The production and perception of tonal variation -: evidence from Tianjin Mandarin

Li, Qian

Cover Page

The handle http://hdl.handle.net/1887/37609 holds various files of this Leiden University dissertation

Author: Qian Li
Title: The production and perception of tonal variation : evidence from Tianjin Mandarin
Issue Date: 2016-02-10

CHAPTER 5 THE ONLINE PERCEPTION OF SANDHI TONES IN TIANJIN MANDARIN – EVIDENCE FROM EYE MOVEMENTS

5.1 Introduction

In tonal languages that utilize melodic pitch contour changes to convey lexical meanings such as Mandarin Chinese, the main acoustic correlate of lexical tones - fundamental frequency (*f0*) changes - varies extensively in connected speech. Part of the variation can be due to the global prosody of the entire utterance (e.g., Xu, 1999; Yuan, 2004; Chen, 2003; Chen & Gussenhoven, 2008; Chen, 2010; Scholz & Chen, 2014a, 2014b; also see reviews in Chen, 2012; Zsiga, 2012). Others are more localized where the *f0* realization of a lexical tone varies as a function of its neighboring tonal contexts. One such contextual tonal alternation process is known as "tone sandhi", commonly regarded as the phonologized change of lexical tones in certain tonal contexts. Tone sandhi usually introduces great changes in the *f0* contours of the sandhi-derived tones (see reviews in Chen, 2000; Zhang, 2010). A well-known case of tone sandhi is the "Third-Tone" sandhi in Standard Chinese. When two dipping tones (Tone 3, hereafter as T3) are combined, the first one surfaces with a rising *f0* contour (hereafter T3$_{sandhi}$), similar to that of the lexical rising tone (Tone 2, hereafter as T2) in the language.

A much-debated issue that is crucial for tonal alternation theories is whether tone sandhi such as the "Third-Tone" sandhi in Standard Chinese involves complete neutralization of otherwise distinctive lexical tones as a result of the tonal sandhi change. Previous perception studies have shown that native listeners cannot reliably distinguish the sandhi variant T3$_{sandhi}$ vs. the lexical T2 (e.g., Wang & Li, 1967; Speer et al., 1989; Chen, 2013; Chen et al., 2015). This leads to the received view of T3 sandhi change in Standard Chinese as the change of T3 to T2, in which T3$_{sandhi}$, as a result, is regarded to completely neutralize with the lexical T2 (e.g., Chen, 2000; Yip, 2002; Chen et al., 2015). This perception-based point of view, however, is challenged by evidence from tone production studies which demonstrate subtle but quite consistent *f0* differences between T3$_{sandhi}$ and the lexical T2 in both read speech (e.g., Zee, 1980; Peng, 1996) and corpus data (Yuan & Chen, 2014). Furthermore, T3$_{sandhi}$ and the lexical T2 have also been shown to be processed differently during speech encoding (Chen et al., 2011; Zhang et al., 2014). If complete neutralization was indeed involved in tone sandhi, such systematic production differences between the sandhi-derived tone (i.e., T3$_{sandhi}$) and the claimed output tone (i.e., T2) should not have been expected. However, it is still unknown so far whether the sandhi-derived tones are processed more similarly to its non-sandhi variants or to the claimed output tones.

In a study related to this issue, Zhou and Marslen-Willson (1997) investigated the representation of tonal sandhi variants in Standard Chinese via two auditory-auditory priming experiments, based on the assumption that T3 changes to T2 before another T3. Results of the first experiment show that, compared to the control condition where primes and targets have unrelated tones and segments, the processing of disyllabic T3$_{sandhi}$T3

targets were significantly facilitated by a T3Tx prime, where T3$_{sandhi}$ in the target and T3 in the prime share the toneme but differ in the surface $f0$ contour (Toneme+, Contour-). However, the targets were significantly inhibited by a T2Tx prime which surfaces with a similar $f0$ with T3$_{sandhi}$ in the targets but has a different toneme (Toneme-, Contour+). The first experiment thus suggests that T3$_{sandhi}$ should be represented similarly as the non-sandhi T3 rather than the lexical T2 as hypothesized by the authors (following the literature). In the following experiment with T2Tx as targets and T3Tx, T3$_{sandhi}$T3 and T2T3 as auditory primes, their results show that compared to the control condition, T3Tx (Toneme-, Contour-) showed a significantly more inhibitory effect than T3$_{sandhi}$T3 (Toneme-, Contour+) and T2T3 (Toneme+, Contour+), suggesting similarity between T2 with T3$_{sandhi}$ but not T3. Assuming that T3 changes to T2 in T3T3 sequences (i.e., T3$_{sandhi}$=T2), we find the results reported here rather puzzling. The relationship between T3$_{sandhi}$ and T2 (or T3) thus is worthy of further research efforts to be clarified.

It is to be noted that while the $f0$ information of lexical tones has been shown to be processed in an incremental fashion as segments in online speech recognition (Malins & Joanisse, 2010; Shen et al., 2013; Li & Chen, 2015), most previous studies on tone sandhi perception have focused on whether native listeners are able to discriminate between the sandhi tones (e.g., T3$_{sandhi}$ in Standard Chinese) and the so-called output tones (e.g., lexical T2 in Standard Chinese) in traditional meta-linguistic tone discrimination paradigms (e.g., Wang & Li, 1967; Speer et al., 1989). Zhou and Marslen-Wilson (1997) used an online auditory-auditory priming lexical decision paradigm, but only end-state responses were recorded. On the one hand, data of only end-state responses are inadequate to reveal the dynamic time-course differences in spoken word recognition prior to the end-state judgement (see further discussion in Malins & Joanisse, 2010). On the other hand, the end-state responses are likely to exhibit a different pattern from that of the real-time processing data, which could potentially lead to completely different understandings of the same phenomenon (Spivey, 2007). Further online speech recognition studies are therefore in great need to shed more light on how listeners recognize sandhi tones in real time.

