BOUNDARY PERCEPTION AS INTERPLAY BY ACOUSTIC CUES, SYNTACTIC CONSTITUENCY, AND PROMINENCE

Jianjing Kuang, May Pik Yu Chan, Nari Rhee

University of Pennsylvania
kuangj@sas.upenn.edu, pikyu@sas.upenn.edu, nrhee@sas.upenn.edu

ABSTRACT

Boundary perception is a complex process related to multiple factors, such as acoustic boundary cues, syntactic constituency, and prominence, however, empirical studies are still lacking to explore how they work together in determining listeners’ boundary perception; and it is unclear whether these factors carry the same weights in different languages. This study aims to address these questions by investigating the boundary perception rates of utterances from continuous corpora in Mandarin and English. Boundary and prominence ratings were collected with a crowd-sourcing perception experiment. The relative strength of the syntactic boundary of both the left and right sides of the constituents was extracted from the syntactic parsing annotations. A wide range of acoustic cues of both prosodic domain-final and domain-initial positions were examined. Results confirmed that boundary perception is shaped by complex interactions among acoustic cues, syntactic parsing, and prominence perception. More importantly, there is cross-linguistic variation between Mandarin and English.

Keywords: speech prosody, boundary perception, prominence, syntax-prosody interface

1. INTRODUCTION

Being able to parse speech into meaningful junctures is an important step in speech processing by human listeners. Boundary perception is a rather complex process that is related to several levels of factors, such as prosodic cues, syntactic structure, and information structure. However, there are still relatively few empirical studies to explore how these factors work together to determine the location and strength of the perceived boundaries. This study intends to provide important evidence for a better understanding of the effects of acoustic boundary cues, syntactic constituency, and prominence on boundary perception.

It has been well-established that prosodic boundaries correlate with a number of acoustic cues. For example, domain-final positions can often be indicated by pauses, final lengthening [1, 2, 3, 4], pitch declination and final lowering [5, 6, 7, 8], creakier voice quality [4, 9, 10, 11]; while domain-initial positions are correlated with domain-initial strengthening and pitch reset [12, 13]. However, although these acoustic cues are widely used among languages, their relative cue weighting may vary across languages.

Prosodic phrasing has a tight link with syntactic parsing. Studies have shown prosodic boundaries are useful for locating syntactic boundaries [14, 15], and resolving syntactically ambiguous sentences [16, 17]. However, the effects of syntax on prosodic boundary perception are much less well understood, though [18, 19] suggest that syntactic cues significantly influence the detection of prosodic boundaries. It is also not clear how the left and right edges of the syntactic constituents align with the prosodic domains (c.f.[20]) in different languages.

Moreover, prominence and boundary as the two important prosodic aspects have an interdependent relationship. Prominence has effects on boundary perception, and vice versa. For example, phrase-final position appears to be a preferred position for prominence perception. Some studies [21, 22] found that phrase-final words are more likely to be perceived to be prominent; with similar effects being found for Spanish, French, and English [21], despite the three languages being quite different in terms of their phrasal prosody. It remains to be seen whether this domain-final effect is universal across other languages as well. When resolving syntactically ambiguous sentences, English speakers treat the prominent words as the end of the syntactic phrase [17]; by contrast, Mandarin listeners parse prominence as the beginning of the phrase [23]. Therefore, although the interdependence between prominence and boundary appears to be common among languages, it is possible that the direction of the alignment between prominence and boundary varies across languages. More cross-linguistic empirical studies are needed to clarify these issues.

This study aims to address these questions
by investigating the boundary perception rates of utterances from continuous corpora in Mandarin and English. Boundary and prominence ratings were collected with a crowd-sourcing perception experiment. The relative strength of the syntactic boundary of both the left and right sides of the constituents was extracted from the syntactic parsing annotations. A wide range of acoustic cues of both prosodic domain-final and domain-initial positions were examined.

2. METHODS

2.1. Participants and materials

Sentences from syntactically-parsed read speech corpora were used as the stimuli for the experiments for both Mandarin and English. For English, a female native speaker’s reading of Jane Austen’s Emma was obtained from LibriVox [24]. We chose a female speaker’s recording of Volume II Chapter 10 of the book for our study, a chapter also in the 2nd edition of The Penn Parsed Corpus of Modern British English (PPCMBE2) [25], which contains Penn Treebank-style annotated brackets. For Mandarin, readings of news articles by one male and one female speaker were retrieved from the Chinese Tree Bank speech corpus [26]. The text of this corpus is based on Chinese Tree Bank 9.0 [27], which included news article texts parsed and annotated with Penn Treebank-style labeled brackets.

A total of 76 participants were recruited for the perception study, including 47 native speakers of English (18-25 years; 29 female) and 29 native speakers of Mandarin (18-35 years; 18 female). All were recruited from the university student community and completed the experiment for partial course credit. From the aforementioned corpora, sentences of around 30 seconds or less were selected for the study (24 for English, 22 for Mandarin). Participants listened to the sentences in their native language and were asked to identify and annotate prominent words and prosodic boundaries. There were up to 382 potential boundaries and 408 potential prominent words in the English condition, and 458 potential boundaries and 480 potential prominent words in the Chinese condition that participants could rate.

