Laboratory Phonology
Journal of the Association for Laboratory Phonology

Lai, W & Li, A. 2022. Integrating phonological and phonetic aspects of Mandarin Tone 3 sandhi in auditory sentence disambiguation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 13(1): 8, pp. 1–31. DOI: https://doi.org/10.16995/labphon.6416

Open Library of Humanities

Integrating phonological and phonetic aspects of Mandarin Tone 3 sandhi in auditory sentence disambiguation

Wei Lai, Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA, wei.lai@vanderbilt.edu

Aini Li, Department of Linguistics, University of Pennsylvania, Philadelphia, PA, USA, liaini@sas.upenn.edu

This study investigates whether Mandarin listeners integrate a prosody-covarying phonological variable, the Chinese Tone 3 sandhi (T3S), into auditory sentence disambiguation. The T3S process changes the first of two consecutive low tones (T3) into a rising tone. It applies obligatorily within a foot and optionally across feet. When T3S is optional, it is more likely to apply to Tone 3 syllables across smaller prosodic boundaries than larger ones; the smaller the boundary, the sharper the T3S pitch rise. Participants listened to twenty-seven structurally ambiguous sentences and identified from two written interpretations the one consistent with what they heard. Each sentence contains two consecutive Tone 3 syllables, and posing different prosodic boundaries between the Tone 3 syllables would result in different interpretations. The first Tone 3 syllable was manipulated into three tone shapes (sharp-rising, shallow-rising, low) and two duration types (long, short). The results show higher major-juncture interpretation rates when the first Tone 3 is long than short, when T3S does not apply than when it applies, and when T3S has a shallower than sharper pitch slope. The tone effect further interacts with the foot formation of Tone 3 syllables in each sentence. We propose that listeners have sophisticated knowledge of prosodic variables and use it efficiently in linguistically meaningful contexts.

1. Introduction

Prosody serves important functions in auditory speech processing at various levels by modulating the tonal and rhythmic aspects of speech (for a review, see Cutler, Dahan, & Van Donselaar, 1997; Pratt, 2017). One of these functions is to facilitate syntactic computation, namely, to help listeners figure out the domain of syntactic constituents and their attachment to one another in the hierarchy. Although syntax and prosody serve different purposes, they are closely related and exhibit substantial overlap in various languages (e.g., Chow, 2018; Domínguez, 2004; Elfner, 2012; Surányi, Ishihara, & Schubö, 2012; Vaissière, 1983). As the speech stream unfolds, listeners can immediately use prosodic cues to segment linguistic units and infer how closely the current unit is associated with its preceding one. The influence of prosody on syntactic processing is evidenced by its role in structural disambiguation (Baek, 2019; Buxó-Lugo & Watson, 2016; Kraljic & Brennan, 2005; Pynte, 1996; Speer, Kjelgaard, & Dobroth, 1996, to name a few). Empirical studies have found that prosody aids the resolution of both global ambiguity arising at the sentence level (e.g., Schafer, Carlson, Clifton Jr, & Frazier, 2000) and local ambiguity arising from the final articulated part of a sentence in online processing (e.g., Beach, 1991). When multiple interpretations are available for a sentence, listeners tend to prefer the analysis that is consistent with the perceived prosodic organization (e.g., Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Schafer, Speer, Warren, & White, 2000).

The disambiguating function of prosody can be realized by either cueing a boundary (Pynte, 1996; Speer et al., 1996) or assigning a focal accent (Kuang, 2010; Schafer, Carlson, et al., 2000). This paper focuses on the former, the effect of prosodic boundary. A prosodic boundary is usually signaled by the modulation of suprasegmental and segmental cues at the edges of a prosodic unit. The beginning of a prosodic unit often has greater articulatory strength in the vocal folds and the vocal tract, leading to domain-initial glottalization (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996) and segment strengthening (Cho, 2004; Fougeron & Keating 1997). The end of a prosodic unit may co-occur with distinctive pitch and tone movements (Beckman & Pierrehumbert, 1986; Hart, Collier, & Cohen, 2006; Lehiste, 1973), intensity reduction (Kim, Yoon, Cole, & Hasegawa-Johnson, 2006; Shen, 1993), creak (Kuang, 2017; H. Zhang, 2016), pauses (Macdonald, 1976; O'Malley, Kloker, & Dara-Abrams, 1973), and preboundary lengthening (Lehiste, 1973; Scott, 1982; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). In contrast to the lengthening and strengthening of speech at the edges of prosodic domains, domain-medial speech materials are usually susceptible to shortening (C. Lai, Sui, & Yuan, 2010; Turk & Shattuck-Hufnagel, 2000) and reduction (W. Lai & Kuang, 2016; Yuan & Liberman, 2015).

Preboundary lengthening refers to a temporal stretch of segmental materials before prosodic boundaries. Acoustic measures indicate the importance of relative duration measures in addition

to absolute ones as a description of preboundary lengthening (e.g., Price et al. 1991). In other words, this phenomenon is signaled by the longer syllable duration of domain-final words than of their domain-medial ones. Preboundary lengthening has been observed at both word (Nakatani, O'Connor, & Aston, 1981; Oller, 1973) and phrase (e.g., Klatt, 1975) boundaries in English and many other languages (Cambier-Langeveld, 1997; Chow, 2004, 2008; Hofhuis, Gussenhoven, & Rietveld, 1995; C. Lai et al., 2010, etc.). Perception studies have shown that domain-final pause is usually a robust cue for parsing (e.g., Baek, 2019; Shen, 1993). In the absence of pauses, pitch movement (Beach, 1991; Chow, 2006; Yang & Wang, 2002) and preboundary lengthening (Price et al., 1991; Scott, 1982; Streeter, 1978; Wightman et al., 1992) also substantially affect listeners' parsing decisions. Some studies argue that preboundary lengthening is more powerful than pitch movement in cueing boundary strength in production (Wang, Xu, & Ding, 2018) and perception (Streeter, 1978), because pitch movement serves other functional purposes of topic encoding and signaling new information. Intensity is reported to be less influential on parsing than pitch movement and lengthening (Streeter, 1978).

Studies on the role of prosodic variables in auditory sentence processing mostly focus on phonetic features rather than phonological ones. Apart from the features reviewed above, prosodic structure is also known to impose categorical constraints on phonological alternations, including segment epenthesis (Itô, 1989), vowel harmony (Van der Hulst, 2016), French liaison (Hsu, 2013), and tone sandhi (Chen, 2000; Shih, 1986; H. M. Zhang, 2016). However, the influence of phonological variables of prosody on sentence parsing has been rarely investigated in the psycholinguistics literature. This paper investigates whether listeners integrate tone sandhi variables in auditory sentence parsing and how they interact with the temporal cue of preboundary lengthening. Tone sandhi refers to a phonological change occurring in tonal languages, in which a lexical tone transforms into a different category as a function of adjacent tones, prosodic position, and morphosyntactic structure (e.g., Chen, 2000). This paper uses the Mandarin Tone 3 sandhi as a testing ground to evaluate whether listeners exploit the phonetic and phonological aspects of tonal variants to aid structural disambiguation in auditory sentence perception.

2. Background

Mandarin has four full lexical tones, which are reported to have typical pitch values of High (Tone 1), Rising (Tone 2), Dipping/Low (Tone 3), and Falling (Tone 4). Following the tradition of a five-level scale notion (Chao, 1968), the pitch targets of the four tones are transcribed as [55], [35], [214]/[21], and [51], where the numbers from [5] to [1] refer to pitch levels from high to low. **Figure 1** shows the mean F_0 contours of the four Mandarin tones read in isolation.

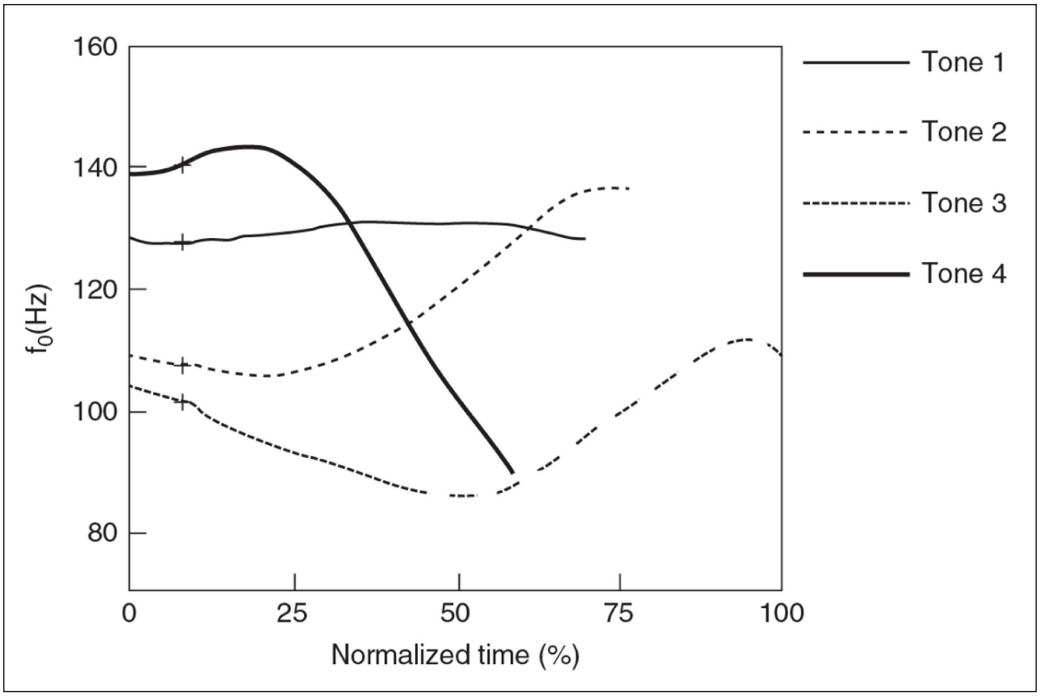

Figure 1: F_0 contours of the four Mandarin tones produced in citation forms adapted from Xu (1997). The dashed rising tail of Tone 3 is a property of the citation form and does not occur word internally.

