Interspeech 2022
18-22 September 2022, Incheon, Korea

The mapping between syntactic and prosodic phrasing in English and Mandarin

Jianjing Kuang[1], *May Pik Yu Chan*[1], *Nari Rhee*[1], *Mark Liberman*[1], *Hongwei Ding*[2]

[1] Department of Linguistics, University of Pennsylvania, USA
[2] Speech-Language-Hearing Center, Shanghai Jiao Tong University, China

kuangj@ling.upenn.edu, pikyu@sas.upenn.edu, nrhee@sas.upenn.edu, myl@cis.upenn.edu, hwding@sjtu.edu.cn

Abstract

To achieve a better understanding of the relationship between syntactic parsing and prosodic phrasing in speech production cross-linguistically, we investigated how syntactic constituents map onto a high-dimensional acoustic space of prosodic phrasing in two read-speech corpora of Mandarin and English with syntactic annotations. The left and right edges of the constituents from the syntactic parsings were used as a proxy for the relative strength of the syntactic boundaries. A wide range of acoustic cues capturing pauses, duration cues, F0, energy, and voice quality cues were extracted. Our results showed that there is a clear correlation between the strength of syntactic boundary and prosodic phrasing, and the syntax-prosody mapping is much stronger for the right boundaries than for the left boundaries. Moreover, the prosodic realization of syntactic boundaries is gradient (especially for right boundaries), and acoustic cues scale up or down collectively to indicate different extents of phrasing, rather than being specific to certain levels of phrasing. We discuss the findings' implications in relation to the prosodic hierarchy and the nature of the prosody-syntax interface.

Index Terms: Prosodic phrasing, Prosody-Syntax interface, syntactic parsing

1. Introduction

Prosody plays important roles in speech parsing and understanding. For human speakers, prosody helps with locating the major syntactic boundaries [1, 2, 3], resolving syntactic ambiguity [4, 5], and syntactic bootstrapping during language acquisition [6]. For speech technology, prosody helps improve the accuracy in parsing [7]. Even though a close link between syntactic boundaries and prosodic boundaries has been recognized in prosodic theories [8, 9, 10, 11, 12], the nature of the mapping between prosodic boundaries and syntactic boundaries remains an unsettled question. In particular, the mapping between syntax and prosody appears to be non-isomorphic, and mismatches between syntax and prosody are very common. This challenge is related to the representation of prosodic structure and syntactic structure.

On the one hand, traditionally, prosodic phrasing hierarchy has been defined as discrete categories [13, 14]. Since many acoustic cues (e.g., duration) of prosodic boundaries are rather gradient in nature, it is often up to debate how many phrasing levels are in the hierarchical structure and how each level is defined acoustically. Both Mainstream American English and Standard Mandarin are considered to have at least three levels of prosodic phrasing: prosodic word, intermediate phrase (ip), and intonational phrase (IP) [13] (Mandarin is proposed to have two levels of intermediate phrases [15]), where each level of prosodic phrasing is defined with specific boundary cues. For example, in English, prosodic words are expected to have no final lengthening and no pause. Between the two higher prosodic hierarchies, IP is primarily associated with pause, and ip is mostly associated with final lengthening. However, studies with non-expert annotators suggest the agreement for boundary detection is not high, especially for intermediate levels of phrasing [16].

On the other hand, syntactic phrasal structures are often modelled via a tree structure of branching XP's (X-Phrases), where X stands for the syntactic category of the head component (e.g. Noun, Preposition, Verb, Complementizer, etc.), and are thus defined as a hierarchy of categories. However, due to the recursiveness and compositionality of syntactic phrases, an XP of a certain syntactic category can be found at both higher and lower levels of the syntactic tree. Conceivably, the syntactic categories of the phrases do not have a good correlation with prosodic boundaries. Indeed, as noted by [17] and [18], almost any syntactic edge is a potential location for a prosodic phrase boundary. Due to the recursiveness, a better proxy for measuring the strength of the syntactic boundary is the depth of the constituency trees (i.e. whether the boundary is located at a major branch split).

