

The role of syntactic structure in guiding prosody perception in spontaneous speech  
Jennifer Cole, Yoonsook Mo and Soondo Baek  
University of Illinois at Urbana-Champaign

Prosodic phrasing serves to break continuous speech into smaller, meaningful chunks and defines an interface between phonology and other dimensions of linguistic organization, including syntax and discourse structure. This paper reports on a large-scale study of prosody in spontaneous speech that explores the relationship between prosody and other aspects of the linguistic structure of utterances, and the reflexes of prosody in the acoustic signal. One goal of this study is to develop a method for studying prosody perception in spontaneous speech that can be used to determine how prosody perception relates to the structural properties of the utterance, and to the acoustic prosodic features. Through a comparison of prosody transcriptions (representing perception) with acoustic measures (representing production) and syntactic features of a given utterance, this study asks if prosody perception is primarily signal-driven (sensitive to acoustic cues), or if the listener's perceptual experience is influenced by his/her awareness of the syntactic structure of an utterance. A finding that syntactic information influences prosody perception quasi-independently of the presence of acoustic cues to prosody would support the view that prosody perception reflects the listener's online processing of speech, and/or that prosody perception is biased by expectations driven by the syntactic parse of the utterance.

This work is based on data from our study of naïve prosody transcription. Untrained transcribers listen to excerpts from the Buckeye corpus of American English interview speech (Pitt et al. 2007) and mark up a printed transcript for the location of boundaries between "chunks" of speech, as judged from their auditory impressions. The results from about 20 transcribers per excerpt yield a probabilistic boundary score (B-score) for each word. Counting boundaries with a B-score above the mean, we consider the intervals between boundaries as representative of perceived prosodic phrases. Analysis of the syntactic structure of the words before and after a perceived prosodic boundary reveals strong evidence that perceived prosodic phrases are anchored in syntactic structure. 45% of perceived phrases comprise a major syntactic constituent and another 28% are phrases consisting of a single word, discourse marker, or filled pause. 16% of perceived phrases are syntactic fragments, most of which consist of two or more complete syntactic constituents. We also observe a bias favoring syntactic alignment of the left edge of a perceived prosodic phrase relative to right edge alignment, and further edge asymmetries in the more frequent location of conjunctions at the left edge of a perceived phrase than at the right edge. Acoustic correlates of prosody (vowel duration, intensity, spectral emphasis and pause) reveal significant effects of perceived prosodic boundaries, especially in duration measures and pause, and regression analyses confirm a significant correlation between acoustic measures and boundary scores (e.g., the duration of the stressed vowel of a word is positively correlated with its score). These findings suggest that transcribers are sensitive to the acoustic correlates of prosody, and tend to mark prosodic boundaries in locations where speakers produce durational cues. Analyses underway explore the relationship between acoustic measures and syntactic features to determine if the syntactic features function at all independently of information from the acoustic signal to influence the listener's perception of prosody. Preliminary evidence suggests that prosody perception is partly driven by listeners' expectations based on recognition of the words in the utterance and their syntactic properties. The overall finding is that prosody perception is both signal-driven and expectation-driven, and syntactic processing is one factor influencing listeners' expectations of prosodic phrasing.