Prosody perception by naïve listeners: Evidence from a large multi-transcriber reliability study

Yoonsook Mo, Jennifer Cole, Eun-Kyung Lee

The Beckman Institute and the Department of Linguistics
University of Illinois at Urbana-Champaign

Q1) How do ordinary listeners perceive the location of prosodic boundaries and prominences?
Q2) How can we measure the reliability of prosodic transcription across multiple listeners?

Prosodic features play an important role in communication,
1. A prosodic boundary demarcates a speech chunk, which is usually a semantically coherent unit of speech.
2. A prosodic prominence highlights a word or phrase in speech, signaling its information contribution.

Goals of this study:
1. To find an appropriate measure of transcription reliability between naïve transcribers in a task of real-time prosody transcription
2. To look at variation across listeners and across speakers

Prior studies test the consistency of prosody perception across transcribers, but with certain limitations.
2. Transcribers: few, trained in prosody and phonetics (Yoon et al. 2004)
3. Procedure (Buhmann et al. 2002; Yoon et al. 2004)
   – Aided by visual inspection
   – Complex annotation scheme
   – Replays allowed as many times as listeners want
4. Analysis: simple % agreement scores or Cohen’s pairwise agreement statistic.

Methodology

Materials
1. Selected from the Buckeye corpus of American English spontaneous speech (Pitt et al. 2007)
2. 38 short excerpts (about 20 sec): 19 speakers * 2 excerpts each
3. Sound files are blocked into two groups, one for prominence labeling and the other for boundary labeling, and then randomized within a group.
4. Materials are prosodically transcribed by 74 subjects, undergraduates at UIUC.

<table>
<thead>
<tr>
<th># of transcribers</th>
<th>Exp. 1</th>
<th>Exp. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>prominence</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>boundary</td>
<td>16</td>
<td>20</td>
</tr>
</tbody>
</table>

Procedures
1. Simple definitions of prominence and boundary and short instructions are provided
2. Subjects see a transcript of each excerpt without any punctuation or capitalization
3. Subjects listened twice to sound files in real time, through headphones, with no visual speech display.
4. Half of the listeners listened for prominence first and the other half listened for boundary first.
5. Listeners marked up a transcript of each excerpt
   – Word word word
   – Word | word | word

Results

Probabilistic Prominence and Boundary scores

Distribution of boundary and prominence intervals across listeners.

Distribution of kappa scores for boundary and prominence, as calculated for each pair of listeners.

Discussion and conclusion
1. Statistically significant Fleiss’ multi-rater agreement scores confirm that listeners’ responses are systematic and relate to linguistic information in the utterance.
2. Higher agreement scores for boundary than for prominence show that listeners are more consistent in their perception of boundary location than for prominence location.
3. Speaker-dependent variation in the perception of prosody indicates that speakers vary in how they structure an utterance prosodically and/or in how effectively they cue prosodic structures in the speech signal.
4. Listener-dependent variation indicates that listeners differ in their sensitivity to the locations of prosodic prominence and boundary, for the same set of speakers.

Acknowledgements
This study is supported by NSF IIS-0309284 and 04-0674147. We would like to thank Steve Winter, Zaki Hilstrom, our participants and the members of the Beckman Institute Prosody and ASR research group for their comments.

References