Although Tone 3 shows a dipping pitch contour in the above figure, this contour usually occurs at phrase-final positions or in isolation. We changed the latter half of Tone 3 into dashed lines to reflect this. A more common variant of Tone 3 is [21], a low pitch contour without the rising tail. The [21] variant of Tone 3 is obligatory at word-internal positions (J. Zhang & Lai, 2010). It is more frequently used than the dipping form at non-final or non-prepausal positions (Xu & Wang, 2001), and it is in free variation with [214] at boundary positions. Therefore, Tone 3 is considered a 'pure low tone' because the meaningful tonal target of Tone 3 in speech perception and production is 'low' rather than 'dipping' (e.g., Xu & Wang, 2001; Zhu, Zhang, & Yi, 2012).

The Tone 3 sandhi (T3S) refers to the process where Tone 3 becomes rising [35] when it precedes another Tone 3. The sandhi variant is similar to Tone 2 in perception, and therefore the sandhi rule is sometimes formulated as Tone 3 → Tone 2 / __ Tone 3. However, acoustic measures show differences in pitch shape between T3S and Tone 2, indicating that this is not a complete neutralization (e.g., Yuan & Chen, 2014). In addition, the specific F_0 realization of the sandhi variant covaries with quite a few factors, including wordhood (J. Zhang & Lai, 2010), word frequency (Yuan & Chen, 2014), and prosodic position (Kuang & Wang, 2006). These factors are to be reviewed later in this section.

2.1 Tone 3 sandhi domain

A substantial body of work has investigated the nature of the domain where T3S is derived. Earlier grammatical analyses of T3S indicate a degree of syntactic dependency (e.g., Cheng, 1973). A well-established finding is that the derivation of T3S with trisyllabic structures is dependent on syntactic branching. This property is exemplified in (1), as adapted from Duanmu (2005). In this and the following examples, digits represent the underlying tonal category, and "s" stands for the sandhi variant. The output tonal sequence of (1) differs depending on the syntactic branching of the structure: A left branching sequence (1a) has only one grammatical output ("s s 3"), whereas a right-branching sequence (1b) has two possible output tonal sequences ("s s 3" and "3 s 3").

(1) a.

	[[mai	hao]	jiu]
	"finished buying wine"		
	buy	done	wine
input	((3	3)	3)
output	((s	s)	3)
	*((3	s)	3)

b.

	[mai	[hao	jiu]
	"buy good wine"		
	buy	good	wine
input	(3	(3	3))
output1	(3	(s	3))
output2	(s	(s	3))

Although the syntactic constraint on T3S derivation is robust with trisyllabic structures, examinations of longer structures suggest that the sandhi domain is not determined only by syntax and does not necessarily reflect syntactic junctures. (2) demonstrates such an example (Chen, 2000). The four-syllable utterance in (2) is broken into two binary units, and the successive Tone 3 syllables "nao" and "jian" are separately organized in the two units. However, the boundary does not block the application of the tone sandhi rule.

(2)

	[tou	nao]	[jian	dan]	"simple-minded"
	head		simple		
input	(2	3)	(3	1)	
output1	(2	3)	(3	1)	
output2	(2	s)	(3	1)	

Phonological studies have also looked into longer strings of Tone 3 syllable sequences and have asked how they are grouped into sandhi domains. Feng (1998) examines Tone 3 syllable sequences in non-branching structures of number strings and finds that they fall apart into sandhi domains that are two or three syllables long, as shown in (3). He defines these domains as the 'natural foot' in Mandarin Chinese.

(3)

	wu	wu	wu	wu	wu	"five five five five five"
input	3	3	3	3	3	
output	(s	3)	(s	s	3)	

On a related note, Kuang (2005) shows that within left-branching structures of five Tone 3 syllables in a row, T3S does not apply across the board as predicted by the branching. Instead, again, they are grouped into disyllabic and trisyllabic sandhi domains. In the example of Feng (1998) and Kuang (2005), the T3S domain is no longer a reflection of syntactic structure but rather a low-level prosodic grouping process similar to syllable footing. The small prosodic units composed of syllables are, in turn, grouped into larger, higher-level units. As a prosodic unit grows larger, it is increasingly likely to be followed by a pause—which is another prosodic factor that interacts with T3S. Cheng (1973) points out that the existence of an intentional pause is likely to block the application of T3S.

To reflect the coexisting syntactic and prosodic dependencies of T3S, more sophisticated views are put forward to define the T3S domain as a prosodic organization derived under syntactic constraints (e.g., Chen, 2000; Shih, 1986). The leading view regards T3S application as a diagnostic process of the 'foot' (Shih, 1986, p. 102), namely, the minimal rhythmic unit in Mandarin defined by the binary-foot rule and syntactic branching (but cf. Hung, 1987; Z.-s. Zhang, 1988). The basic principles of foot formation are proposed by Shih (1986, p. 110) and are shown in (4).

(4) Mandarin Foot Formation Rule:
   I. Foot Construction:
      a Immediate Constituency: Link immediate constituents into disyllabic feet;
      b Duple Meter: Scanning from left to right, string together unpaired syllables into binary feet, unless they branch to the opposite direction;
   II. Super-foot Construction: Join any leftover monosyllable to a neighboring binary foot according to the direction of syntactic branching.

According to this theory, T3S applies cyclically based on a hierarchically organized prosodic structure. The units of this structure include foot (Ft, composed of a disyllabic word or two adjacent monosyllabic words), superfoot (Ft', composed of disyllabic feet and adjacent, as yet unadjoined syllables), phrase (Ph, composed of feet and super feet), and utterance (Utt, followed by a pause). The foot is the obligatory domain for T3S. Failing to apply the rule within a foot yields unacceptable readings in any style and at any speech rate. The utterance is the upper boundary of the sandhi rule, across which T3S application is prohibited. At the intermediate levels (superfoot, phrase), T3S applies optionally across feet in an utterance.

Shih's approach is sufficient to account for most cases of T3S-related grammaticality judgments. It successfully resolves the phenomenon that T3S is never blocked within disyllabic structures, regardless of the type of syntactic juncture in between. In addition, it addresses the dependency of T3S on syntactic branching in trisyllabic structures by formulating cyclic application rules for T3S.

2.2 Covariation between Tone 3 sandhi and prosodic structure

The T3S rule proposed in Shih (1986) involves a degree of covariation between the T3S domain and prosodic 'closeness' (Speer, Shih, & Slowiaczek, 1989), i.e., prosodic boundary strength. That is, the upper bound and the lower bound of the T3S domain correspond to specific types of prosodic boundaries. The Tone 3 syllables are maximally prosodically close when they fall into the same foot where the T3S rule must apply; the rule becomes optional when the syllables are more prosodically distant, with higher-level-prosodic boundaries intervening in between.

The covariation between T3S and boundary strength is also influenced by the probabilistic distribution of T3S across the intermediate prosodic levels: When T3S is optional, it is less likely to apply if the two adjacent Tone 3 syllables are separated by a larger prosodic boundary. This relationship is evidenced by both production studies (Kuang & Wang 2006) and perception studies (Speer et al., 1989). Kuang (2005) compares the frequency of sandhi application to successive Tone 3 syllables across different levels of prosodic junctures in 150 designed sentences produced by native Beijing speakers. She finds that although T3S is optional both across feet within a prosodic phrase and across different prosodic phrases within an utterance, the sandhi rule is highly likely to apply in the former case (across feet within a phrase) with only a few exceptions, whereas the frequency decreases in the latter case (across phrases within an utterance). Speer et al. (1989) evaluate the likelihood of listeners perceiving a T3S as opposed to Tone 2 for the same recording excerpt (Tone 2 syllable) embedded in sentences of different lengths. They find that listeners are more likely to perceive Tone 2 instead of T3S in longer sentences where the two target syllables are pulled apart into two separate phrases by sentential materials, and T3S is more likely to be perceived when the sentence has fewer syllables. This finding suggests that listeners may have knowledge of the probabilistic distribution of T3S, conditioned by prosodic structure, and they employ it in speech perception.

When T3S applies, sandhi variants at different levels of prosodic boundaries also vary in their features, such as the range of F_0 rising (Yuan & Chen, 2014), F_0 slope, duration, and intensity (Kuang & Wang, 2006). Kuang and Wang (2006) provide a detailed description of the acoustic manifestation of T3S at different prosodic boundaries: for T3S applying within a prosodic phrase where there is no obvious pause and lengthening between the two Tone 3 syllables, sandhied syllables within a foot have sharper rising pitch and shorter duration than those located at a foot boundary. For T3S applying across phrases with lengthening but no pause, the first syllable is much longer than the second one, and the rising pitch is obviously lengthened, leading to a shallow pitch rise. For consecutive Tone 3 syllables across utterances, the first syllable is short, and the intensity reduces quickly; pitch and intensity are reset in the subsequent syllable. T3S rules cannot apply across this kind of boundary.

Figure 2 provides an illustration of the F_0 and duration of consecutive Tone 3 syllables in three kinds of prosodic environments, namely, within foot (left), across feet (middle), and across utterances (right).[1] Tone sandhi applies in the left and the middle facets (rising plus low) but not in the right facet (low plus low). Gradient differences can be observed between foot-medial (left) versus foot-final sandhi realizations (middle): Compared to the foot-final sandhi, the foot-medial one has an earlier turning point, larger F_0 increase, and shorter duration whereas the foot-final one has a later turning point, longer duration, and a smoother rising F_0 contour.

