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 pressure level (SPL), (5) the alpha ratio (the level difference between the 1k-5kHz region and the 50-1kHz region), (6) L1-L0 (the level difference between the F1 region (defined as 300-800 Hz) and the F0 region (defined as 0-300 Hz), and (7) Cepstral Peak Prominence-Smoothed (CPPS). (3)-(7) were taken from voiced segments of the word. (5)-(7) as measures of voice quality are discussed in [31].

The data was then restructured such that each observation corresponded to a boundary, and acoustic measurements of the pre-/post-boundary words were reformalized as follows: The continuous acoustic measurements were first z-scored by speaker, then a difference in acoustic measure between pre- and post-boundary word was calculated except for (1) and (2). To capture the effects of phrase initial strengthening, the minimum of the pre-boundary word was subtracted from the maximum of the post-boundary word for F0 (3) and SPL (4) measurements. We also subtracted the acoustic measurements' mean value of the pre-boundary word from the post-boundary word for measures (5)-(7). We coded pause (1) as a binary grouped variable instead of a continuous measure as pause thresholds were defined arbitrarily by the aligners. Furthermore, the effect of pause was unlikely to be linear as most words do not have pauses between them (86% for English, 79% for Mandarin).

3. RESULTS AND DISCUSSION

We evaluate the effect of various variables on boundary perception using a linear mixed effects model in R in the lme4 package. Boundary perception ratings as a dependent variable were included in the model as the number of boundary responses for each token divided by the maximum number of boundary responses in a given language (numeric). We included the main effects of the proportion of pre- and post-boundary prominence responses in each given language (numeric), the pre- and post-boundary word's average syllable duration, F0, SPL, alpha, L1-L0 and CPPS differences (numeric), the proportion of left and right brackets (numeric), presence of pause (0 or 1) and language (2 levels: English [reference level], Mandarin). Two-way interactions between pause and every other main effect, as well as between language and every other main effect were included. And three-way interactions between pause, language, and all the acoustic, syntactic, and prominence perception variables were also included in the model, as were random intercepts for talkers. We computed p-values via Satterthwaite's degrees of freedom method using the lmerTest package.

As expected, a number of acoustic cues contribute to the boundary perception; importantly, the effects of acoustic cues are modulated by pause and language, as there are significant two-way interactions between acoustic cues (post-boundary syllable duration ($\beta = 8.33, p = 0.004$), F0 difference ($\beta = 12.35, p < 0.0001$)) and pause, and three-way interactions between language, acoustic cues (pre-boundary syllable duration ($\beta = -16.56, p < 0.0001$), F0 difference ($\beta = -8.80, p = 0.008$), SPL difference ($\beta = 7.00, p = 0.015$), L1-L0 difference ($\beta = 7.59, p = 0.029$) and pause. As illustrated in Figure 1, when there is no pause, longer pre-boundary syllable duration is related to higher rates of perceived boundaries for both languages; but when there is a pause, unlike English listeners, Mandarin listeners expect shorter pre-boundary syllables. As shown in Figure 2, for both languages, reset effects are generally stronger when there is a pause, but the effect size is greater for English listeners. In general, the presence of pauses is the dominant acoustic cue for Mandarin listeners, as exemplified by significant interaction effects between pause and language ($\beta = 62.18, p < 0.0001$).

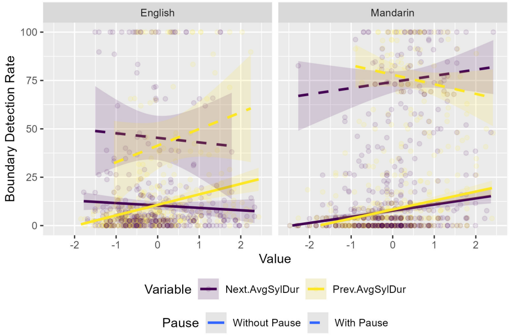

Figure 1: The relationship between average syllable duration and boundary perception.

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

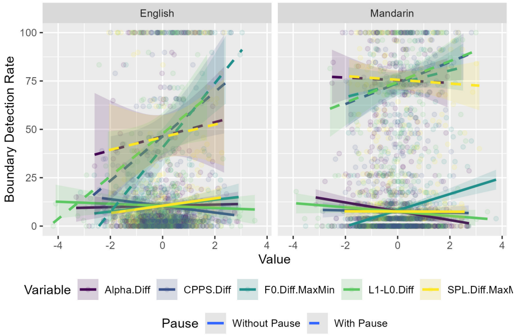

Figure 2: The relationship between other acoustic measures and boundary perception.

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.

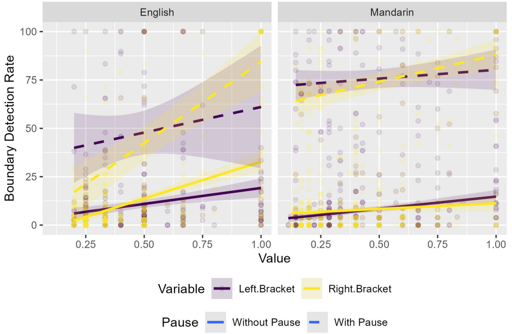

Figure 3: The relationship between syntactic brackets and boundary perception.

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.

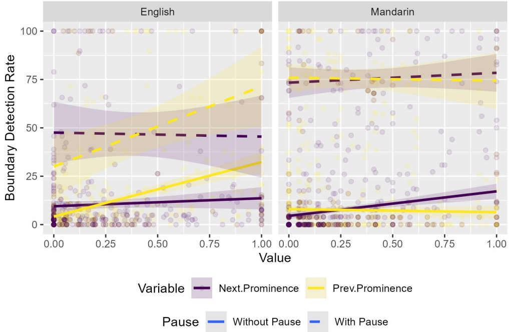

Figure 4: The relationship between promiennce and boundary perception separated by presence of pause.