Moreover, the syntax-prosody mapping is likely to be language-specific. At the syntax level, [10, 19] proposed that languages differ parametrically as to whether the left or right edges (or both) of syntactic constituents are aligned with the prosodic domains [10, 19]. For example, Japanese aligns to the left boundary, while English aligns to the right. Empirically, [16] indeed found that English listeners are more consistent in detecting the right boundaries. At the acoustic level, it is also well-known that the acoustic encoding of prosodic boundaries is highly variable across languages.

Overall, there has been much theoretical discussion on the nature of syntax-prosody mapping, but empirical studies on naturalistic speech that directly address this issue are still very few. Most studies were done based on a small number of sentences, and very few production studies based on large-scale continuous speech corpora have been done, because of the lack of syntactically-parsed speech corpora. In this study, we built two such corpora for English and Mandarin, and extracted the acoustic and syntactic features for both languages. By exploring how syntactic phrasing is mapping onto a high-dimensional acoustic space of prosodic phrasing, we will provide valuable empirical insight into the nature of syntax-prosody mapping. Specifically, we test 1) whether there is a reliable correlation between syntactic phrasing and prosodic phrasing; 2) whether there are discrete hierarchies in the mapping between syntax and prosody; 3) whether there is any boundary-alignment preference in the syntax-prosody mapping.

10.21437/Interspeech.2022-10726

To obtain a comprehensive understanding of the acoustic space of prosodic phrasing, a wide range of acoustic cues were examined in this study. We included domain-final cues such as phrase-final lengthening and pause [14, 20, 21, 22], pitch scaling cues [23, 13, 24, 25] as well as voice quality cues [21, 26, 27, 28]. We also included domain-initial relevant cues, such as pitch reset and other domain-initial strengthening effects on duration and voice quality [29, 30].

2. Methods

The lack of syntactically parsed speech corpora has made it difficult for researchers to conduct a large-scale cross-linguistic analysis of the prosody-syntax interface. In this study, we approached this problem by building speech corpora of text in English and Mandarin that have already been fully syntactically parsed according to the Penn Treebank [31] syntactic annotation guidelines. The speech corpora for this study are fluent read speech, which allows us to investigate sentences with relatively complex syntactic structure, and at the same time rule out the factor of disfluency which is a salient feature of conversational speech.

2.1. The English Corpus

For English, because Penn Treebank-style [31] syntactic annotations were available from the 2nd edition of The Penn Parsed Corpus of Modern British English (PPCMBE2) [32], a chapter of Jane Austen's Emma (volume II, chapter 10) has been chosen for analysis. Fluent readings of the book by multiple readers were available from LibriVox [33], a free public domain of audiobooks. The specific chapter has been read by 8 female and 1 male (7 female speakers read the entire chapter, while 1 female and 1 male divided up parts to read.) native English speakers, with a recording sampling frequency of 44.1kHz and sample depth of 16-bit, except for one speaker (Speaker S8C2), whose recording had a sampling rate of 22kHz. A total of 1680 sentences (210 sentences each) were analyzed in this study.

2.2. The Chinese Corpus

A spoken corpus of the Chinese Tree Bank, consisting of segmented, annotated, and parsed news article texts, was chosen as the Mandarin corpus in this study. Similar to the English corpus, Penn Treebank-style [31] syntactic annotations were available from the Chinese Treebank 9.0 [34]. Because not all speakers read the same number of passages and sentences, a subset consisting of 9 passages (107 sentences long) read by all 15 speakers (9 female) were chosen for acoustic analysis. Matching the size of the English corpus, a total of 1448 sentences were analyzed for Mandarin. Speakers in the corpus were all native Mandarin speakers who achieved Class 2 Level 1 or better on the Putonghua Shuiping Ceshi (the national standard Mandarin proficiency test). The recordings were made in a sound-treated booth at Shanghai JiaoTong University, using the software SpeechRecorder. The sampling frequency of the recordings was 44.1kHz, with a sample depth of 16-bit.