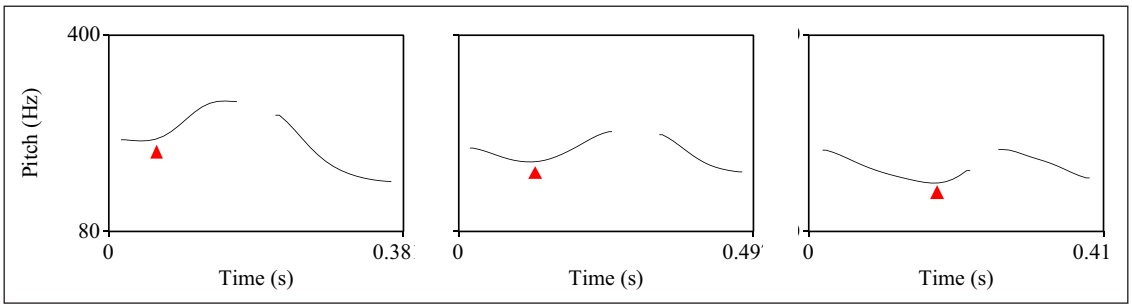

Figure 2: Pitch contours of two consecutive Tone 3 syllables (x-axis: time in seconds; y-axis: F_0 in Hz) within foot (left), cross feet (middle), and cross utterances (right). The red triangle indicates turning point of the first syllable. T3S applies in the left and the middle facets but not in the right facet.

To summarize, Mandarin T3S is influenced by factors at various linguistic levels, including grammatical and prosodic factors. Across prosodic boundaries larger than the foot, the incidence of sandhi is probabilistic rather than deterministic, reflecting prosodic boundary strength. Besides, the acoustic realizations of sandhi are gradient rather than categorical. Building on these findings, in this paper, we further investigate how variants of the T3S are integrated into sentence parsing in speech perception. We expect answers to this question to shed light on the encoding of linguistic variation in listeners' language knowledge and its facilitative role in language processing and comprehension.

[1] The F_0 is extracted from a recording of the first author pronouncing the same Tone 3 sequence, "hao hen," in three stimuli where these two syllables are separated by either a syllable boundary, a foot boundary, or an utterance boundary (defined by a pause). The stimuli, where the same two syllables are embedded in these different prosodic contexts, are adopted from Kuang and Wang (2006, p. 10–11). The three stimuli used for Figure 2 are as follows (T3 syllables and their glosses are in bold):

a. *within foot:* zhe4-me0 gai3-yi4-gai3 jiu4 [**hao3 hen3**-duo1] le0. "These changes make it **much better**."
b. *cross feet*: [pei4-he2 **hao3**][**hen3** nan2]. "To cooperate **well** / (is) **very** hard."
c. *cross utterances*: [da4-jia1 **hao3**], [**hen3** gao1-xing4 ren4-shi0 ni3-men0]. "Everyone **hello** / **very** glad to know you all."

3. The current paper

This paper investigates whether Mandarin Chinese listeners integrate auditory forms of T3S variants to guide their parsing decisions for structurally ambiguous sentences in speech perception. In the light of previous findings that T3S covaries with prosodic structure through both phonological alternation and phonetic variation, we evaluate whether listeners integrate the application of the sandhi rule and the pitch shape of sandhi variants to perceive prosodic boundaries. Furthermore, the pitch shape of a T3S variant is also dependent on the length of its bearing syllable, and the bearing syllable is further susceptible to preboundary lengthening. Therefore, we also evaluate the interaction between T3S cues and duration cues in sentence parsing.

We break down the research goal of this paper into the following specific questions:

1. Would listeners attend to the application versus non-application of T3S and use this information to guide sentence parsing and disambiguation? If so, does this effect vary with different prosodic groupings of the critical Tone 3 syllables in each sentence?
2. When T3S applies and when it changes the first low tone into a rising tone, would listeners attend to the pitch slope of the rising variant and use it to guide sentence parsing and disambiguation?
3. How does the integration of T3S variants interact with the length of the preboundary Tone 3 syllable in perception? When tone and duration are manipulated in parallel, will they affect boundary perception independently or interactively?

3.1 Stimuli

Twenty-seven structurally ambiguous sentences were constructed to serve as critical stimuli. The length of these sentences varies from four to eight syllables. Each sentence contains two consecutive Tone 3 syllables. Different interpretations can be derived depending on whether the two Tone 3 syllables are separated by a strong prosodic juncture or a weak one. The two optional interpretations are therefore named a ‘major-juncture interpretation’ (with a strong intervening boundary) and a ‘minor-juncture interpretation’ (with a weak intervening boundary). The breakdown of the critical sentences is described in more detail later in this section.

Twenty-seven ambiguous filler sentences were added to the experiment to increase variability. They are 4-13 syllables in length and do not contain consecutive Tone 3 syllables. The ambiguity of filler sentences is triggered by either lexical, structural, or pragmatic factors. A full list of the critical sentences and the filler sentences is provided in the Appendix.

3.1.1 Tone 3 sandhi in sentence disambiguation

We provide several examples of our critical sentences to show how variants of Tone 3 sandhi would affect the interpretation of these sentences. (5) adopts an ambiguous structure of *[Verb NP1 DE NP2]*, which is ambiguous between a complement-clause interpretation and a relative-clause interpretation (e.g., Hsieh, Boland, Zhang, & Yan, 2009). The two interpretations are provided along with their corresponding syntactic and prosodic structures. The critical consecutive Tone 3 syllables are placed in the VP-final and the NP1-initial positions. They are bolded in the syntactic analyses and represented with bold-italicized "σ" in the prosodic analyses.

(5) dai4-**bu3** **wo3**-men0 de0 jian4-die2
"arrest" 1-PPL POSS/RC "spy"

(5a) major-juncture interpretation: "Arrest our spy."
syntax: $[_{VP}\ [_{VP}$ dai4-**bu3**$]\ [_{NP}\ [_{PossP}\ [_{NP}$ **wo3**-men0$]$ de0$]\ [_{NP}$ jian4-die2$]]]$
prosody: $(\ (\sigma\ \boldsymbol{\sigma})_{Ft})_{Ph}\ (\ (\boldsymbol{\sigma}\ \sigma\ \sigma)_{Ft'}\ (\sigma\ \sigma)_{Ft}\ )_{Ph}$

(5b) minor-juncture interpretation: "The spy that arrests us."
syntax: $[_{NP}\ [_{CP}\ [_{VP}\ [_{VP}$ dai4-**bu3**$]\ [_{NP}$ **wo3**-men0$]]$ de0$]\ [_{NP}$ jian4-die2$]]$
prosody: $(\ (\sigma\ \boldsymbol{\sigma})_{Ft}\ (\boldsymbol{\sigma}\ \sigma)_{Ft}\ \sigma)_{Ph}\ (\ (\sigma\ \sigma)_{Ft}\ )_{Ph}$

(5) has two interpretations that differ in whether there is a major syntactic and prosodic juncture between VP and NP1. The sentence means "arrest our spy" when the two Tone 3 syllables fall apart into two different phrases, as in (5a); it means "the spy that arrests us" when the Tone 3 syllables are grouped in the same phrase, as in (5b). Since the two Tone 3 syllables are separated by a phrase boundary in (5a) but a foot boundary in (5b), we refer to (5a) as the major-juncture interpretation and (5b) as the minor-juncture interpretation. Since the prosodic juncture across the two Tone 3 syllables is larger in (5a) than in (5b), we expect tone sandhi to be less likely to apply in (5a) than in (5b). If tone sandhi applies in both conditions, we expect the slope of the T3S variant to be shallower in (5a) and sharper in (5b).

Similarly, we predict that sentences where T3S applies would cause fewer major-juncture interpretations than sentences where T3S does not apply. For sentences where T3S applies, we predict that a sandhi variant with a shallower-rising F_0 contour is more likely to cause a major-juncture interpretation than a sandhi variant with a sharper-rising F_0 contour.

3.1.2 Coding the minimal Tone 3 grouping unit

Like (5), each of the 27 critical sentences has a major-juncture interpretation and a minor-juncture interpretation. The specific interpretation that each listener arrives at depends on the presence or absence of a major prosodic boundary between the Tone 3 syllables. However, since

the specific type of the involved prosodic boundary or domain differs across sentences, this has additional influence on disambiguation responses. Consider (6) for such an example. Unlike (5) whose interpretation depends on whether the two Tone 3 syllables are in the same *phrase* or not, (6)'s interpretation depends on whether the two syllables are in the same *foot* or not.

(6) **wo3** **xie3** bu4 hao3
1PSG "write" NEG "good/well"

(6a) major-juncture reading: "I cannot write (it) well."
prosody: ***σ*** (***σ*** σ σ)$_{Ft'}$ / syntax: [$_{NP}$ **wo3**][$_{VP}$ **xie3** bu4 hao3]

(6b) minor-juncture reading: "It is not good that I write (it)."
prosody: (***σ σ***)$_{Ft}$ (σ σ)$_{Ft}$ / syntax: [$_{CP}$ **wo3 xie3**][$_{NegP}$ bu4 hao3]

As elaborated in Section 2.1, the grouping domain of the foot posits strict constraints on T3S application: The sandhi process becomes mandatory in order for the sentence to be grammatical. In our case here, T3S is mandatory in (6b) but optional in (6a). In other words, if T3S applies, the sentence is still ambiguous between the two interpretations; but if it does not apply, then (6a) becomes the only possible interpretation, since (6b), without T3S, is ungrammatical. Therefore, sentences like (6) should be treated differently from sentences like (5), as listeners can exploit the grammaticality properties of T3S in addition to its probabilistic and acoustic aspects for the disambiguation of (6). However, this grammaticality cue is not useful for interpreting (5).

To distinguish the above two situations, for each critical sentence, we coded its minimal Tone 3 grouping unit, namely, the lowest-level prosodic domain that the two Tone 3 syllables share under a minor-juncture reading. Sentences similar to (6) were coded to have a minimal Tone 3 grouping unit of foot (Ft), whereas those similar to (5) were coded to have a minimal grouping unit of phrase (Ph). Apart from the two types of units discussed above, it is also possible that the Tone 3 syllables are grouped into a superfoot (Ft') at the lowerest level. An example is provided in (7).