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

[1] S.-A. Jun, *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press on Demand, 2007, vol. 1.

[2] D. R. Scott, "Duration as a cue to the perception of a phrase boundary," *The Journal of the Acoustical Society of America*, vol. 71, no. 4, pp. 996–1007, 1982.

[3] A. E. Turk and S. Shattuck-Hufnagel, "Multiple targets of phrase-final lengthening in American English words," *Journal of Phonetics*, vol. 35, no. 4, pp. 445–472, 2007.

[4] S. Chavarria, T.-J. Yoon, J. Cole, and M. Hasegawa-Johnson, "Acoustic differentiation of ip and IP boundary levels: Comparison of l-and ll% in the switchboard corpus," in *Speech Prosody 2004, International Conference*, 2004.

[5] J. B. Pierrehumbert, "The phonology and phonetics of English intonation," Ph.D. dissertation, Massachusetts Institute of Technology, 1980.

[6] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in Japanese and English," *Phonology*, vol. 3, pp. 255–309, 1986.

[7] B. Connell and D. R. Ladd, "Aspects of pitch realisation in Yoruba," *Phonology*, vol. 7, no. 1, pp. 1–29, 1990.

[8] D. R. Ladd, *Intonational phonology*. Cambridge University Press, 2008.

[9] E. Bird and M. Garellek, "Dynamics of voice quality over the course of the English utterance," in *Proceedings of the 19th International Congress of Phonetic Sciences*, 2019, pp. 2406–2410.

[10] J. Slifka, "Some physiological correlates to regular and irregular phonation at the end of an utterance," *Journal of voice*, vol. 20, no. 2, pp. 171–186, 2006.

[11] J. Kuang, "The influence of tonal categories and prosodic boundaries on the creakiness in Mandarin," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. EL509–EL515, 2018.

[12] P. Keating, T. Cho, C. Fougeron, and C.-S. Hsu, "Domain-initial articulatory strengthening in four languages," *Phonetic interpretation: Papers in laboratory phonology VI*, pp. 143–161, 2004.

[13] T. Cho, J. M. McQueen, and E. A. Cox, "Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English," *Journal of Phonetics*, vol. 35, no. 2, pp. 210–243, 2007.

[14] W. E. Cooper and J. Paccia-Cooper, *Syntax and Speech*. Harvard University Press, 1980.

[15] D. Watson and E. Gibson, "The relationship between intonational phrasing and syntactic structure in language production," *Language and Cognitive Processes*, vol. 19, no. 6, pp. 713–755, 2004.

[16] A. J. Schafer, S. R. Speer, P. Warren, and S. D. White, "Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task," *Approaches to studying world-situated language use*, pp. 209–225, 2005.

[17] S.-A. Jun and J. Bishop, "Priming implicit prosody: prosodic boundaries and individual differences," *Language and speech*, vol. 58, no. 4, pp. 459–473, 2015.

[18] J. Cole, Y. Mo, and S. Baek, "The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech," *Language and Cognitive Processes*, vol. 25, no. 7-9, pp. 1141–1177, 2010.

[19] A. Buxó-Lugo and D. G. Watson, "Evidence for the influence of syntax on prosodic parsing," *Journal of Memory and Language*, vol. 90, pp. 1–13, 2016.

[20] E. Selkirk and T. Shen, "Prosodic domains in Shanghai Chinese," *The phonology-syntax connection*, vol. 313, p. 337, 1990.

[21] J. Cole, J. I. Hualde, C. L. Smith, C. Eager, T. Mahrt, and R. N. de Souza, "Sound, structure and meaning: The bases of prominence ratings in English, french and Spanish," *Journal of Phonetics*, vol. 75, pp. 113–147, 2019.

[22] J. Bishop, G. Kuo, and B. Kim, "Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from rapid prosody transcription," *Journal of Phonetics*, vol. 82, p. 100977, 2020.

[23] J. Kuang, "Prosodic grouping and relative clause disambiguation in Mandarin," in *Eleventh annual conference of the international speech communication association*, 2010.

[24] H. McGuire, "Librivox," Aug 2005. [Online]. Available: https://librivox.org/

[25] A. Kroch, B. Santorini, and A. Diertani, "The Penn Parsed Corpus of Modern British English (PPCMBE2)," *Philadelphia: Department of Linguistics, University of Pennsylvania.*, 2010.

[26] J. Kuang, M. P. Y. Chan, N. Rhee, M. Liberman, and H. Ding, "The mapping between syntactic and prosodic phrasing in English and Mandarin," *Proc. Interspeech 2022*, pp. 3443–3447, 2022.

[27] N. Xue, X. Zhang, Z. Jiang, M. Palmer, F. Xia, F.-D. Chiou, and M. Chang, "Chinese Treebank 9.0 LDC2016T13," *Philadelphia: Linguistic Data Consortium*, 2016.

[28] J. Cole, T. Mahrt, and J. Roy, "Crowd-sourcing prosodic annotation," *Computer Speech & Language*, vol. 45, pp. 300–325, 2017.

[29] J. Yuan, N. Ryant, and M. Liberman, "Automatic phonetic segmentation in Mandarin Chinese: Boundary models, glottal features and tone," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 2539–2543.

[30] J. Yuan, M. Liberman *et al.*, "Speaker identification on the SCOTUS corpus," *Journal of the Acoustical Society of America*, vol. 123, no. 5, p. 3878, 2008.

[31] A. E. da Silva Antonetti, V. V. Ribeiro, A. G. Brasolotto, and K. C. A. Silverio, "Effects of performance time of the voiced high-frequency oscillation and lax vox technique in vocally healthy subjects," *Journal of Voice*, 2020.