2.3. Syntactic constituency and boundary strength

Under the Penn Treebank guidelines, the constituency parsings of the sentences are represented by left and right brackets that group words that belong to the same constituents and syntactic phrases in the sentences. An example of an English sentence's Penn Treebank-style parsings, as well as a tree representation of the sentence is shown in Figure 1.

```
(IP-MAT-SPE (NP-SBJ-1 (PRO It))
            (MD would)
            (BE be)
            (ADJP (ADJ dreadful))
            (IP-INF-SPE-1 (TO to)
                   (BE be)
                   (VAG standing)
                   (ADJP-LOC (ADVR so) (ADJ close))))
```

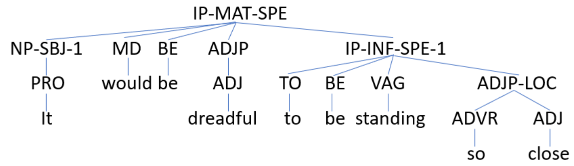

Figure 1: *An example of the syntactic constituency parsings (top) and a tree representation (bottom) of a sentence from Jane Austen's Emma. (ID AUSTEN-1815-2,162.288)*

As can be seen in this example, since each word is a constituent of its own, every word has at least one left and right bracket. Any additional brackets for the words come from additional branching nodes. Words located at the edges of larger syntactic constituents have a greater number of brackets. We therefore counted the number of left and right brackets between each pair of consecutive words to use as proxies for the depth of the syntactic structure and the strength of the syntactic edges.

2.4. Acoustic features

Praat was used to extract acoustic measures. The HMM-based Mandarin forced-aligner [35] was used to align the Mandarin speech, and the Penn Phonetics Lab Forced Aligner [36] was used to force align English speech. A wide range of acoustic measures that are related to prosodic phrasing was extracted: (i) the pause duration at the boundary min-max normalized for each passage; the average syllable duration of the (ii) pre-boundary and (iii) post-boundary words, standardized for each speaker and passage; (iv-v) the post-boundary word's maximum F0 (fundamental frequency)/SPL (sound pressure level) minus the pre-boundary word's minimum F0/SPL, standardized for each speaker and passage to capture effects of phrase initial strengthening (e.g. pitch reset); and the difference between the mean spectral tilt and (vi) Cepstral Peak Prominence-Smoothed (CPPS) of the pre- and post- boundary word, standardized for each speaker and passage. Spectral tilt was measured by the (vii) alpha ratio—the level difference between the 1kHz-5kHz region and that of the 50 Hz-1kHz region, and (viii) L1-L0—the level difference between the first formant region (defined as between 300 Hz and 800 Hz) and the fundamental frequency region (defined as between 0 Hz and 300 Hz). Methods for extracting (vi-viii) and their effectiveness as measures of voice quality are explored in [37].

3. Results

The high-dimensional acoustic space for prosodic phrasing is modeled with principal component analysis (PCA), with the number of left or right brackets introduced in the plots as qualitative variables, to investigate how the variance explained by the first few principal components (PCs) aligns with the strength of the syntactic boundaries. Since Mandarin speech data and English speech data in this study are in different styles (news vs. novel), separate PCA models were fitted for each language.

3.1. English

The first three PCs account for 22.1%, 18.7% and 12.5% of the variance. Due to space limitations, here we focus on the first two PCs. Figure 2 plots the first two principal components for English phrasing.[1] To illustrate the effect of syntactic brackets, the acoustic space in Figure 2 is color-coded twice, by left and right bracket counts. Variations in the data points are represented using ellipses drawing a 95% confidence interval.