(7) **deng3** **da3**-qiu2 de0 tong2-xue2
"wait" "play basketball" POSS/RC "classmate"

(7a) major-juncture reading: "Wait for the classmate playing basketball."
syntax: [$_{VP}$ [$_{VP}$ **deng3**] [$_{NP}$ [$_{CP}$ [$_{VP}$ **da3**-qiu2] de0] [$_{NP}$ tong2-xue2]]]
prosody: ***σ*** (((***σ*** σ)$_{Ft}$ σ)$_{Ft'}$ (σ σ)$_{Ft}$)$_{Ph}$

(7b) minor-juncture reading: "The classmate waiting for playing basketball."
syntax: [$_{NP}$ [$_{CP}$ [$_{VP}$ **deng3** [$_{CP}$ **da3**-qiu2]] de0] [$_{NP}$ tong2-xue2]]
prosody: ((***σ*** (***σ*** σ)$_{Ft}$)$_{Ft'}$ σ)$_{Ph}$ ((σ σ)$_{Ft}$)$_{Ph}$

Among the 27 sentences, 'phrase' was coded as the minimal Tone 3 grouping unit for ten sentences, 'superfoot' for 13 sentences, and 'foot' for the remaining four sentences. Note that in our sample, the minimal Tone 3 grouping unit is highly correlated with sentence length and syntactic structure. Sentences labeled with a minimal grouping unit of 'phrase' are seven to eight syllables long; sentences labeled with 'superfoot' are all six syllables long; sentences labeled as 'foot' have a length ranging from four to five syllables. Most of the sentences in the first two categories (i.e., with a minimal Tone 3 grouping unit of 'phrase' and 'foot') have a *[Verb NP1 DE NP2]* structure, whereas none of the sentences in the third category has such a structure. The factors of sentence length, minimal Tone 3 grouping unit, and syntactic structure are included as predictors in our statistical model in Section 4.

3.1.3 Norming study

A norming study was conducted to evaluate Mandarin speakers' preferences for any specific interpretations of the critical sentences in silent reading. Twenty native Mandarin speakers were recruited from the UPenn undergraduate subject pool and from mainland China to participate in a norming experiment through Qualtrics (Qualtrics, 2020) in exchange for credit or payment. They were instructed to read 54 ambiguous sentences presented in Chinese characters and choose from two unambiguous paraphrases the one that was more consistent with their reading. The 54 ambiguous sentences consisted of the 27 critical sentences and the 27 filler sentences, which were randomized for each participant and presented one at a time. The order of the two choices in each trial was also randomized by participant. Note that these participants did not participate in our main auditory sentence perception experiment. Their responses therefore should reflect parsing decisions based on their implicit prosody (e.g., Bishop, 2020) rather than the acoustics of our stimuli.

A mean major-juncture interpretation rate of each critical sentence was calculated using the number of major-juncture interpretations divided by the total number of responses. On average, participants reported a major-juncture interpretation 38.6% of the time and a minor-juncture interpretation 61.4% of the time. We also checked the norming results for sentences with different syntactic structures and Tone 3 grouping units. In the aggregate, sentences with a *[Verb NP1 DE NP2]* structure received a major-juncture interpretation 52.6% of the time, which is higher than sentences with other types of structures (32.8%). Sentences with a minimal Tone 3 grouping unit of foot received a major-juncture interpretation 55.3% of the time, followed by sentences with a minimal unit of phrase (37.5%) and those with a minimal unit of superfoot (33.7%).

The mean major-juncture interpretation rate of each sentence was used as an index of the sentence-level comprehension bias, which was later included in our statistical model as a predictor of the responses in the auditory sentence perception experiment in Section 4.

3.2 Recording and manipulation

The critical sentences and the fillers were produced by a female native Mandarin speaker (the first author) and were recorded in a professional recording booth at the University of Pennsylvania. The 27 critical sentences were each repeated six times with sandhi and six times without sandhi. Among the six repetitions in the sandhi/non-sandhi condition, three of them were read in favor of a major-juncture interpretation and the other three were read in favor of a minor-juncture interpretation. For each critical sentence, one well-articulated sandhied production was chosen to be further manipulated into the experimental stimuli. We chose recordings produced with T3S instead of those with two low tones because sandhi forms of the first Tone 3 syllable were less susceptible to creak than low forms.

We first manipulated the duration of all the segments in the 27 selected sentences to neutralize any potential bias in their timing pattern, in order to not favor one interpretation over the other. To do this, we first calculated the mean duration of each segment in each sentence across the 3 (repetition) × 2 (sandhi) × 2 (semantics) = 12 repetitions of that sentence. We forced-aligned all the 27 (sentence) × 12 (repetitions) = 324 recordings using the Penn Phonetics Lab Forced Aligner (Yuan & Liberman, 2008). The output were 324 Praat TextGrid files, each containing the information of the identified phoneme and word boundaries of a sentence. After manually checking the accuracy of the boundary information, we extracted the duration of all the segments from the TextGrid files and calculated the mean duration of each segment in each sentence. Then, the duration of each segment of the selected sentence was manipulated to match its mean duration across the 12 repetitions using the PSOLA algorithm in Praat.

Next, for the critical Tone 3 syllables in the selected sentences, F_0 and timing were manipulated. F_0 was manipulated using Praat's PitchTier function, through which the F_0 contours of the recordings were displayed, modified, and added back to the recordings through overlap-add speech resynthesis. For each critical sentence, we adjusted the F_0 of the consecutive Tone 3 syllables into three tone patterns: 'sharp-rising low' (SharpR), 'shallow-rising low' (ShallowR), and 'low low' (Low). The tone conditions were defined by the following perceptual criteria as judged by the authors. First, the output sentences sound unambiguous to the authors regarding whether tone sandhi applies or not. Second, the Tone 3 sequences in the SharpR condition sound identical to Tone 2 plus Tone 3 in Mandarin Chinese. Third, the Tone 3 sequences in the ShallowR condition sound distinguishable from their counterparts in the SharpR condition. This was achieved by adjusting the turning point of a SharpR variant to be later in timing and lower in F_0, and by compressing the rise of F_0 after the turning point (as shown in **Figure 3**).

Following previous studies on the perception of prosodic boundary (e.g., Buxó-Lugo & Watson, 2016; X. Zhang, 2012), we further manipulated the duration of the preboundary materials as a

cue to boundary strength. In our case, the preboundary unit of interest is the first Tone 3 syllable, which was manipulated into two timing conditions (Short and Long) for each sentence. The Long condition favors the perception of a boundary between the Tone 3 syllables by inducing a level of preboundary lengthening. In contrast, the Short condition favors the absence of a prosodic boundary in perception. Since the timing pattern derived from the mean syllable lengths already has the first Tone 3 syllable outranking the second one in length, this was directly used to serve as our Long timing condition. Stimuli of the Short condition were generalized by compressing the first Tone 3 syllable to 0.7 of its length in the Long condition using the PSOLA algorithm in Praat. This is a uniform compression across syllable lengths, which vary with one another depending on their phoneme classes.

The filler sentences were read with two prosodic patterns, each in favor of a different interpretation. These sentences were not further manipulated in F_0 and timing. Finally, all of the critical stimuli and the fillers were scaled to 70 dB.

3.3 Acoustic measures

The mean F_0 contours of the Tone 3 sequences in the three tone conditions and the two timing conditions are shown in **Figure 3**. Each contour represents the mean of the manipulated F_0 contours of the 27 sentences in one specific tone and timing condition, and the shaded areas represent their deviations by one standard error. The four dashed vertical lines indicate the beginning and the end of the two syllables. As mentioned above, the first Tone 3 syllable (mean syllable duration: 187 ms; mean pitch duration: 170 ms) is longer than the second one (mean syllable duration: 151 ms; mean pitch duration: 133 ms). The mean duration of the first syllables is 131 ms after temporal compression. The compressed F_0 contours in the Short timing condition are presented in red and are aligned with those in the Long condition by the right edge.

In **Figure 3**, each of the two Tone 3 syllables in the Low Tone condition bears a falling F_0 with no F_0 reset in between. By contrast, in the SharpR and ShallowR Tone conditions, the first syllable bears a falling-rising F_0 contour whereas the second syllable bears a falling F_0 contour.[2] The acoustic differences between the first Tone 3 syllables in the SharpR condition and those in the ShallowR condition were examined with the following features: Ending F_0 (the F_0 height at the end of the syllable), Valley F_0 (the F_0 height of the turning point), ΔF_0 (the amount of rising F_0 after the turning point), and Valley Loc. (the location of the turning point in the syllable, indexed by the distance between the syllable onset and the turning point divided by the entire syllable length). A summary of these features of the first tone in the three tone conditions is

[2] The initial falling part of the F_0 contour of the first syllable is due to the carry-over effect of Chinese tone coarticulation (e.g., Xu, 1997).

provided in **Table 1**. Notably, these acoustic parameters remain identical in the two timing conditions, although the rising F_0 slope is generally sharper in the Short condition than in the Long condition. This is due to the realization of the same amount of F_0 movement within a shorter time window.

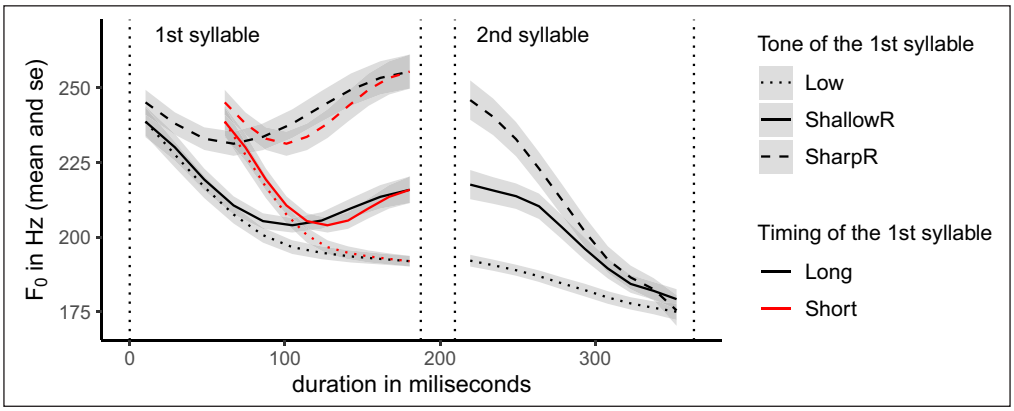

Figure 3: The mean F_0 contours of the two Tone 3 syllables in three tone conditions and two timing conditions.

