Integrating phonetic and phonological aspects of Mandarin third tone sandhi in auditory sentence disambiguation (Wei Lai & Ani Li, University of Pennsylvania)

Psycholinguistic studies have shown that prosodic manifestations affect sentence comprehension by signaling syntactic breaks and constituent-internal grouping [1]. This study investigated whether listeners could integrate variants of the Mandarin third tone sandhi (T3S), a prosody-covarying phonological variable, to facilitate structural disambiguation in auditory sentence comprehension. T3S is the process where the first of two consecutive low tones changes into a rising tone. It is sensitive to prosodic structure both in phonological derivation and in phonetic realization. Phonologically, T3S applies mandatorily within foot and optionally across feet [2], with a tendency to occur more frequently across higher-level prosodic junctures than lower-level ones [3]; phonetically, T3S variants at smaller prosodic boundaries tend to have a sharper rising pitch whereas those at bigger prosodic boundaries have a shallower rising pitch [3].

This study examined whether Mandarin listeners would integrate the covariation between T3S variants and prosodic structure to assist syntactic parsing and disambiguation, by attending to a) whether T3S applies, and b) the pitch shape of T3S variants if it applies. 60 participants from mainland China attended an auditory sentence comprehension experiment online, where they heard ambiguous sentences and needed to choose from two written interpretations the one consistent with what they heard. The critical stimuli were 27 ambiguous sentences, each containing two consecutive T3 syllables. Crucially, the two interpretations for each of them differ in whether the two T3 syllables were intervened with a major syntactic juncture (see (1) for an example). Each critical stimulus was manipulated into 3 tonal (sharp-rising, shallow-rising, and low; Fig. 1) by 2 timing (whether the first T3 syllable was shortened or not) conditions. Three lists were constructed, across which the tonal conditions of stimuli were counterbalanced. Listeners were assigned randomly to one of the lists such that each listener only heard each stimulus in one tonal condition.

We predicted that listeners were more likely to obtain a major-juncture interpretation when tone sandhi did not apply (*low*) than when it did, and, in the latter case, when it had a *shallow-rising* than a *sharp-rising* pitch. This is what we found (Fig. 2). A mixed-effects logistic model was conducted with the major-juncture reading response as the dependent variable, tone-nested-in-timing as the main effect (timing: shortened-0 normal-1; tone: repeated coded, sharp-shallow-low), and sentence and subject-nested-in-group as random intercepts. The result revealed a significant timing effect (β= 0.32, $p<.001$), meaning that listeners reported fewer major-juncture interpretations when the boundary-final syllable was shortened. In the normal timing condition, the major-juncture reading was significantly more frequent in the *low* than *shallow-rising* tone condition (β=0.3, p=.002), and marginally more frequent in the *shallow-rising* than *sharp-rising* tone condition (β=0.23, p=.09). In the shortened timing condition, the major-juncture reading rate was significantly higher in the *shallow-rising* than *sharp-rising* tone condition (β= 0.35, p=.01), but showed no difference between *shallow-rising* and *low* conditions (β= 0.09).

The integration of T3S was further found to interact with sentence length. Fig. 3 shows that the cue of phonological derivation (low versus non-low) was robust for short sentences (4-5 syllables long), but not for longer (more than 5 syllables). This is because short sentences tend to break down prosodically into two feet (2+2) or a foot plus a superfoot (2+3, 3+2), where T3S is mandatory. In other words, the grouping of T3 syllables mandatorily causes T3S to apply in short sentences, which makes the non-application of T3S a strong cue of prosodic segmentation (see (2) tor an example). These findings suggest that listeners have a sophisticated knowledge of speech variation from both phonological derivations and phonetic implementations, and made efficient use of this knowledge once it became linguistically relevant.