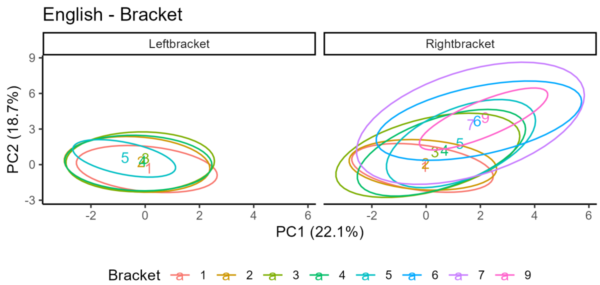

Figure 2: *English PCA model by Left and Right bracket counts. Labels represent the number of brackets. Figures with PC3 are available online in the link in footnote 1.*

As shown in Figure 2, although the acoustic space is generally continuous and gradient, the effects of the number of syntactic brackets can be clearly observed. For right brackets, there is a clear trend that the bracket counts line up in a linear order in the space along both PC1 and PC2. As summarized in Table 1, F0, intensity, voice quality cues in PC1 and durational cues (pause and final lengthening) in PC2 all significantly contribute to the phrasing. Left brackets mostly map onto PC1, although syntactic boundary has a rather weak effect. There is a cluster for bracket counts 2-4, with 1 and 5 brackets on either end of the cluster. Both results suggest that boundary edge strength could be modelled in the acoustic space, along both PC1 and PC2.

Since pause is a highly salient cue in boundary perception [38] and also a highly relevant cue for PC2, we removed tokens with pauses to evaluate whether the lining up of syntactic boundaries is pause driven. The PCA results with pause tokens excluded are illustrated in Figure 3. PC1 (25.0%) and PC2 (17.5%) account for 42.5% of the variance in the data. Results are similar to the model with pauses. Left brackets show a cluster for brackets 2-4, with 1 and 5 on either ends differing in both PCs. Results are less linear for right brackets; however, clustering among lower bracket counts and higher bracket counts is still apparent. For the model with pauses excluded, as summarized in Table 1, PC1 is still mostly correlated to voice quality cues (Diff.CPPS: r = 0.56; Diff:L1L0: r = 0.45; Diff.Alpha: 0.36), while PC2 is most correlated with phrase initial strengthening related measures (Diff.MaxMinSPL: r = 0.55; Diff.MaxMinF0: r = 0.53).

The effects of syntactic bracketing with and without pause tokens were further examined using linear mixed effects models. We predict the first two PCs with Left brackets and Right brackets as numeric main effects, random intercepts by speakers were included. Results showed significant main effects of left (PC1:β = -1.439e-01, SE = 1.489e-02, p <2e-16; PC2: β = 3.327e-01, SE = 1.279e-02, p <2e-16) and right (PC1: β = 2.347e-01, SE = 1.312e-02, p <2e-16; PC2: β = 4.842e-01, SE = 1.127e-02, p <2e-16) brackets for both PC1 and PC2 in the pause included model. The same significant main effects of left (PC1:β = -3.052e-01, SE = 1.545e-02, p < 2e-16; PC2: β = 1.095e-01, SE = 1.300e-02, p < 2e-16) and right (PC1: β = -8.257e-02, SE = 1.829e-02, p = 6.4e-06; PC2: β = 2.031e-01, SE = 1.539e-02, p < 2e-16) brackets were also found in the pause excluded models. These results suggest that acoustic cues other than pauses are effective in indicating syntactic phrasing.