	Ending F_0 (Hz)		Valley F_0 (Hz)		ΔF_0 (Hz)		Valley Loc.	
	Mean	***SD***	***Mean***	***SD***	***Mean***	***SD***	***Mean***	***SD***
SharpR	256	30	227	21	29	18	0.40	0.22
ShallowR	216	23	199	15	17	15	0.63	0.21
Low	192	9	191	10	0	3	0.89	0.20

Table 1: Features of the F_0 contour of the first Tone 3 syllable in the three tone conditions.

3.4 Participants

Eighty-two participants were recruited to participate in an online experiment of Chinese auditory sentence comprehension through Qualtrics. Thirty-nine of them were recruited from mainland China, and 43 were recruited from the undergraduate subject pool at the University of Pennsylvania. According to the results of a brief demographic survey given before the experiment, we only used data from participants who reported themselves to have grown up in mainland China (excluding Hongkong, Macau, Taiwan), speak Mandarin as their first language, and have normal hearing. Sixty-five participants turned out to satisfy all the above criteria. Two participants were further excluded in order to fulfill the counterbalancing design of the experiment (see Section 3.5). The remaining participants were 38 women and 25 men, aged from 18 to 33 years old ($Mean = 23$, $SD = 3.6$).

3.5 Procedure

Three experimental lists were constructed. Each list was made up of 54 experimental stimuli and 46 filler items. The 54 experimental stimuli were nine low-tone items, nine sharp-rising sandhi items, and nine shallow-rising sandhi items, each occurring twice in different temporal conditions. The pairing of items with tone conditions was counterbalanced across lists, such that each participant heard each experimental stimulus in only one of the three tone conditions, and heard them once in each of the Short and Long timing conditions. Participants were randomly assigned to one of the three lists, with an equal number of participants ($N = 20$) hearing each list. Each participant was asked to provide information about their age, gender, birthplace, language background, and whether they have hearing deficits.

Regarding the experimental task, listeners were told that they would hear sentences that were ambiguous in meaning. For each sentence, they would see two options of interpretations rewritten in unambiguous ways. The order of the two options of interpretations was randomized for each trial. Participants were instructed to identify the interpretation that was consistent with the speech they heard. Each participant went through three practice trials of filler sentences before they proceeded to the experiment. All of the 100 trials (54 critical stimuli and 46 fillers) were presented in a single block, with the order of trials and the two choices in each trial randomized by participant.

4. Results

Analyses were conducted using the *R* Statistical environment (R Core Team, 2014). Mixed effects logistic regression was run using the *lme4* library (Bates, Mächler, Bolker, & Walker, 2015), and plots were created using *ggplot* (Wickham, 2016). Post-hoc pairwise differences were assessed using estimated marginal means (Searle, Speed, & Milliken, 1980), which were extracted from the fitted model with the *emmeans* package version 1.4.3 (Lenth, Singmann, Love, Buerkner, & Herve, 2019) in R. The data and analysis scripts of this paper are available at https://osf.io/ubgpw/.

We first fit a mixed-effects logistic regression model to predict listeners' responses of sentence interpretation, with the following five predictors as fixed effects[3]:

[3] In earlier versions of this paper, we also included SentenceLength as a model predictor. Due to the high collinearity between SentenceLength and MinimalUnit, we created a parameter of Length.Resid by evaluating a linear regression model to fit SentenceLength with MinimalUnit and subtracting the predicted values from SentenceLength. We thank a guest editor who brought to us critiques about minimizing collinearity via residualization (e.g., Wurm & Fisicaro, 2014) and dropped sentence length entirely from our current modeling. Researchers interested in the potential effect of sentence length could look into this issue in future research.

- Tone: ternary predictor indicating the tone shape of the first of two consective Tone 3 syllables. Levels: Sharp rising (SharpR), Shallow rising (ShallowR), Low.
- Timing: binary predictor indicating whether the first Tone 3 syllable is shortened by temporal compression. Levels: Long, Short.
- MinimalUnit: ternary predictor indicating the minimal prosodic domain that the two Tone 3 syllables are in. Levels: Foot (Ft), Superfoot (Ft'), Phrase (Ph).
- NormingBias: continuous predictor indicating the mean major-juncture interpretation rate of each sentence in silent reading.
- Syntax: binary predictor indicating whether the stimulus sentence has a [Verb NP1 DE NP2] structure or not. Levels: [Verb NP1 DE NP2], Other.

All the continuous variables were centered and scaled using the *scale()* function in R. All the categorical variables were sum-coded, allowing us to compare the responses under specific experimental conditions with the grand mean.[4] Interaction items included all the possible two-way interactions involving Tone or Timing (Tone × Timing, Tone × MinimalUnit, Tone × NormingBias, Tone × Syntax, Timing × MinimalUnit, Timing × NormingBias, Timing × Syntax) and all the possible three-way interactions involving both Tone and Timing (Tone × Timing × MinimalUnit, Tone × Timing × NormingBias, Tone × Timing × Syntax).

Model comparisons were performed using log-likelihood ratio tests to diagnose non-significant factors and find the model with the optimal fit. Model selection began from a maximal model and proceeded in a backward stepwise manner by removing insignificant factors one at a time and comparing the reduced model with the superset model at each step. A Chi-square test was used to calculate whether the superset model fits significantly better or worse than the reduced model. If the model comparison resulted in a value of $p < 0.05$, the model that provided the better fit was chosen. Otherwise, if the two models were not significantly different, the simpler model was chosen.

The model selection process started with evaluations of the benefits of random factors. At each step, we proceeded by removing the random factor that captured the smallest amount of variance according to the model output. The full model included Subject and Sentence as two random intercepts and Tone by Subject and Timing by Subject as two random slopes. Through the model selection process, we successively removed the random slopes of Timing by Subject (χ^2 (4) = 0.08, p = 1) and Tone by Subject (χ^2 (5) = 4.94, p = 0.42), because they did not

[4] We did not adopt treatment coding for the model reported in Table 2 such that the results can represent the estimates based on the whole set of our data instead of a subset defined by the 'baseline' conditions. In places where a direct comparison between different levels of a variable is desirable, we treatment coded the variable and ran this model again to resolve comparisons of interest.

significantly improve the model fit. The two random intercepts were retained (Subject: χ^2 (1) = 8.13, p = 0.004; Sentence: χ^2 (1) = 331.22, p < 0.001) and were used consistently in all the following models to be compared.

Turning to the main predictors, we started with higher-order variables (i.e., interaction terms) and proceeded to lower-order ones. For variables at the same level, we examined insignificant variables before marginally significant and significant ones. For variables matched in both criteria, we conducted a F-test using *anova()* to obtain the F-value of each variable. We examined variables with smaller F-values before those with large F-values. Each time when a new optimal model was obtained, we repeated the above process to decide on the variable to be removed from that specific optimal model. No decision was made based on the orders decided at earlier steps of the comparison process.

For the three-way interaction items, the model selection results showed that none of them significantly improved the model fit. They were removed from the model in the order of Tone × Timing × NormingBias (χ^2 (2) = 1.06, p = 0.59), Tone × Timing × Syntax (χ^2 (2) = 0.72, p = 0.70), and Tone × Timing × MinimalUnit (χ^2 (4) = 2.10, p = 0.72). Similarly, four of the two-way interactions were removed for not significantly improving the model fit. They were removed in the order of Timing × NormingBias (χ^2 (1) = 0.42, p = 0.52), Tone × Timing (χ^2 (2) = 1.78, p = 0.41), Tone × Syntax (χ^2 (2) = 0.49, p = 0.79), and Tone × NormingBias (χ^2 (2) = 4.32, p = 0.12).

The remaining predictors were retained in the final model because the inclusion of either themselves or their higher-order interaction items led to significant improvements of the model fit. The final model[5] took Tone (χ^2 (6) = 46.14, p < 0.001), Timing (χ^2 (4) = 18.86, p < 0.001), NormingBias (χ^2 (2) = 5.28, p = 0.02), MinimalUnit (χ^2 (8) = 24.94, p < 0.01), and Syntax[6] (χ^2 (2) = 4.30, p = 0.12) as fixed effects, Tone × MinimalUnit (χ^2 (4) = 16.92, p = 0.002), Timing × MinimalUnit (χ^2 (2) = 6.76, p = 0.03), and Timing × Syntax (χ^2 (1) = 4.09, p = 0.04) as interaction items, and Speaker and Sentence as random intercepts, to predict listeners' interpretation responses (minor-juncture = 0, major-juncture = 1).

The estimates of the fixed effects of this model are summarized in **Table 2**. A positive estimate in the model tables indicates that the major-juncture interpretation is more likely, and a negative estimate indicates such an interpretation is less likely.

[5] We thank the editors for pointing us to a *buildmer()* R function (Voeten, 2020), which takes the full model as an argument and performs model selection automatically. We tried *buildmer()* with our full model, and it arrived at the same optimal model as the one we reported here.

[6] Syntax was retained in the final model due to its involvement in significant higher-order predictors.

Fixed Effects	Estimate	*SE*	z value	Pr (>|z|)
(Intercept)	–0.04	0.20	–0.18	0.86
Tone (SharpR)	–0.33	0.06	–5.15	<0.001
Tone (Low)	0.38	0.07	5.58	<0.001
MinimalUnit (Ft)	0.43	0.43	1.01	0.31
MinimalUnit (Ph)	–0.29	0.28	–1.04	0.30
NormingBias	0.44	0.18	2.43	0.02
Timing (Short)	–0.14	0.05	2.96	0.003
Syntax (Other)	0.13	0.28	0.49	0.63
Tone (SharpR) × MinimalUnit (Ft)	–0.15	0.10	–1.45	0.15
Tone (Low) × MinimalUnit (Ft)	0.44	0.11	3.89	<0.001
Tone (SharpR) × MinimalUnit (Ph)	0.04	0.09	0.48	0.63
Tone (Low) × MinimalUnit (Ph)	–0.20	0.09	–2.27	0.02
Timing (Short) × MinimalUnit (Ft)	0.13	0.10	1.30	0.19
Timing (Short) × MinimalUnit (Ph)	0.04	0.07	0.67	0.50
Timing (Short) × Syntax (Other)	0.13	0.07	2.02	0.04

Table 2: Fixed effects of the final model: Response ~ Tone + Timing + MinimalUnit + NormingBias + Syntax + Tone × MinimalUnit + Timing × Syntax + Timing × MinimalUnit + (1| Subject) + (1| Sentence).