To tap into this issue, the present study investigates the time course of online recognition of sandhi tones in Tianjin Mandarin. As will be introduced below, Tianjin Mandarin serves as an interesting test case here for its rich layers of contextual tonal sandhi variation patterns over disyllabic tonal sequences (e.g., Li & Liu, 1985; Chen, 2000; Wee et al., 2005; Zhang & Liu, 2011; Li & Chen, 2016).

5.2 Tianjin Mandarin

Tianjin Mandarin is a dialect of Mandarin Chinese, which is mainly spoken in the urban areas of Tianjin Metropolis, China. There are four lexical tones in Tianjin Mandarin: Tone 1 (T1) is a low-falling tone, Tone 2 (T2) a high-rising tone, Tone 3 (T3) a dipping tone and Tone 4 (T4) a high-falling tone (Li & Chen, 2016). Figure 5.1 illustrates the $f0$ realization of the four lexical tones uttered in isolation, based on 50 samples for each tone, produced by a male speaker in his 20s at the time of recording (born in 1983).
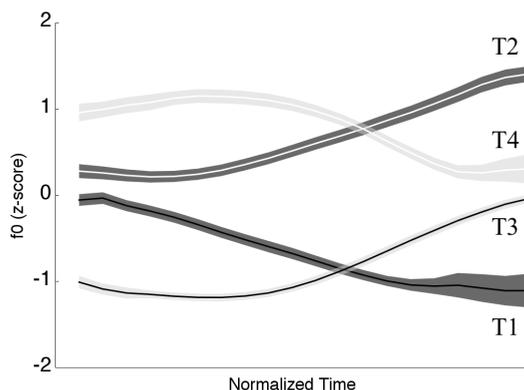
Figure 5.1 *Four lexical tones in Tianjin Mandarin produced in isolation (with normalized time). Lines stand for the mean. Gray areas stand for ±1 standard error of the mean.*

When two tones are combined, several of the disyllabic tonal sequences undergo tone sandhi changes from one tone to another in Tianjin Mandarin. For example, when two T3s are combined, the first one is realized with a high rising *f0*, resembling that of the lexical T2 with only subtle differences (Zhang & Liu, 2011; Li & Chen, 2016; but see e.g., Li & Liu, 1985; Chen, 2000; Hyman, 2007 which proposed a categorical change of lexical T3 to lexical T2). Furthermore, another type of tone sandhi has been reported for the tonal sequences of T1T1 and T4T1, where T1$_{sandhi}$ and T4$_{sandhi}$ are significantly different from their respective targeted output tones as claimed in the literature (i.e., T3 and T2, respectively) (Li & Chen, 2016; but see e.g., Li & Liu, 1985; Chen, 2000; Hyman, 2007 which categorize these tonal variants also as the categorical changes of lexical tones, i.e. T1 and T4, to T3 and T2, respectively). Given the different degrees of resemblance to another lexical tone within the tonal inventory (as illustrated in Figure 5.2 below), T3T3 has been classified as Near-Merger Sandhi while T1T1 and T4T1 as No-Merger Sandhi (Li & Chen, 2016).

Figure 5.2 illustrates the *f0* contours of the three tone sandhi sequences compared to that of their respective targeted output sequences as claimed in the literature. The data were averaged across 72 tokens (six speakers, three items, two informational status, two repetitions). White lines with dark areas represent the *f0* realization of the tone sandhi sequences; and black lines with light areas represent the claimed target output sequences. In the Near-Merger Sandhi case (i.e., the T3T3 sequence in Figure 5.2*a*), the first T3 is realized with a high-rising *f0* contour, which greatly resembles that of the first T2 in T2T3. The *f0* realization of T3T3 tonal sequence, therefore, shows great similarity to that of the claimed targeted output sequence (T2T3) in the literature despite that no complete neutralization is involved (Zhang & Liu, 2011; Li & Chen, 2016). In the No-Merger Sandhi cases (i.e., T1T1 and T4T1), both T1$_{sandhi}$ and T4$_{sandhi}$ surface with a much altered *f0* contour, while still maintaining tonal distinctiveness from their respective targeted output tonal contours claimed in the literature. Specifically, T1T1 (Figure 5.2*b*) has been described to change into T3T1 in the literature (e.g., Li & Liu, 1985; Chen, 2000; Hyman, 2007). As shown in Figure 5.2*b*, although the first T1 in T1T1 is realized similarly to T3 as in T3T1,

its *f0* realization is clearly different from that of T3, suggesting no merging of T1$_{sandhi}$ with T3. For T4T1 (Figure 5.2*c*), similarly, although the first T4 in T4T1 is similar to the claimed targeted output - T2 (e.g., Li & Liu, 1985; Chen, 2000; Hyman, 2007), its *f0* realization is clearly different from that of lexical T2.
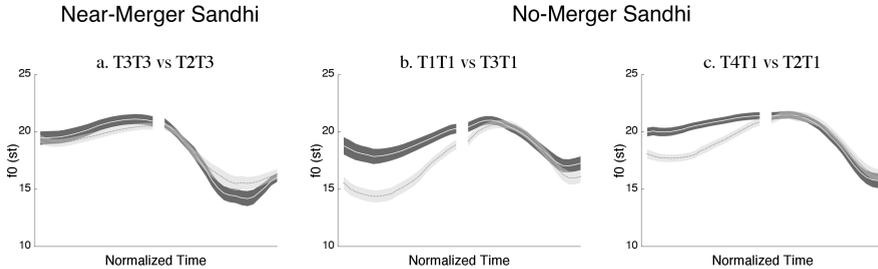
Near-Merger Sandhi   No-Merger Sandhi