[1] Supplementary plots for PC1 through PC3 may be found here: https://osf.io/b84kq/

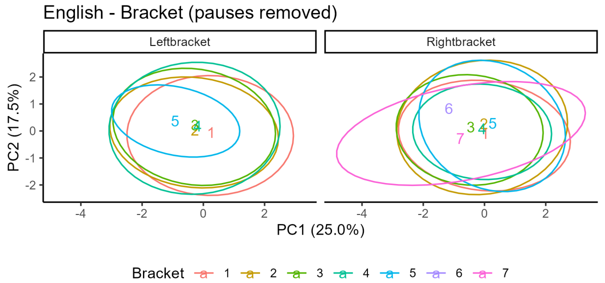

Figure 3: *English PCA model with pauses removed, by Left and Right bracket counts. Labels represent the number of brackets. Plots with PC3 (14.5%) are available online.*

3.2. Mandarin

The same PCA models were run based on the Mandarin data set to explore the mapping between the syntactic brackets and the acoustic space. The first two PCs (PC1:25.0%; PC2:17.1%) of the full PCA model with pauses included are illustrated in Figure 4, color-coded by left and right brackets.

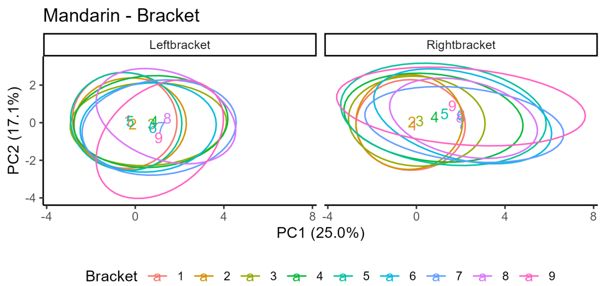

Figure 4: *Mandarin PCA model by Left and Right bracket counts. Labels represent the number of brackets. Plots with PC3 (13.1%) are available online.*

Figure 4 shows that left bracket counts are also clustered by low and high bracket counts, similar to results from English. One complication is that 5 brackets pattern with 1-2 brackets. For right brackets, a clear linear order by bracket count varying mostly along PC1 can be observed. As summarized in Table 1, PC1 is correlated with the maximum intensity and F0 difference across the boundary (i.e., reset effect), pause duration (Norm.Pause: r = 0.50), and pre-boundary syllable duration (Prev.AvgSylDur: r = 0.30). On the other hand, PC2, which is most correlated with spectral slope cues (Diff.Alpha: r = 0.70; Diff.L1L0: r = 0.65), is less relevant to the syntactic bracketing.

To evaluate whether the boundary effects remain without pauses, PCA model (PC1:21.4%; PC2:19.7%) results with pauses excluded are illustrated in Figure 5. Results show that similar clustering effects are found for left brackets, with 1, 2, and 5 brackets forming an initial cluster, and higher bracket counts differ further away along PC1. A general ordering of right brackets in the acoustic space is also observed. For both left and right brackets, 8-9 bracket counts vary further away from the rest of the brackets. With pause tokens excluded, left brackets mostly vary along PC1, which is correlated with phrase initial strengthening related measures (Diff.MaxMinF0: r = 0.63; Diff.MaxMinSPL: r = 0.63); right brackets mostly vary along PC2, which is correlated with spectral slope cues (Diff.L1L0: r = 0.67; Diff.Alpha: r = 0.66).

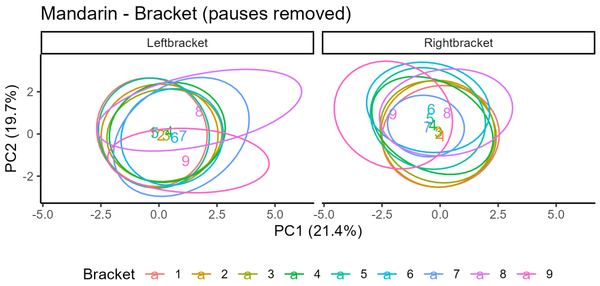

Figure 5: *Mandarin PCA model with pauses removed, by Left and Right bracket counts. Labels represent the number of brackets. Plots with PC3 (15.5%) are available online.*