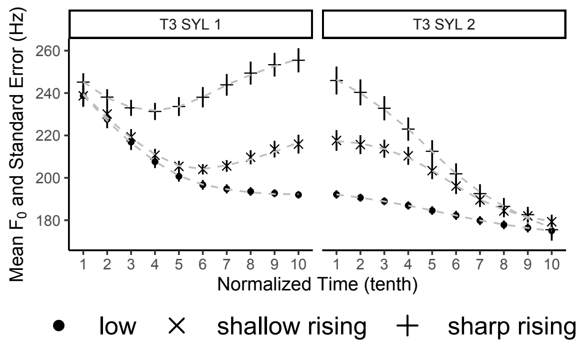


Fig. 1. The F_0 manipulation of the T3 syllables

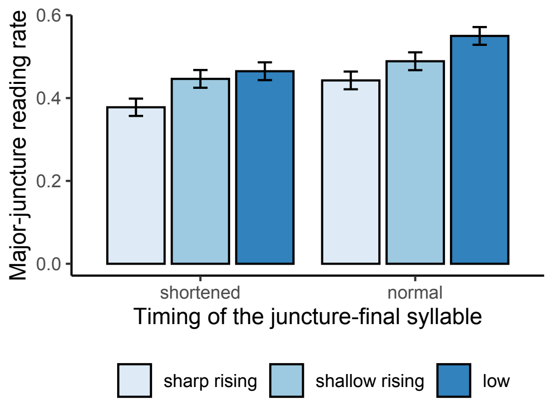
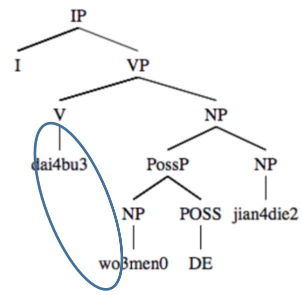


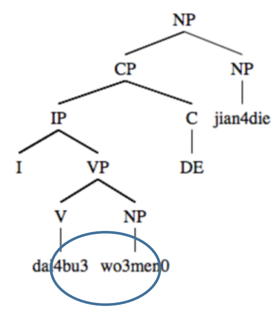
Fig. 2. Major-juncture reading rates on different conditions

(1)	逮捕	我们	的	间谍
	dai4bu3	wo3men	de	jian4die2
	arrest	1PPL	POSS/RC	*spy*

1a. Major-juncture: 'Arrest our spy'



1b. Minor-juncture: 'The spy that arrest us.'



(2)	Pinyin:	wo3	xie3	bu4	hao3	(我写不好)
	Gloss:	1pps	write	NEG	good/well	
	Parsing 1:	(wo3	xie3)	(bu4	hao3): 'It is not good that I write.' (mandatory T3S)	
	Parsing 2:	(wo3)	(xie3	bu4	hao3): 'I cannot write (it) well.' (optional T3S)	

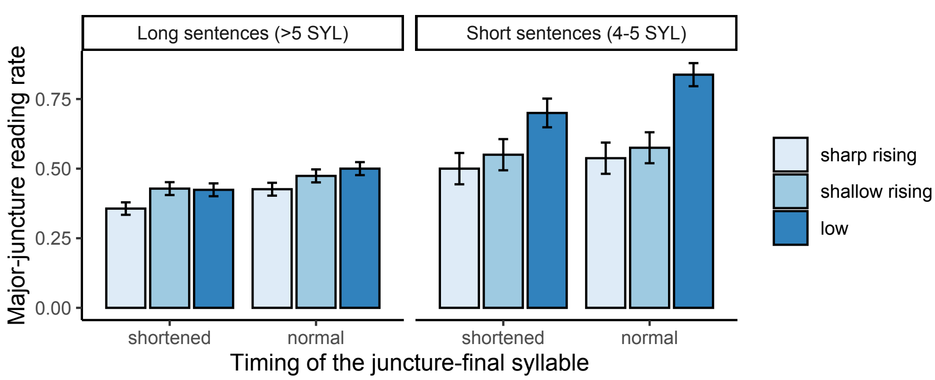


Fig. 2. Major-juncture reading rates in different conditions for short and long sentences

Selected Reference: 1. Cutler et al. (1997). Prosody in the comprehension of spoken language: A literature review. 2. Shih. (1997). Mandarin third tone sandhi and prosodic structure. 3. Kuang, Jianjing & Hongjun Wang (2006). T3 sandhi at the boundaries of different prosodic hierarchies. 4. Hsieh et al. (2009) Limited syntactic parallelism in Chinese ambiguity resolution.