctuation or capitalization.

3. 97 transcribers: UIUC undergraduates, untrained and unfamiliar with the phonetics and phonology of prosody.

4. After being given simple instructions and definitions of prominence and boundary, four groups of 12-20 subjects marked the locations of prosodic prominence and boundaries on the printed transcriptions in a separate task in real time, based only on auditory impression (no visual speech display).

5. Collecting the transcription data from all subjects, each word in the set of excerpts was assigned probabilistic P prominences) and B (boundary) scores depending on the number of transcribers who marked the word as prominent or as followed by a juncture.

Reliability Tests

Distribution of P- and B-scores

<table>
<thead>
<tr>
<th>Speaker</th>
<th>P-scores</th>
<th>B-scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker 26</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

- RPT prosody annotation is reliable across transcribers by Fleiss’ multi-rater’s kappa agreement scores.
- Higher agreement for boundary than for prominence.

Acoustic measurements

- Extracted from the lexically stressed vowels for prominence and the word final lexically stressed vowels for boundary.
- Stress indented according to the ISLE dictionary. [Ref: 2007.]
- Normalized acoustic measures:
- temporal measures (Vdur and pause)
- intensity measures (overall and subband intensities in 0-0.5, 0.5-1.0, 1.0-2.0, and 2.0-4.0 kHz)
- F0 measures (local F0 maximum and F0 at the right edge)
- formant measures (F1) "Y2 measures are not reported here."

Spearman’s non-parametric correlation analysis

- Prominence:
  - Across speakers, prosodic prominence is positively correlated with a large subset of the acoustic parameters and prosodic phrase boundary is positively correlated with temporal measures (V, duration and silent pause).
  - Speaker variability:
    - Across speakers, there is no single acoustic parameter that is correlated with P-scores and the strength of correlation is not uniform.
    - Speakers vary in which acoustic parameters other than temporal parameters are correlated with B-scores and the direction and the strength of correlation is not uniform and varies across speakers.

Speakers-dependent regression models of prosody

- Speaker-dependent regression models account for 12 - 54% of variation in listeners’ response to prominence and 18 - 87% of variation in boundary perception by speaker.
- Speakers vary in the set of cues used to encode prominence. Some speakers rely more on F0 and duration as cues, but other speakers rely heavily or exclusively on intensity.
- The temporal measures of vowel duration and silent pause play a primary role in cueing boundaries across speakers. Yet, one speaker does not rely on temporal parameters at all and some speakers employ other acoustic parameters to cue for prosodic phrase boundary.

Discussion and Conclusion

1. Uniformity comes from the fact that:
   - the phonetic characteristics are enhanced when a word is prominent and therefore, most or all acoustic parameters change in the direction of enhancing phonetic distinctiveness.
   - speech tempo reduces in the vicinity of a prosodic phrase boundary, which directly results in the elongation of the pre-boundary vowel and the following pause.

2. Variability comes from the fact that:
   - speakers vary in their acoustic encoding of prominence in terms of the correlation strength
   - speakers vary in their acoustic encoding of boundary in terms of both the strength and direction
   - the contribution of each acoustic cue to listeners’ prosody perception varies depending on speakers.

3. Prominence is manifested through the general acoustic enhancement of speech. Boundary is manifested through slowing down speech.

References


Acknowledgements

This research is supported by NSF grant IIS-0702924 to Jennifer Cole and Mark Hasegawa-Johnson. Special thanks to the Prosody-RPT group members for their comments.