

Listeners' perceptions of phrasing produced by native and non-native readers in English

Caroline L. Smith and Paul Edmunds

University of New Mexico

caroline@unm.edu, pedmunds@unm.edu

The method of "Rapid Prosody Transcription" (e.g., Cole, Mo & Baek 2010) has been used to access listeners' perceptions of spontaneous speech prosody. Listeners hear an audio recording while following along on an orthographic, unpunctuated transcript. They indicate their perception of phrasal boundaries or prominent words by marking them on the transcript in real time. Here we adapt this technique to examine native English listeners' perceptions of prosodic phrasing as produced by native and non-native English speakers. Given the importance of phrasing in communication, it could be useful to evaluate non-native speakers' success at producing it. Comparison of phrasing by speakers with different language backgrounds can help to distinguish properties of boundaries that are language-specific from those that characterize many or all languages. Our prediction was that since non-native speakers do not fully control the cues used by native English speakers to indicate boundaries, we expected listeners to show a lower rate of agreement when responding to the speech of non-natives. Also, the non-native speakers seemed to have frequent hesitations that listeners might interpret as boundaries. We thus expected our listeners to mark phrasal boundaries more frequently in non-natives' speech. Listeners responded to recordings of the Rainbow Passage read by twelve native speakers of American English and twelve non-native speakers whose native language is Spanish. Eleven listeners listened to recordings of the native speakers, and a different eleven listeners heard the recordings of the non-natives. For all but two speakers, there were reading errors in the recordings. These were transcribed accurately in the transcripts, so the number of words in each recording varied from a minimum of 329 to a maximum of 382. The non-native speakers averaged 351 words in 149 s, the natives 345 words in 116 s.

Results. The data were examined in two ways: by comparing listeners' responses to different speakers, in order to see if the speakers differ from each other, and by comparing different listeners' patterns of responding, to detect individual differences. We first compare responses between the native and non-native groups of speakers. As predicted, listener agreement was lower for non-natives' speech. The mean kappa for listeners who heard native speakers was 0.91; the mean with non-native speakers was 0.87. This difference was significant in a t-test. In addition, listeners marked significantly more boundaries in recordings of non-native speakers (mean 33.6) than for native speakers (mean 27.6). This difference was also significant.

Another analysis examined how many listeners marked boundaries at different locations in the recordings. Very similar numbers of locations were marked by all or almost all listeners for both speaker groups, but there were more locations in the non-native readings where only a few listeners marked boundaries. As a result there were more different locations in the non-natives' readings where at least one listener marked a boundary, which is consistent with the finding that listeners marked more frequently (but less consistently) in the non-natives' recordings.

Conclusions. Native English listeners were less consistent in their perceptions of phrasal boundaries in reading aloud by non-native speakers of English than by other native English speakers. They tended to hear more boundaries in the non-natives' speech, but at many locations just a few listeners marked a boundary. The many short pauses in the non-natives' productions may have confused listeners as to when a speaker truly intended a boundary. The real-time nature of the task may challenge listeners who must decide how to interpret any break in the flow of speech, a difficulty that is particularly acute in listening to less fluent non-natives. The methodology provides a simple, effective means of comparing different groups of speakers' effectiveness at communicating prosodic phrasing, but can only offer hints about the source of differences among them.