Linear mixed effects models with the same specifications as the English models were fitted to test whether bracket count can predict each PC. For the pause included model, significant main effects of both left (PC1: β = 1.615e-01, SE = 5.670e-03, p <2e-16, PC2: β = -3.130e-02, SE = 4.872e-03, p <0.001) and right (PC1: β = 2.912e-01, SE = 6.543e-03, p <2e-16, PC2: β = 1.153e-01, SE = 5.622e-0, p <2e-16) brackets for both PCs were found. For the models without pauses, significant main effects of left (β = 9.692e-02, SE = 5.732e-03, p <2e-16) and right (β = -1.071e-01, SE = 8.030e-03, p <2e-16) brackets were found for PC1, but only a significant main effect of right bracket (β = 1.521e-01, SE = 7.727e-03, p <2e-16) but not left (β = -8.680e-03, SE = 5.515e-03, p = 0.116) was found for PC2.

4. Discussion and Conclusion

This study provides empirical understandings of the mapping between the strength of syntactic boundary and prosodic phrasing by examining syntactically-parsed speech corpora of Mandarin and English. First of all, the acoustic space of the prosodic phrasing is rather continuous and gradient in general, and high dimensional cues (pause, final-lengthening, pitch reset, intensity, voice quality) collectively contribute to the phrasing. Secondly, for both English and Mandarin, there is a clear correlation between syntactic bracketing and prosodic cues – larger syntactic boundaries are produced with stronger prosodic boundary cues. And for both English and Mandarin, the syntax-prosody mapping is much stronger for the right boundaries than for the left boundaries. This finding is consistent with the notion that Mandarin and English syntactic boundaries are aligned with the right edges of the prosodic domains, and also consistent with our previous perception results that English and Mandarin listeners are more consistent in perceiving the right boundaries [38]. Moreover, left and right edges of the syntactic boundaries seem to have different effects on the prosodic phrasing. For

Table 1: *Summary of top 5 loadings for PC1-3 in 4 PCA models.*

	English	Mandarin
	Pauses Included	
PC1	Diff.CPPS: 0.47 Diff.L1L0: 0.46 Diff.MaxMinSPL: 0.42 Diff.MaxMinF0: 0.38 Diff.Alpha: 0.31	Diff.MaxMinSPL: 0.55 Diff.MaxMinF0: 0.51 Norm.Pause: 0.50 Prev.AvgSylDur: 0. 30 Diff.CPPS: 0.25
PC2	Prev.AvgSylDur: 0.56 Norm.Pause: 0.49 Diff.CPPS: -0.34 Diff.Alpha: -0.32 Diff.MaxMinF0: 0.31	Diff.Alpha: 0.70 Diff.L1L0: 0.65 Next.AvgSylDur: -0.25 Diff.CPPS: 0.16 Diff.MaxMinSPL: -0.07
PC3	Diff.Alpha: 0.68 Diff.MaxMinSPL: -0.44 Prev.AvgSylDur: 0.30 Diff.L1L0: 0.29 Next.AvgSylDur: -0.25	Diff.CPPS: -0.73 Prev.AvgSylDur: 0.61 Diff.L1L0: 0.19 Diff.MaxMinF0: -0.16 Next.AvgSylDur: -0.10
	Pauses Excluded	
PC1	Diff.CPPS: 0.56 Diff.L1L0: 0.45 Diff.Alpha: 0.36 Diff.MaxMinSPL: 0.32 Next.AvgSylDur: 0.32	Diff.MaxMinF0: 0.63 Diff.MaxMinSPL: 0.63 Next.AvgSylDur: 0.26 Diff.CPPS: 0.24 Diff.Alpha: -0.20
PC2	Diff.MaxMinSPL: 0.55 Diff.MaxMinF0: 0.53 Diff.Alpha: -0.48 Prev.AvgSylDur: 0.32 Diff.L1L0: 0.17	Diff.L1L0: 0.67 Diff.Alpha: 0.66 Next.AvgSylDur: -0.25 Diff.CPPS: 0.20 Diff.MaxMinSPL: 0.08
PC3	Diff.L1L0: 0.55 Diff.Alpha: 0.44 Diff.AvgSylDur: -0.43 Diff.MaxMinSPL: -0.34 Diff.CPPS: -0.28	Prev.AvgSylDur: 0.68 Diff.CPPS: -0.67 Diff.L1L0: 0.23 Diff.MaxMinSPL: 0.10 Diff.MaxMinF0: -0.09