Figure 5.2 *f0 realization of three disyllabic sandhi sequences in Tianjin Mandarin (T3T3 in a, T1T1 in b and T4T1 in c) compared to their respective sandhi target sequences (T2T3 in a, T3T1 in b, T2T1 in c) as claimed in the literature. T3T3 is Near-Merger Sandhi; T1T1 and T4T1 are No-Merger Sandhi. Thick white lines indicate the mean f0 of the sandhi sequences (dark gray areas for ±1 standard error of mean); black lines indicate the mean f0 of the claimed target sequences (light gray areas for ±1 standard error of mean). Normalized time.*

The differences between the two sandhi types have been found to significantly influence the recognition of the lexical tones as in our recent eye-tracking study (Li & Chen, 2015 and Chapter 4). The eye movement data show that listeners have more difficulty in processing Near-Merger sandhi tone than No-Merger sandhi tones, due to the fact that for Near-Merger sandhi, the sandhi-derived tone is more similar to another lexical tone within the lexical tonal inventory in terms of the *f0* realization, while the No-Merger sandhi tones remain distinctive from the lexical tones with the most similar *f0* contours (see below for further details).

This study taps further into the nature of tone sandhi recognition. To be specific, we ask whether the online processing of sandhi tones is in a similar way to that of the non-sandhi variants (Toneme+, Contour-) or another lexical tone which has been claimed as the output target of the tone sandhi rules due to *f0* contour similarity (Toneme-, Contour+), and whether different sandhi types might differ in how the sandhi tones are recognized. These should shed further light on how tonal variants are represented in the mental lexicon.

5.3 Visual world paradigm

To investigate the recognition of tone sandhi, we employed the "Visual World Paradigm" (Tanenhaus et al., 1995) with an auditory word-recognition task following Li and Chen (2015), in which the eye movements of listeners are tracked while they listen to auditory stimuli. As in many experiments with the Visual World Paradigm, participants are presented with multiple objects (either as images or in written words) on a computer screen. They are required to follow instructions to complete certain tasks with one of these

objects (i.e., the target) upon hearing an auditory stimulus corresponding to the object. Their eye movements are recorded simultaneously for later analyses. A typical display of visual stimuli consists of a target, a competitor as well as two distractors. The participants are asked to identify the words they have just heard and click on it with a mouse (see Huettig et al., 2011 for a review).

This paradigm has enabled us to tap into the time course of auditory word recognition based on the assumption that eye movements are closely time-locked to the spoken-word processing in an incremental fashion (e.g., Tanenhaus et al., 1995; Allopenna et al., 1998; Dahan et al. 2001a, 2001b; McMurray, Tanenhaus & Aslin, 2002; Beddor et al., 2013; see also reviews in Tanenhaus et al., 2000; Huettig et al., 2011). Furthermore, tasks involved in this paradigm are more natural than meta-linguistic judgement tasks in terms of perception experience (Spivey, 2007).

An important dependent variable in the visual world paradigm is the proportion of looks to each visual object over time - the likelihood of certain object being looked at during the time window. Two indices are found to effectively reflect how auditory stimuli are processed.

One is concerned with how fast listeners fixate on the target object, which is related to how early the target starts to be activated. For example, in Tanenhaus et al. (1995), it has been found that when the target object (e.g., *candy*) was presented together with another object whose name has phonological overlapping with the target (e.g., *candle*), participants took longer time (i.e., 230ms) to initiate an eye movement to the target object *candy* than when no phonologically similar object was presented (i.e., 145ms). This shows that when there exists a phonologically overlapping competitor, the activation of the target is delayed due to stronger competition effect in target recognition compared to when there is no such competitor.

The other important index is how likely listeners are looking at the targets (or competitors), which is believed to reflect the level of activation of the targets (or competitors). In Allopenna et al. (1998), participants were presented with four images on the computer screen, one corresponding to the auditory stimuli (target, e.g., *beaker*), one for "onset competitor" which has onset overlapping with the target (e.g., *beetle*, which shares the onset with *beaker*), and the rest two for distractors that are phonologically unrelated with the target (e.g., *dolphin*, *carriage*). When participants heard the instruction "*Pick up the beaker*", the proportion of looks to both *beaker* and *beetle* started to increase initially, indicating the start of activation for both words. However, as the speech unfolded, *beetle* was gradually deactivated and consequently, the proportion of looks to it started to decrease; while the target picture gained more looks given the increased activation of the target word.