2.2. Rapid Prosody Transcription Task

We conducted a listening experiment where participants provided ratings of prominences and boundaries as in the Rapid Prosody Transcription task used in [18, 28], and the boundary detection task in [19]. Participants heard selected sentences one at a time while reading the transcription that was displayed simultaneously. Once the participant entered a trial, the audio started automatically, though participants could replay the audio as many times as they wished. Participants were asked to select where they “think there are boundaries” and where they “think there are prominent words” within the sentence. Participants could not select the beginning or end of the sentence as boundary markers, though any word in the sentence could be selected as a prominent word. The experiment was administered on Qualtrics.

Before the experiment, participants first heard three sample recordings involving the same sentence read with different prosodic focus and structures as examples. This was to familiarize participants with the experimental interface. No further explanations for what constitutes a “boundary” or “prominent” word were given. For both prominence and boundaries, the response rate was calculated by the response count divided by the maximum number of corresponding prominence/boundary responses in a given language condition.

2.3. Syntactic parsing and acoustic measurements

Syntactic annotations following the Penn Treebank guidelines were available for both the Chinese Treebank 9.0 [27] and Penn Parsed Corpus of Modern British English (PPCMBE2) [25], where syntactic parsings for the chosen sentences were extracted. The numbers of left and right brackets between each pair of consecutive words were used as a proxy for the depth of the syntactic structure, capturing the additive strength of syntactic boundaries at the left and right edges of the constituent structure. By the design of the annotation guidelines, each word is wrapped by at least one left and one right bracket. To normalize the varied sentence lengths, the number of left or right brackets was divided by the maximum number of left or right brackets in the sentence.

The chosen sentences were aligned using the HMM-based Mandarin forced-aligner [29] and the Penn Phonetics Lab Forced Aligner [30] for Mandarin and English respectively. We took a range of acoustic measurements in Praat, including: (1) Whether there is a pause at the boundary junction as determined by whether silent portions were identified by the corresponding aligners; (2) the average syllable duration of each word; (3) the minimum and maximum fundamental frequency (F0), (4) the minimum and maximum sound
4. Speech Prosody

were random intercepts for talkers. We computed variables were also included in the model, as acoustic, syntactic, and prominence perception way interactions between pause, language, and all other main effect were included. And three-way interactions between pause and every other main effect, as well as between language and all three main effects, as between language and pause. As for syntactic constituency, significant main effects were found for both left ($\beta = 9.77, p = 0.013$) and right ($\beta = 32.76, p < 0.0001$) syntactic brackets, and for both languages, right brackets have a stronger effect. Therefore, both Mandarin and English align the right edge of syntactic boundaries with prosodic boundaries. Moreover, the syntactic effects are modulated by pause and language, indicated by two-way interactions between pause and brackets (left: $\beta = 31.18, p = 0.012$, right: $\beta = 29.04, p = 0.002$), between brackets and language (right: $\beta = -23.27, p < 0.0001$), as well as...
three-way interactions among brackets, pause and language (left: $\beta = -36.87, p = 0.011$). As illustrated in Figure 3, the effect of right syntactic brackets is stronger for English listeners, and for Mandarin listeners, syntactic boundaries largely co-vary with pause. Overall, syntactic cues play more important roles in detecting prosodic boundaries for English listeners, and pause plays a more important role for Mandarin listeners.

Finally, Figure 4 illustrates the effects of perceived pre-boundary prominence and post-boundary prominence on boundary perception. With English as the reference level, significant main effects of both prominence perception of the previous ($\beta = 0.19, p < 0.0001$) and following word ($\beta = 0.09, p = 0.009$) were found. There were also significant interactions between the pre-boundary prominence and language ($\beta = -0.22, p < 0.0001$). These results indicate that, for English listeners, prominence perception of the pre-boundary word has a positive effect on boundary perception, the effect of which appears stronger when accompanied by a pause. And there is a much weaker effect of the post-boundary word on boundary perception. Mandarin patterns differently from English, showing much weaker effects of prominence perception. More importantly, post-boundary prominence instead of pre-boundary prominence is positively correlated with boundary perception, showing the opposite direction from English. In other words, prominence marks the end of the domain for English listeners, whereas signals the beginning of the domain for Mandarin. This result is consistent with [23, 17] with a different experiment paradigm.

4. CONCLUSION

This study examined the effects of acoustic boundary cues, syntactic constituency, and perceived prominence on boundary perception by investigating the boundary perception rates of utterances from continuous corpora in Mandarin and English. Results confirmed that boundary perception is shaped by complex interactions among acoustic cues, syntactic parsing, and prominence perception. Importantly, how these factors contribute to prosodic phrasing is language specific, as we show Mandarin and English pattern differently at various levels: 1) Pause is the determinant cue for Mandarin listeners, and final lengthening is not expected when a pause is present; but for English, final lengthening and reset effects are stronger when there is a pause. 2) English overall has stronger syntactic effects than Mandarin; syntactic boundaries largely co-vary with pauses in Mandarin. 3) Perceived prominence marks the right boundary for English, but indicates the left boundary for Mandarin. These findings open up more questions about the interface between prosody and other linguistic processing, such as how information structure and syntactic heading are encoded in prosody for different languages.
5. REFERENCES