both languages, while left brackets show clustering of large vs. small boundaries, right brackets exhibit a more gradient scaling effect. In addition, since pause was found to be a determinant cue in boundary perception, we examined the phonetic spaces with and without pause tokens. The results suggest that pause is an important cue for both left boundary and right boundary, as the phonetic differences among syntactic boundaries are much smaller when pause tokens were excluded; but syntactic boundary effects were still maintained. Furthermore, acoustic cues are not specified for different syntactic levels. For example, pause is rather gradient, and contributes to the phrasing of all levels of syntactic boundaries. Cue trading between pause and other acoustic cues was not observed. Longer pauses are correlated with longer final lengthening and greater pitch reset. All acoustic cues scale up or down collectively for relative larger or smaller phrasing boundaries. In sum, there is a direct mapping between syntactic and prosodic phrasing, but the mapping appears to be more on a gradient scale. Finally, although English and Mandarin largely share similar mechanisms, there is also some language variation. For example, pause is more strongly correlated with syntactic boundaries for Mandarin than for English, suggesting that the stronger role of pause in boundary perception for Mandarin [38] is rooted in production.

5. References

[1] C. M. Beach, "The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations," *Journal of Memory and Language*, vol. 30, no. 6, pp. 644–663, Dec. 1991.

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

[3] 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.

[4] 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.

[5] 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.

[6] A. Christophe, S. Millotte, S. Bernal, and J. Lidz, "Bootstrapping lexical and syntactic acquisition," *Language and speech*, vol. 51, no. 1-2, pp. 61–75, 2008.

[7] T. Tran, S. Toshniwal, M. Bansal, K. Gimpel, K. Livescu, and M. Ostendorf, "Parsing speech: a neural approach to integrating lexical and acoustic-prosodic information," *arXiv preprint arXiv:1704.07287*, 2017.

[8] M. Nespor and I. Vogel, *Prosodic Phonology: With a New Foreword*. De Gruyter Mouton, 2012.

[9] E. Selkirk, "Syntax and phonology: The relation between sound and structure," 1984.

[10] ——, "On derived domains in sentence phonology," *Phonology*, vol. 3, pp. 371–405, 1986.

[11] B. Hayes, "Precompiled phrasal phonology," *The phonology-syntax connection*, vol. 85, p. 108, 1990.

[12] H. Truckenbrodt, "On the relation between syntactic phrases and phonological phrases," *Linguistic inquiry*, vol. 30, no. 2, pp. 219–255, 1999.

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

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

[15] S.-h. Peng, M. K. Chan, C.-y. Tseng, T. Huang, O. J. Lee, and M. E. Beckman, "Towards a Pan-Mandarin system for prosodic transcription," *Prosodic typology: The phonology of intonation and phrasing*, pp. 230–270, 2005.

[16] 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.

[17] L. Frazier, C. Clifton Jr, and K. Carlson, "Don't break, or do: prosodic boundary preferences," *Lingua*, vol. 114, no. 1, pp. 3–27, 2004.

[18] J. Pynte, "Phrasing effects in comprehending pp constructions," *Journal of psycholinguistic research*, vol. 35, no. 3, pp. 245–265, 2006.

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

[20] 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.

[21] 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.

[22] 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.

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

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

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

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] A. Taylor, M. Marcus, and B. Santorini, "The Penn Treebank: an overview," *Treebanks*, pp. 5–22, 2003.

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

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

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] J. Kuang, M. P. Y. Chan, and N. Rhee, "The effects of syntactic and acoustic cues on the perception of prosodic boundaries," *Speech Prosody*, 2022.