This paradigm has been successfully used in recent studies on lexical tone perception, which demonstrate that listeners are sensitive to fine-grained *f0* details (Malins & Joanisse, 2010; Shen et al., 2013; Li & Chen, 2015). For example, Li and Chen (2015) investigated the effect of contextual tonal variation on the perception of lexical tones of Tianjin Mandarin in connected speech. Participants heard disyllabic auditory target stimuli, which undergo three types of tonal variation: Near-Merger Sandhi, No-Merger Sandhi and No

Sandhi coarticulation. Upon hearing each auditory stimulus, participants saw four disyllabic collocations on the computer screen: a Target which corresponds to the auditory stimulus (e.g., *zhi³ fa³* 'fingering'), a Critical Competitor of which the first syllable shares the segments but has an unrelated tone with the target (e.g., *zhi⁴ hou⁴* 'lag'), and two phonologically unrelated distractors (e.g., *san¹ jiao³* 'triangle'; *gua⁴ hao⁴* 'register'). Listeners were asked to click, with a mouse, on the target amongst the four possibilities upon hearing each auditory stimulus. Results suggest that No-Sandhi coarticulation was the easiest to recognize, as in an earlier time window, participants' fixation on the targets increased at the fastest rate among the three conditions. Between the two sandhi conditions, Near-Merger Sandhi was more difficult to process than No-Merger Sandhi, which was reflected in the overall less proportion of looks to target in the Near-Merger Sandhi condition within a later time window.

In the present study, we set out to investigate the recognition time course of different target stimuli when the targets are presented with different types of competitors. Specifically, target stimuli in the present study included two types of tone sandhi, i.e., Near-Merger Sandhi (i.e., T3T3) and No-Merger Sandhi (i.e., T1T1, T4T1), as well as those with No Sandhi coarticulation (i.e., T4T4, T3T2, T3T4) as a control. Three types of disyllabic competitors were included for these targets which all have segmental overlap with the target for the first syllable: 1) Toneme Competitor whose first tone shares the underlying toneme with that of the target but with distinctively different *f0* realization from the target, i.e., Toneme+, Contour-; 2) Contour Competitor whose first tone has a different toneme but surfaces with a similar *f0* contour with that of the target, i.e., Toneme-, Contour+; 3) Segmental Competitor whose first tone does not share toneme or contour with the target, i.e., Toneme-, Contour-. Taking the target *zhi³ fa³* ('fingering') for example, the Toneme Competitor is *zhi³ xiang¹* ('carton') whose first tone shares the T3 toneme with that of the target, the Contour Competitor is *zhi² cheng¹* ('professional title') in which the *f0* contour of the first tone is similar to that of the target, and the Segmental Competitor is *zhi⁴ hou⁴* ('lag') whose first tone is unrelated to that of the target. We are interested in how the proportion of looks to the target is affected by these three different competitor types, and how the effect of different competitors might interact with different target types.

5.3 Method

5.3.1 Participants

All participants in Chapter 4 participated in this experiment. The data from one subject who wore cosmetic contact lens were excluded. Data from another two subjects who were shortsighted without wearing glasses for correction were also excluded. All the remaining 31 subjects had normal or corrected-to-normal vision. All participants provided written informed consent.

5.3.2 Stimuli

The target stimuli in this experiment were the same as those in Chapter 4, including 18 near-merger sandhi targets (i.e., T3T3), 18 no-merger sandhi targets (i.e., T1T1, T4T1), and 18 no sandhi target (i.e., T4T4, T3T2, T3T4)[5].

For each target, a corresponding toneme competitor, a contour competitor and a segmental competitor were chosen (a within-item design). All competitors share segmental information with the targets at the first syllable, with manipulation only in tones. The first syllable of the toneme competitors shared the underlying toneme with that of the targets. The first tone of the contour competitors is the claimed output tone of that in the targets according to the sandhi rules claimed in the literature, whose *f0* contour resembles that of the targets. The segmental competitor does not share the underlying toneme or the surface *f0* realization with that of the targets, which are therefore unrelated to that of the targets. The second syllables of the targets and competitors are different in terms of both tone and segment across all conditions. Each target-competitor pair also has two distractors within the Visual World Paradigm. These distractors do not share any tone (toneme or surface contour) or segments with the target or the competitor. Table 5.1 illustrates the experimental design and sample stimuli.

The targets and competitors were further controlled to be closely matched in terms of lexical frequency based on Cai and Brysbaert (2010) as well as orthographic complexity (following Shen et al., 2013) as we presented Chinese characters instead of pictures. There was no significant difference between the targets and competitors for lexical frequency ($F(3)=0.108$, $p=0.955$) or visual complexity across conditions ($F(3)=1.526$, $p=0.209$). No participant reported difficulty of recognizing any character. All auditory stimuli were the same as those in Chapter 4.

Table 5.1 *Experimental design and sample stimuli.*

Near-Merger Sandhi			
Target	Toneme Competitor	Distractor 1	Distractor 2
	纸箱	电话	斑马
	(zhi^3 xiang^1)	(dian^4 hua^4)	(ban^1 ma^3)
	carton	*telephone*	*zebra*
	Contour Competitor	Distractor 1	Distractor 2
指法	职称	操场	机场
(zhi^3 fa^3)	(zhi^2 cheng^1)	(cao^1 chang^3)	(ji^1 chang^3)
fingering	*professional title*	*playground*	*airport*
	Segmental Competitor	Distractor 1	Distractor 2
	滞后	三角	挂号
	(zhi^4 hou^4)	(san^1 jiao^3)	(gua^4 hao^4)
	lag	*triangle*	*to register*

[5] No Sandhi stimuli included those claimed to undergo certain tone sandhi in impressionistic studies (e.g., Li & Liu, 1985; Chen, 2000; Wee et al. 2005) but not confirmed by empirical data in Chapter 3 and in Li & Chen (2016).