The model reveals a significant main effect of Tone: Compared to the grand mean, the major-juncture reading becomes less likely when the first Tone 3 syllable bears a sharp-rising tone contour ($\beta = -0.33, p < 0.001$) and more likely when it bears a low tone contour ($\beta = 0.38, p < 0.001$). To resolve comparisons of interest, we further treatment-coded the variable of Tone with ShallowR as the reference level and refit the model. The results further suggest that, compared to the first Tone 3 syllables with a shallow-rising tone contour, those with a sharp-rising contour significantly dampen the major-juncture reading ($\beta = -0.29, p < 0.01$) whereas syllables with a low tone contour significantly boost the major-juncture interpretation ($\beta = 0.42, p < 0.001$).

In addition, **Table 2** also shows a significant effect of the Timing cue on the probability of the major-juncture reading. Across the board, shorter duration of the first Tone 3 syllable leads to lower probabilities of major-juncture interpretations ($\beta = -0.14, p = 0.003$). NormingBias also turns out to be statistically significant ($\beta = 0.44, p = 0.02$), indicating that the major-juncture interpretation rate of a sentence in silent reading is a reliable predictor of major-juncture interpretation responses of that sentence in speech perception. MinimalUnit is found to have no significant main effect on its own.

Two interaction items, Tone × MinimalUnit and Timing × Syntax, proved to be significant in our generalized logistic mixed-effects model. Post-hoc tests were further performed to evaluate the differences between multiple groups in these interactions. **Figure 4** shows the mean major-juncture interpretation rates by tone and the minimal Tone 3 grouping unit of the sentence. **Table 3** shows the results of post-hoc pairwise comparisons between tone conditions for sentences with different minimal Tone 3 grouping units.

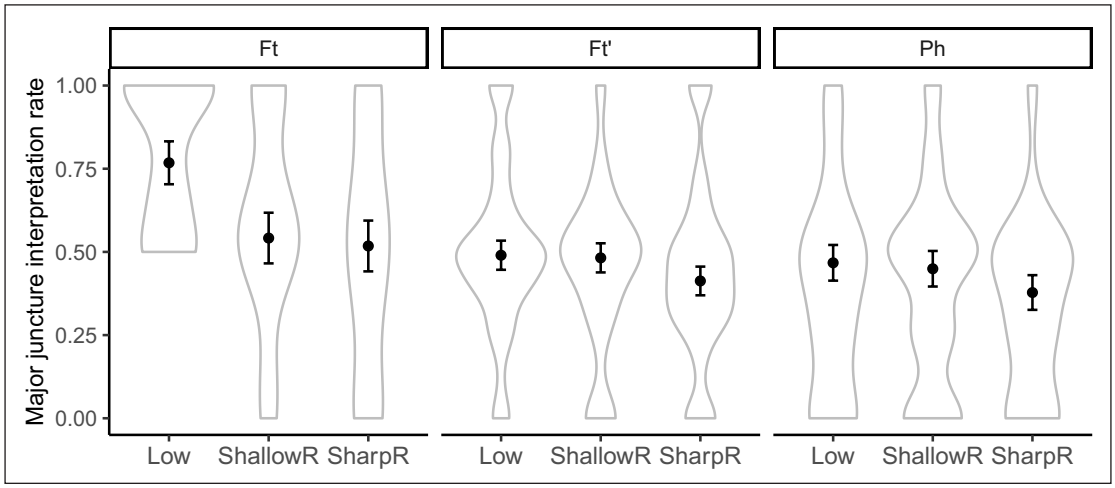

Figure 4: The mean and distribution of major-juncture interpretation rates by tone and minimal Tone 3 grouping unit. Error bars indicate 95% confidence intervals. Violins indicate the distributions of mean major-juncture response rates by participant.

Contrast	Pair	Estimate	*SE*	z.ratio	p.value
Ft	Sharp Rising – Low	–1.31	0.27	–4.79	< 0.001
	Sharp Rising – Shallow Rising	–0.15	0.24	–0.63	0.80
	Low – Shallow Rising	1.15	0.27	4.25	< 0.001
Ph	Sharp Rising – Low	–0.46	0.19	–2.40	0.04
	Sharp Rising – Shallow Rising	–0.41	0.18	–2.25	0.06
	Low – Shallow Rising	0.05	0.19	0.26	0.97
Ft’	Sharp Rising – Low	–0.37	0.15	–2.43	0.04
	Sharp Rising – Shallow Rising	–0.31	0.14	–2.17	0.08
	Low – Shallow Rising	0.06	0.15	0.41	0.91

Table 3: Posthoc pairwise test results of the model fit differences between tone conditions at each MinimalUnit contrast level (with Tukey’s adjustment).

According to these results, for sentences where the Tone 3 syllables can be grouped into a foot, the differences in response are significant both between the Low and the Sharp-rising conditions (*difference* = 25%, p < 0.001) and between the Low and the Shallow-rising conditions (*difference* = 22.6%, p < 0.001). No significant difference was detected between the two rising tone conditions (*difference* = 2.4%, $p = 0.80$). However, this situation is not shared by the superfoot or the phrase condition. In these two MinimalUnit conditions, the only significant difference in pairwise comparisons between the three tone conditions lies between Low and Sharp-rising (Ft': *difference* = 7.7%, $p = 0.04$; Ph: *difference* = 8.9%,, $p = 0.04$). No significant difference was observed between the Low and Shallow-rising conditions (Ft': *difference* = 0.8%, $p = 0.91$; Ph: *difference* = 1.8%, $p = 0.97$) or between the Shallow-rising and Sharp-rising conditions (Ft': *difference* = 6.9%, $p = 0.08$; Ph: *difference* = 7.1%, $p = 0.06$).[7]

Turning to the interaction between Timing and Syntax, **Figure 5** shows the mean major-juncture interpretation rates by Timing and Syntax. **Table 4** shows the results of post-hoc pairwise comparison between timing conditions for sentences in different categories of syntactic structures. The post-hoc pairwise comparison with Tukey correction reveals a significant difference between timing conditions for sentences in a [Verb NP1 DE NP2] structure (*difference* = 7.9%, $p = 0.002$). For sentences with other types of syntactic structures, however, the post-hoc test reveals no significant difference between the timing conditions (*difference* = 4.4%, p = 0.96).

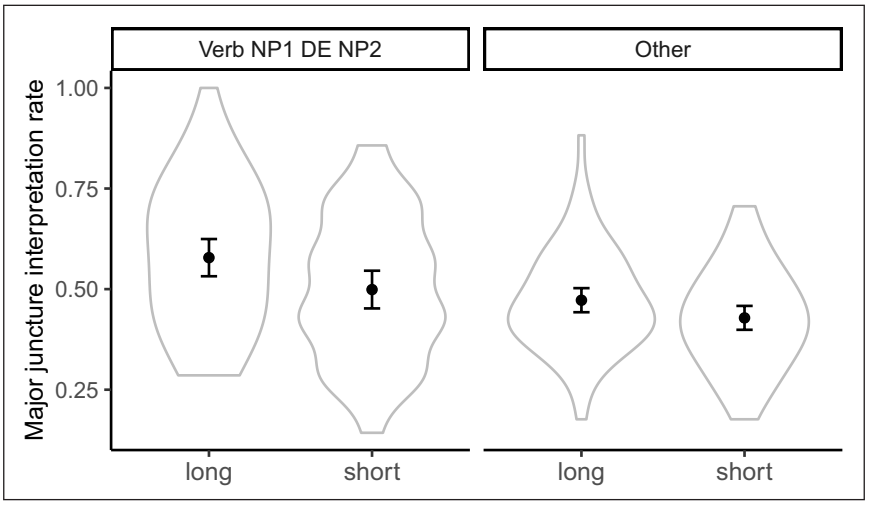

Figure 5: The mean and distribution of major-juncture interpretation rates by timing and syntax. Error bars indicate 95% confidence intervals. Violins indicate the distribution of mean major-juncture response rates by participant.

[7] We also tried this analysis with Bonferroni correction and Holm correction. The results with Bonferroni correction patterned with the current pattern. The results with Holm correction showed significant difference between SharpR and ShallowR for structures with a minimal unit of superfoot and phrase.

Contrast	Pair	Estimate	*SE*	z.ratio	p.value
Verb NP1 DE NP2	Short – Long	–0.54	0.17	3.13	0.002
Other	Short – Long	–0.01	0.15	0.05	0.96

Table 4: Posthoc pairwise test results of the model fit differences between timing conditions at each Syntax contrast level (with Tukey's adjustment).

5. Discussion

In this paper, we investigated whether Chinese listeners integrate the prosody-covarying cue of Tone 3 sandhi (T3S) to guide sentence parsing in speech perception. This question was evaluated with a sentence disambiguation task, in which the interpretations of ambiguous sentences depend on the presence versus absence of a major prosodic juncture between two consecutive Tone 3 syllables in these sentences. The F_0 of the first Tone 3 syllable was manipulated into three types of Tone 3 variants: Low (T3S does not apply), SharpR (T3S applies and bears a Sharp-rising F_0 contour), and ShallowR (T3S applies and bears a shallow-rising F_0 contour). Based on the covariation between Tone 3 sandhi and Chinese prosodic structure (Kuang & Wang, 2006; Shih, 1986), we predicted that listeners' interpretations can be affected by both the phonological alternation and the phonetic variation of T3S. The phonological alternation of T3S refers to whether the sandhi rule applies to change the low tone to a rising tone, whereas the phonetic variation of T3S refers to how sharp the F_0 slope of the rising sandhi variant is when T3S applies. We begin our discussion by responding to our research questions laid out at the beginning of Section 3 with the findings of the current experiment. We also spell out several preliminary findings beyond the scope of our current goals that have the potential to pave the way for future research.