	No-Merger Sandhi		
Target	Toneme Competitor	Distractor 1	Distractor 2
	气球	樱花	香肠
	(qi^4 qiu^2)	(ying^1 hua^1)	(xiang^1 chang^2)
	balloon	*cherry blossom*	*sausage*
	Contour Competitor	Distractor 1	Distractor 2
器官	棋牌	钢琴	接力
(qi^4 guan^1)	(qi^2 pai^2)	(gang^1 qin^2)	(jie^1 li^4)
organ	chess and cards	*piano*	*rally*
	Segmental Competitor	Distractor 1	Distractor 2
	乞丐	餐巾	空气
	(qi^3 gai^4)	(can^1 jin^1)	(kong^1 qi^4)
	beggar	*napkin*	*air*
	No Sandhi		
Target	Toneme Competitor	Distractor 1	Distractor 2
	稳重	办公	记号
	(wen^3 zhong^4)	(ban^4 gong^1)	(ji^4 hao^4)
	steady	*to work*	*marker*
	Contour Competitor	Distractor 1	Distractor 2
吻合	温度	裤袜	外币
(wen^3 he^2)	(wen^1 du^4)	(ku^4 wa^4)	(wai^4 bi^4)
to match	*tempareture*	*leggings*	*foreign currency*
	Segmental Competitor	Distractor 1	Distractor 2
	问好	住址	袋鼠
	(wen^4 hao^3)	(zhu^4 zhi^3)	(dai^4 shu^3)
	to greet	*home address*	*kangaroo*

5.3.3 Procedure and eye movement data analysis

We followed the same experimental procedure and eye movement data analysis as that in Chapter 4. The eye movement data were also reported from the onset of the auditory stimulus to 1400ms post stimulus onset following Chapter 4. We employed the growth curve analyses procedure with the package *lme4* (Bates et al., 2014) in R (R Core Team, 2014) as in Chapter 4.

The critical independent factors included TARGET TYPE and COMPETITOR TYPE. We also included the control factors TARGET FREQUENCY, TARGET STROKE NO., TARGET STRUCTURE, TARGET BIGRAM INFO., COMPETITOR FREQUENCY, COMPETITOR STROKE NO., as well as COMPETITOR STRUCTURE. Our main interest was to see the possible effects of TARGET TYPE and COMPETITOR TYPE after having regressed out possible mainly stimulus-intrinsic factors as listed above. Multiple comparisons with Bonferroni adjustment were conducted with the function *glht* in package *multcomp* (Hothorn et al., 2008) whenever necessary.
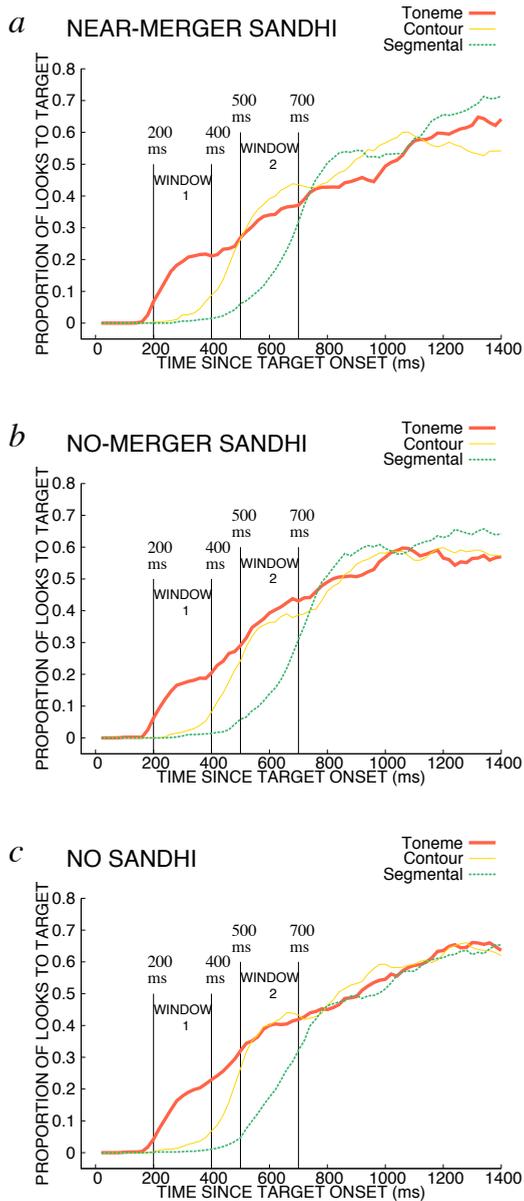
Figure 5.3 *Mean proportion of looks to target when targets of different types were presented with different competitors. Targets with Near-Merger Sandhi were plotted in a, No-Merger Sandhi targets in b, No Sandhi targets in c. Thick solid lines stand for the condition when presented with a Toneme Competitor, thin solid lines when presented with a Contour Competitor, thin dashed lines when presented with a Segmental Competitor. Aggregated across participants and items.*

5.4 Results

To investigate how sandhi tones are recognized, we included analyses for looks to both targets and competitors in this section.

5.4.1 Looks to the target

Figure 5.3 shows the mean proportion of looks to the target when each type of target was presented with different competitors (illustrated with different line types). X-axis stands for the time since auditory target onset, y-axis for the proportion of looks to target/competitor. Targets with Near-Merger Sandhi were plotted in *a*, No-Merger Sandhi targets in *b*, No Sandhi targets in *c*.

We can see from Figure 5.3 that for all target types, there are great differences in the proportion of looks to the target due to different competitors, especially from 200ms to 700ms after auditory stimuli onset. After 700ms, the differences between the three competitor conditions are much more reduced. We further zoomed into two particular time windows which show interesting eye movement patterns in each target type. The first window is from 200-400ms after auditory target onset. Given the 200ms delay to plan and execute an eye movement (Matin et al., 1993), 200-400ms relatively corresponds to 200ms into the auditory stimuli which maximally ensures no information of the second syllable is available during this time window, as the minimal duration of the first syllable is 220ms in our dataset. The second window is from 500ms to 700ms, following Li and Chen (2015) and data reported in Chapter 4 in order to make a direct comparison.

From visual inspection of the 200-400ms time window (WINDOW 1, Figures 5.3*a,b,c*), the proportion of looks to the target in the three competitor conditions seems to start increasing at different time points. Participants took the least time to initiate an eye movement to the target when there is a Toneme Competitor, i.e., around 200ms after the auditory stimulus onset. The proportion of looks to the target with the presentation of the other two types of competitors started to increase at much later time points, i.e., around 300ms for the Contour Competitor condition and after 400ms for the Baseline Competitor condition. This has resulted in a much greater overall mean (around 20%) and a much faster increase in the proportion of looks to the target in Toneme Competitor condition than in the other two competitor conditions within this time window. The Contour Competitor condition and Segmental Competitor condition show only slight differences in the proportion of looks to the target, both less than 10%. Similar patterns could be consistently observed across three target types.

A growth curve analysis was run for the comparison of the proportion of looks to the target among the three competitor conditions within the 200-400ms window across target types. The best-fit model contained the interaction of TIME (up to fourth-order) and COMPETITOR TYPE (three levels: TONEME COMPETITOR, CONTOUR COMPETITOR, SEGMENTAL COMPETITOR) in the fixed factor structure. TARGET TYPE did not significantly improve the model fit ($\chi^2(10)=12.10$, p=0.28, n.s.), which

suggested that different sandhi types in the target did not affect the proportion of looks patterns within the 200-400ms time window. Among all the controlled factors, only COMPETITOR STRUCTURE showed significant improvement of the model ($\chi^2(2)=27.62$, $p<0.001$), and thus was added to the model. SUBJECT was considered as a random factor over all time terms. Table 5.2 shows the estimation of the effect of competitor types.

Table 5.2 *Pairwise comparison of the effect of competitor types on the proportion of looks to target within the time window 200-400ms.*

	a. Contour vs. Toneme	b. Segmental vs. Toneme	c. Segmental vs. Contour
Intercept	β=-0.14 z=-51.75 p<0.001	β=-0.17 z=-64.25 p<0.001	β=-0.03 z=-11.87 p<0.001
Slope	β=-0.03 z=-3.69 p<0.01	β=-0.11 z=-12.41 p<0.001	β=-0.08 z=-8.66 p<0.001
Quadratic	β=0.07 z=8.20 p<0.001	β=0.04 z=4.76 p<0.001	β=-0.03 z=-3.42 p<0.01
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.

The statistic results confirmed the observations made from WINDOW 1 of Figure 5.3. As shown in Table 5.2, the proportion of looks to target in the Toneme Competitor condition had a significantly higher overall mean than that of the other two conditions (see significant results in the intercept of Contour vs. Toneme and Segmental vs. Toneme in Tables 5.2*a,b*). The Contour Competitor condition had a slightly but significantly higher overall mean than that of Segmental Competitor condition (see significant results in the intercept of Baseline vs. Contour in Table 5.2*c*). Furthermore, the Toneme Competitor condition showed a significantly faster proportion increase than the Contour Competitor condition, which was further faster than the Segmental Competitor condition, as reflected by the significant differences in the slope and quadratic terms in Tables 5.2*a,b,c*.

The second interesting time window is from 500ms to 700ms (WINDOW 2, Figures 5.3*a,b,c*). In this window, the proportion of looks to target in Toneme and Contour Competitor conditions keep increasing at a similar rate (both rising from 30% to around 40%). The Segmental Competitor condition, however, just starts to increase but at a relatively faster increasing rate (from less than 10% to around 30%).

Within this window, both COMPETITOR TYPE ($\chi^2(10)=2047.1$, $p<0.001$) and TARGET TYPE ($\chi^2(10)=30.25$, $p<0.001$) significantly improved the model fits. They also showed significant interactions ($\chi^2(20)=50.69$, $p<0.001$). To investigate the effect of competitor in each target type, we further subset the data to run separate analyses within each target type. Estimates of the effect of competitors are listed in Table 5.3.

Two observations can be made from Table 5.3. First, in all three target types, both the

Toneme Competitor condition and the Contour Competitor condition were significantly different from the Segmental Competitor condition in terms of significantly higher overall mean but slower increase in the proportion of looks to the target. This was reflected by the significant results in the intercept, slope as well as the quadratic of Segmental vs. Toneme and Segmental vs. Contour in Tables 5.3I*b,c*, II*b,c*, III*b,c*. Meanwhile, the differences between the Toneme and Contour Competitor conditions were relatively smaller. Only very subtle overall mean differences were observed (see significant results in the intercept only in Tables 5.3I*a*, II*a*, and non-significant results in Table 5.3III*a*).