Our first question regards whether listeners make use of the covariation between prosodic structure and the probability of T3S application and attend to the phonological alternations of Tone 3 to infer prosodic boundary strength. Evidence in favor of a positive answer would be obtained if participants showed higher rates of major-juncture interpretations in the Low tone condition than in the Rising tone conditions. The results of the mixed-effects logistic model support this possibility by showing that the Low Tone condition has a higher major-juncture interpretation rate than the ShallowR Tone condition (under treatment coding) and the grand mean of the three tone conditions (under sum coding). This means that listeners are more likely to arrive at a meaning that entails a major prosodic boundary between the Tone 3 syllables when T3S does not apply.

Further, the model demonstrates an interesting interaction between the Low Tone condition and the minimal grouping unit of Tone 3 syllables (MinimalUnit), which we illustrate in

Figure 4: For sentences with a MinimalUnit of foot, a major-juncture intepretation is chosen overwhelmingly in the Low condition whereas this interpretation occurs statistically at chance in SharpR and ShallowR conditions. For sentences with a MinimalUnit of superfoot or phrase, this preference for major-juncture interpretations in the Low tone condition disappears. This is because T3S is mandatory within the domain of the foot. A low tone on the first syllable indicates that T3S does not apply, and therefore, the Tone 3 syllables must be in different feet in order for the sentence to be grammatical. For sentences with a MinimalUnit of foot, the presence of a foot boundary between the Tone 3 syllables gives rise to a major-juncture interpretation. Note that the strengths of the absence and the presence of T3S are not symmetrical. The lack of T3S (i.e., the Low condition) entails that the two Tone 3 syllables must not be in the same foot. However, the presence of T3S (i.e., the ShallowR and SharpR conditions) does not tell us whether the two syllables are within the same foot or not because T3S can apply in both conditions. That is why only the Low condition, but not the Rising conditions, affects the responses remarkably for sentences with a MinimalUnit of foot. For sentences with a MinimalUnit larger than foot, the absence of T3S becomes less informative because the presence and absence of T3S are both allowed under each interpretation and they only vary in probability. That being said, we want to provide the caveat that this experiment is not designed to evaluate the interaction between T3S and local prosodic grouping precisely. This finding of an interaction between T3S's disambiguation strength and local foot formation possibilities is only preliminary and should be tested with a more careful design in future research.

Our second question regards whether listeners integrate the F_0 slope of T3S variants to infer prosodic boundary strength. Evidence that listeners attend to the phonetic realization of T3S in prosodic parsing would be obtained if participants gave higher rates of major-juncture interpretations in the ShallowR condition than in the SharpR condition. The results of our mixed-effects logistic model lend support to this possibility by showing that the ShallowR Tone condition has a higher major-juncture interpretation rate than the SharpR condition (under treatment coding) and the grand mean of all the three tone conditions (under sum coding). Post-hoc comparisons between the two rising tone conditions show that the SharpR condition has a lower major-juncture interpretation rate than the ShallowR condition when the minimal Tone 3 grouping unit is larger than foot. These findings suggest that listeners attend to the slope of the T3S variants to guide their parsing decisions for sentences with a minimal Tone 3 grouping unit larger than foot. We did not find any F_0 slope effect for sentences where Tone 3 syllables can be grouped into a foot. Considering the interaction between the Low tone effect and foot formation as discussed above, these results suggest that the informativeness of different aspects of Tone 3 variants for parsing may vary with prosodic contexts.

Regarding our third question of whether tone and timing interact with each other in cueing a prosodic boundary, our statistical analysis reveals that these two factors simultaneously have an effect on prosodic parsing and that these effects work in parallel. In the Short timing condition, where the first syllable is temporally compressed, the major-juncture interpretation rate becomes significantly lower than in the Long condition, where the first syllable is not compressed. In terms of whether the effects of tone and timing are independent or interactive, our results seem to suggest that the two mechanisms work in parallel without much interaction. In the model selection process, we find that adding Tone × Timing as a predictor does not significantly improve the model fit, suggesting that the two effects are independent of each other.

Instead of interaction between Tone and Timing, we find that Timing interacts with the Syntax of the stimuli: The effect that a shorter preboundary Tone 3 syllable leads to lower major-juncture interpretation rates is only significant for sentences with a structure of *[Verb NP1 DE NP2]*. This finding is not surprising, given the previous finding that the interaction between syntactic structure and prosodic boundary influences sentence parsing (e.g., Buxó-Lugo & Watson, 2016). At the same time, it opens up a broad line of inquiry about how much this finding can be generalized to other specific types of syntactic junctures and prosodic boundaries in disambiguation contexts. In our case, despite the statistical significance of the interaction between Timing and Syntax, in general, Timing seems to affect the responses in a constant manner across different types of sentences (i.e., the longer the first syllable, the higher the major-juncture rate). It is unclear to what extent the two types of syntactic structures respond to Timing in *qualitatively* different ways. Further, questions remain as to which syntactic structures under the label of 'Other' are the real driving force for this interaction. Unfortunately, given the limitations of our sample size, further cross-tabulation cannot be done at this point. Future studies are needed to follow up on these questions with stimuli well balanced in terms of syntactic structures.

Unsurprisingly, we find that in addition to the above factors, listeners' responses also depend on the intrinsic semantic bias of the critical sentences, which is captured by the comprehension responses of a separate group of participants via silent reading (NormingBias). It is also worth noting that the aggregate pattern, which turns out to be consistent with our prediction, actually masks considerable individual variability. Not all the participants have shown a trend consistent with the aggregate pattern. An interesting further avenue to pursue is to examine the individual variation of the integration of tone sandhi information and evaluate how individuals' strategies are related to their language background and experience.

Finally, we provide several implications of the present findings regarding how they fit in the theoretical context of the previous literature. First, our results shed light on the nature of

the Mandarin Tone 3 sandhi and lend further support to a prosody-based account of the tone sandhi domain. In specific, we show that listeners have good knowledge of the covariation between prosodic structure and Tone 3 sandhi variants in both probabilistic (i.e., the likelihood of application) and phonetic (i.e., the specific pitch shape) aspects and use this knowledge of covariation effectively in sentence parsing. Moreover, these results extend the substantial body of literature on prosody and auditory sentence processing by incorporating tone sandhi into the established inventory of prosodic cues to boundary strength. Compared to commonly investigated prosodic cues such as pitch reset and preboundary lengthening, the integration of tone sandhi as a cue of prosodic strength seems to require sophisticated linguistic abilities, ranging from grammaticality judgments and probability matching to phonetic sensitivity. The multidimensionality of the integration of tone sandhi makes it an appealing candidate to be further investigated along this line of research. Finally, our results add to our current understanding of listeners' knowledge of speech variability. We propose that listeners have a sophisticated knowledge of speech variability derived from both inter-category phonological alternation and intra-category phonetic variation; they make efficient use of this knowledge once it becomes linguistically relevant.

Abbreviations

T3S: The Mandarin Tone 3 sandhi

F_0: Fundamental frequency

Ft: Foot

Ft': Superfoot

Ph: Phrase

POSS: Possessive case (translated as *'s* in the source)

RC: Relative clause

1/2/3 PPL/PSG: First/second/third person plural/singular

Additional file

The additional file for this article can be found as follows:

- **Appendix.** A pdf file of a list of all the test and filler sentences, including their Chinese characters, Pinyin transcription, word-by-word gloss, and the two possible interpretations. DOI: https://doi.org/10.16995/labphon.6416.s1

Acknowledgements

The authors want to thank the audience at the 33rd Annual CUNY Human Sentence Processing Conference and three anonymous reviewers for their useful feedback.

Competing interests

The authors have no competing interests to declare.

Author contributions

The first author was responsible for most of the work on experimental design and stimulus manipulation. The second author did most of the work on experiment setup and data collection. The two authors both contributed to the statistical analysis and theoretical interpretation of the results, as well as the writing of this paper.

References

Baek, H. (2019). A cross-linguistic comparison on the use of prosodic cues for ambiguity resolution. In *Proceedings of meetings on acoustics 177asa* (Vol. 36, p. 060005). DOI: https://doi.org/10.1121/2.0001094

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of memory and language*, *30*(6), 644–663. DOI: https://doi.org/10.1016/0749-596X(91)90030-N

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255–309. DOI: https://doi.org/10.1017/S095267570000066X

Bishop, J. (2020). Exploring the similarity between implicit and explicit prosody: Prosodic phrasing and individual differences. *Language and Speech*, 0023830920972732. DOI: https://doi.org/10.1177/0023830920972732

Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, *90*, 1–13. DOI: https://doi.org/10.1016/j.jml.2016.03.001

Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of Dutch. *Linguistics in the Netherlands*, *14*(1), 13–24. DOI: https://doi.org/10.1075/avt.14.04cam

Chao, Y. R. (1968). *Language and symbolic systems* (Vol. 457). CUP Archive.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486364

Cheng, C.-c. (1973). A quantitative study of Chinese tones. *Journal of Chinese linguistics*, 93–110.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *32*(2), 141–176. DOI: https://doi.org/10.1016/S0095-4470(03)00043-3

Chow, I. (2004). *Prosodic structures of French and Mandarin* (Doctoral dissertation). University of Toronto.

Chow, I. (2006). Interactions between syllabic-level prosody and prosodic-group boundary markers in cantonese. In *Proceedings of the 2006 annual conference of the Canadian linguistic association.*

Chow, I. (2008). Quantitative analysis of preboundary lengthening in cantonese. In *Proceedings of the speech prosody 2008 conference* (pp. 543–546).