Second, it is notable that we observed a significantly larger overall mean of the proportion of looks to target in the Contour Competitor condition than in Toneme Competitor condition for Near-Merger Sandhi targets. This could be seen from the significant results in the Contour vs. Toneme of Table 5.3I*a*.

Table 5.3 *Pairwise comparison of the effect of competitor types on the proportion of looks to target within the time window 500-700ms in different target types.*

	a. Contour vs. Toneme	*b.* Segmental vs. Toneme	*c.* Segmental vs. Contour
I. Near-Merger Sandhi			
Intercept	$\beta=0.04$ $z=5.07$ $p<0.001$	$\beta=-0.17,$ $z=-18.73$ $p<0.001$	$\beta=-0.21$ $z=-23.35$ $p<0.001$
Slope	n.s.	$\beta=0.21$ $z=7.47$ $p<0.001$	$\beta=0.17$ $z=5.87$ $p<0.001$
Quadratic	n.s.	n.s.	$\beta=0.11$ $z=3.71$ $p<0.01$
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.
II. No-Merger Sandhi			
Intercept	$\beta=-0.06$ $z=-6.14$ $p<0.001$	$\beta=-0.21$ $z=-24.79$ $p<0.001$	$\beta=-0.16$ $z=-17.02$ $p<0.001$
Slope	n.s.	$\beta=0.16$ $z=5.85$ $p<0.001$	$\beta=0.20$ $z=7.05$ $p<0.001$
Quadratic	n.s.	$\beta=0.09$ $z=3.19$ $p<0.05$	$\beta=0.10$ $z=3.51$ $p<0.01$
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.

III. No Sandhi		
	a. Contour vs. Toneme	*b.* Segmental vs. Toneme	*c.* Segmental vs. Contour
Intercept	n.s.	β=-0.20 z=-21.86 p<0.001	β=-0.21 z=-22.95 p<0.001
Slope	n.s	β=0.21 z=7.17 p<0.001	β=0.18 z=6.25 p<0.001
Quadratic	n.s.	n.s.	n.s.
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.

5.4.2 Looks to the competitor

To observe where participants were looking if the target is not looked at, this section further analyzes the eye movement patterns over competitors. Looks to competitor were analyzed in a similar fashion to that for looks to target. Figure 5.4 shows the mean proportion of looks to the competitor when each type of target was presented with different competitors (illustrated with different line types). X-axis stands for the time since auditory target onset, y-axis for the proportion of looks to target/competitor. Targets with Near-Merger Sandhi were plotted in *a*, No-Merger Sandhi targets in *b*, No Sandhi targets in *c*. Two observations can be made.

First, it can be seen from WINDOW 1 in Figures 5.4a,b,c that, participants directed more looks to the Contour Competitor and the Segmental Competitor (both above 20%) than to the Toneme Competitor (less than 20%) at the beginning of the auditory stimuli (200-400ms). Meanwhile, the Contour Competitor and the Segmental Competitor are relatively similar to each other during this time window. This pattern was consistently observed across three target types, which was confirmed by the statistic results that TARGET TYPE did significantly improve the model fit of the data. As shown in Table 5.4, results of the best-fit model confirmed this with significant results in the proportion of looks between the Contour Competitor condition vs. the Toneme Competitor condition (see significant results in the intercept in Figure 5.4*a*) and Segmental Competitor vs. Toneme Competitor condition (see significant results in the intercept and slope in Figure 5.4*b*). Contour Competitor condition and Segmental Competitor condition did not differ significantly from each other.
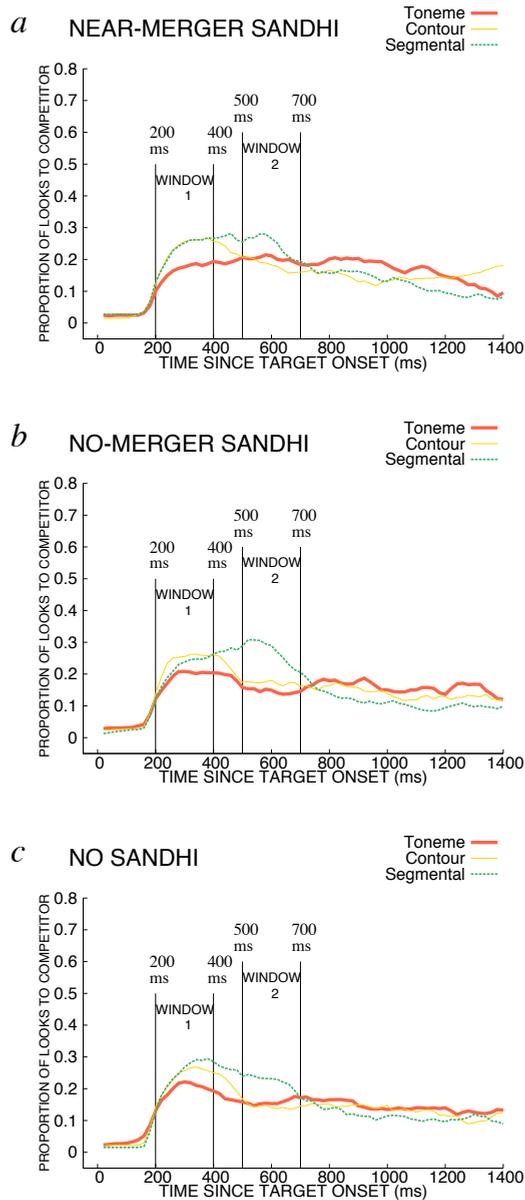
Figure 5.4 *Mean proportion of looks to competitor when targets of different types were presented with different competitors. Targets with Near-Merger Sandhi were plotted in a, No-Merger Sandhi targets in b, No Sandhi targets in c. Thick solid lines stand for the condition when presented with a Toneme Competitor, thin solid lines when presented with a Contour Competitor, thin dashed lines when presented with a Segmental Competitor. Aggregated across participants and items.*