Chow, I. (2018). An investigation on syntactic disambiguation in Mandarin speech perception and the phonological status of the disyllabic foot. *Journal of Chinese Linguistics*, *46*(2), 269–291. DOI: https://doi.org/10.1353/jcl.2018.0010

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, *40*(2), 141–201. DOI: https://doi.org/10.1177/002383099704000203

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of phonetics*, *24*(4), 423–444. DOI: https://doi.org/10.1006/jpho.1996.0023

Domínguez, L. (2004). *Mapping focus: The syntax and prosody of focus in spanish* (Unpublished doctoral dissertation). Boston University.

Duanmu, S. (2005). The tone-syntax interface in Chinese: some recent controversies. In *Proceedings of the symposium cross-linguistic studies of tonal phenomena* (pp. 221–54).

Elfner, E. (2012). *Syntax-prosody interactions in Irish* (Doctoral dissertation). University of Massachusetts Amherst, Amherst, MA.

Feng, S. (1998). On natural foot in Chinese. *Zhongguo Yuwen*, *1*, 40–47.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The journal of the acoustical society of America*, *101*(6), 3728–3740. DOI: https://doi.org/10.1121/1.418332

Hart, J., Collier, R., & Cohen, A. (2006). *A perceptual study of intonation: an experimentalphonetic approach to speech melody*. Cambridge University Press.

Hofhuis, E. M., Gussenhoven, C., & Rietveld, T. (1995). Final lengthening at prosodic boundaries in Dutch. In E. Elenius & P. Branderud (Eds.), *Proceedings of the xiiith international congress of phonetic sciences* (p. 154–157). Stockholm: Stockholm University.

Hsieh, Y., Boland, J. E., Zhang, Y., & Yan, M. (2009). Limited syntactic parallelism in Chinese ambiguity resolution. *Language and Cognitive Processes*, *24*(7–8), 1227–1264. DOI: https://doi.org/10.1080/01690960802050375

Hsu, B. (2013). Constraining exceptionality as prosody-morphology mismatch: a study of French nasal vowels. *Proceedings of CLS*, *49*.

Hung, T. T. (1987). *Syntactic and semantic aspects of Chinese tone sandhi*. University of California, San Diego, Department of Linguistics.

Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language & Linguistic Theory*, *7*(2), 217–259. DOI: https://doi.org/10.1007/BF00138077

Kim, H., Yoon, T., Cole, J., & Hasegawa-Johnson, M. (2006). Acoustic differentiation of l-and ll% in switchboard and radio news speech. In *Proceedings of speech prosody* (pp. 214–7).

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of phonetics*, *3*(3), 129–140. DOI: https://doi.org/10.1016/S0095-4470(19)31360-9

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive psychology*, *50*(2), 194–231. DOI: https://doi.org/10.1016/j.cogpsych.2004.08.002

Kuang, J. (2005). *Rhythm organization for t3 sandhi of Mandarin Chinese: syntax rules and prosodic rules* (Undergraduate thesis). Peking University, Columbus, OH.

Kuang, J. (2010). Prosodic grouping and relative clause disambiguation in Mandarin. In *Eleventh annual conference of the international speech communication association*. DOI: https://doi.org/10.21437/Interspeech.2010-501

Kuang, J. (2017). Creaky voice as a function of tonal categories and prosodic boundaries. In *Interspeech* (pp. 3216–3220). DOI: https://doi.org/10.21437/Interspeech.2017-1578

Kuang, J., & Wang, H. (2006). T3 sandhi at the boundaries of different prosodic hierarchies. *J. Chinese Phonetics*, *1*, 125–131.

Lai, C., Sui, Y., & Yuan, J. (2010). A corpus study of the prosody of polysyllabic words in Mandarin Chinese. In *Speech prosody 2010-fifth international conference.*

Lai, W., & Kuang, J. (2016). Prosodic grouping in Chinese trisyllabic structures by multiple cues–tone coarticulation, tone sandhi and consonant lenition. *Tonal Aspects of Languages 2016*, 157–161. DOI: https://doi.org/10.21437/TAL.2016-34

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *The Journal of the Acoustical Society of America*, *53*(1), 380–380. DOI: https://doi.org/10.1121/1.1982702

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). *Emmeans: Estimated marginal means, aka least-squares means. 2018; r package version 1.3. 1.*

Macdonald, N. (1976). Duration as a syntactic boundary cue in ambiguous sentences. In *Icassp'76. ieee international conference on acoustics, speech, and signal processing* (Vol. 1, pp. 569–572). DOI: https://doi.org/10.1109/ICASSP.1976.1170024

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of experimental psychology*, *45*(1), 73–87. DOI: https://doi.org/10.1080/14640749208401316

Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica*, *38*(1–3), 84–105. DOI: https://doi.org/10.1159/000260016

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *The journal of the Acoustical Society of America*, *54*(5), 1235–1247. DOI: https://doi.org/10.1121/1.1914393

O'Malley, M., Kloker, D., & Dara-Abrams, B. (1973). Recovering parentheses from spoken algebraic expressions. *IEEE Transactions on audio and electroacoustics*, *21*(3), 217–220. DOI: https://doi.org/10.1109/TAU.1973.1162449

Pratt, E. (2017). Prosody in sentence processing. *The Handbook of Psycholinguistics*, 365–391. DOI: https://doi.org/10.1002/9781118829516.ch16

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *the Journal of the Acoustical Society of America*, *90*(6), 2956–2970. DOI: https://doi.org/10.1121/1.401770

Pynte, J. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and cognitive processes*, *11*(1–2), 165–192. DOI: https://doi.org/10.1080/016909696387259

Qualtrics. (2020). *Qualtrics.* Retrieved from https://www.qualtrics.com (Version 2020-12-4)

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

Schafer, A., Carlson, K., Clifton Jr, C., & Frazier, L. (2000). Focus and the interpretation of pitch accent: Disambiguating embedded questions. *Language and speech*, *43*(1), 75–105. DOI: https://doi.org/10.1177/00238309000430010301

Schafer, A., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of psycholinguistic research*, *29*(2), 169–182. DOI: https://doi.org/10.1023/A:1005192911512

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *The Journal of the Acoustical Society of America*, *71*(4), 996–1007. DOI: https://doi.org/10.1121/1.387581

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population marginal means in the linear model: an alternative to least squares means. *The American Statistician*, *34*(4), 216–221. DOI: https://doi.org/10.1080/00031305.1980.10483031

Shen, X. S. (1993). The use of prosody in disambiguation in Mandarin. *Phonetica*, *50*(4), 261–271. DOI: https://doi.org/10.1159/000261946

Shih, C. (1986). *The prosodic domain of tone sandhi in Chinese*. University of California, San Diego.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of psycholinguistic research*, *25*(2), 249–271. DOI: https://doi.org/10.1007/BF01708573

Speer, S. R., Shih, C.-L., & Slowiaczek, M. L. (1989). Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Language and Speech*, *32*(4), 337–354. DOI: https://doi.org/10.1177/002383098903200403

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America*, *64*(6), 1582–1592. DOI: https://doi.org/10.1121/1.382142

Surányi, B., Ishihara, S., & Schubö, F. (2012). Syntax-prosody mapping, topic-comment structure and stress-focus correspondence in hungarian. In *Prosody and meaning* (pp. 35–72). De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110261790.35

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*(4), 397–440. DOI: https://doi.org/10.1006/jpho.2000.0123

Vaissière, J. (1983). Language-independent prosodic features. In *Prosody: Models and measurements* (pp. 53–66). Springer. DOI: https://doi.org/10.1007/978-3-642-69103-4_5

Van der Hulst, H. (2016). Vowel harmony. In *Oxford research encyclopedia of linguistics.* Voeten, C. C. (2020). *Using 'buildmer'to automatically find & compare maximal (mixed) models*. DOI: https://doi.org/10.1093/acrefore/9780199384655.013.38

Wang, B., Xu, Y., & Ding, Q. (2018). Interactive prosodic marking of focus, boundary and newness in Mandarin. *Phonetica*, *75*(1), 24–56. DOI: https://doi.org/10.1159/000453082

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. Retrieved from http://ggplot2.org. DOI: https://doi.org/10.1007/978-3-319-24277-4

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717. DOI: https://doi.org/10.1121/1.402450

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of memory and language*, *72*, 37–48. DOI: https://doi.org/10.1016/j.jml.2013.12.003

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of phonetics*, *25*(1), 61–83. DOI: https://doi.org/10.1006/jpho.1996.0034

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech communication, 33*(4), 319–337. DOI: https://doi.org/10.1016/S0167-6393(00)00063-7

Yang, Y., & Wang, B. (2002). Acoustic correlates of hierarchical prosodic boundary in Mandarin. In *Speech prosody 2002, international conference.*

Yuan, J., & Chen, Y. (2014). 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics*, *42*(1), 218–237.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics 2008*, 5687–5690. DOI: https://doi.org/10.1121/1.2935783

Yuan, J., & Liberman, M. (2015). Investigating consonant reduction in Mandarin Chinese with improved forced alignment. In *Sixteenth annual conference of the international speech communication association.* DOI: https://doi.org/10.21437/Interspeech.2015-401

Zhang, H. (2016). Boundary effects on allophonic creaky voice: A case study of Mandarin lexical tones. *Tonal Aspects of Languages 2016*, 94–98. DOI: https://doi.org/10.21437/TAL.2016-20

Zhang, H. M. (2016). *Syntax-phonology interface: argumentation from tone sandhi in Chinese dialects*. Routledge. DOI: https://doi.org/10.4324/9781317389019

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 153–201. DOI: https://doi.org/10.1017/S0952675710000060

Zhang, X. (2012). *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese.* (Unpublished doctoral dissertation). University of Michigan.

Zhang, Z.-s. (1988). *Tone and tone sandhi in Chinese* (Unpublished doctoral dissertation). The Ohio State University.

Zhu, X., Zhang, T., & Yi, L. (2012). A classification of dipping tones [j]. *Studies of the Chinese Language, 5*.