Table 5.4 *Pairwise comparison of the effect of competitor types on the proportion of looks to competitor within the time window 200-400ms.*

	a. Contour vs. Toneme	*b.* Segmental vs. Toneme	*c.* Segmental vs. Contour
Intercept	β=0.06 z=11.92 p<0.001	β=0.05 z=11.30 p<0.001	n.s.
Slope	n.s.	β=0.06 z=3.94 p<0.01	n.s.
Quadratic	n.s.	n.s.	n.s.
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.

Second, within the 500-700ms window (WINDOW 2, Figures 5.4*a,b,c*), the participants looked at the Segmental Competitor (above 20%) more than at the Toneme and Contour Competitors (both less than 20%). This was reflected by the significant results in both the intercept and the slope of Segmental vs Toneme and Segmental vs. Contour in Figures 5.5Ib,c, IIb,c, IIIb,c. The Toneme Competitor condition and the Contour Competitor condition further differ from each other according to different target types. For both No-Merger Sandhi and No Sandhi targets, participants' looks to the Toneme and Contour Competitors clearly drop compared to that within the previous time window (WINDOW 1, Figures 5.4b,c). While for Near-Merger Sandhi targets, the proportion of looks to the competitor in the Toneme Competitor condition remains as that in the previous time window (WINDOW 1, Figure 5.4a). This was confirmed by the significantly higher overall mean in the looks to the Toneme Competitor than Contour Competitor (see significant results in the intercept of Contour vs. Toneme in Table 5.5I*a*).

Table 5.5 *Pairwise comparison of the effect of competitor types on the proportion of looks to competitor within the time window 500-700ms in different target types.*

	I. Near-Merger Sandhi		
	a. Contour vs. Toneme	*b.* Segmental vs. Toneme	*c.* Segmental vs. Contour
Intercept	β=-0.02 z=3.33 p<0.05	β=0.03 z=3.77 p<0.01	β=0.06 z=6.83 p<0.001
Slope	n.s.	β=-0.08 z=-3.21 p<0.05	n.s.
Quadratic	n.s.	n.s.	n.s.
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.

	II. No-Merger Sandhi		
	a. Contour vs. Toneme	*b.* Segmental vs. Toneme	*c.* Segmental vs. Contour
Intercept	β=0.03 z=3.63 p<0.01	β=0.12 z=15.58 p<0.001	β=0.09 z=11.18 p<0.001
Slope	n.s.	β=-0.13 z=-5.30 p<0.001	β=-0.12 z=-4.99 p<0.001
Quadratic	n.s.	n.s.	n.s.
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.
	III. No Sandhi		
	a. Contour vs. Toneme	*b.* Segmental vs. Toneme	*c.* Segmental vs. Contour
Intercept	β=-0.02 z=-3.423 p<0.01	β=0.05 z=7.16 p<0.001	β=0.08 z=10.54 p<0.001
Slope	n.s	β=-0.11 z=-4.57 p<0.001	β=-0.07 z=-3.06 p<0.05
Quadratic	n.s.	n.s.	n.s.
Cubic	n.s.	n.s.	n.s.
Quartic	n.s.	n.s.	n.s.

5.5 Discussion & conclusion

This study taps into the online recognition of tonal sandhi variants in Tianjin Mandarin via the Visual World Paradigm. We examined the participants' looks to the target and competitor when they were presented with three types of competitors upon hearing two types of tone sandhi (i.e., Near-Merger Sandhi and No-Merger Sandhi) as well as stimuli with tonal coarticulation and no sandhi involved (i.e., No Sandhi) as a baseline for comparison. For each target stimulus, we included a Toneme Competitor which has toneme overlap with the target, a Contour Competitor whose *f0* contour resembles that of the target but of a different toneme, as well as a Segmental Competitor with an unrelated tone to that of the target. Results show that in all target types, participants were sensitive to the types of competitors together presented with the targets, which was reflected in the different eye movement patterns when participants were presented with different types of competitors.

First, regardless of the target types, there is a Toneme Competitor benefit across all target types within an earlier time window (200-400ms). During this stage, participants directed significantly more looks to the target (WINDOW 1, Figure 5.3) and less looks to the competitor (WINDOW 1, Figure 5.4) in the Toneme Competitor condition. This was reflected by significantly higher overall mean and significantly faster increase of the proportion of looks to target as well as the significantly lower overall mean of the

proportion of looks to competitor in the Toneme Competitor condition. The other two competitor conditions (i.e., Contour Competitor condition and Contour Competitor condition) exhibit relatively similar eye movement patterns (reflected in the gaze pattern for both target and competitor). This has thus suggested that a Contour Competitor (Toneme-, Contour+) plays a similar role in the processing of the target to that of the Segmental Competitor (Toneme-, Contour-) at the first 200ms of the target stimuli, both of which show stronger activation than the Toneme Competitor (Toneme+, Contour-).

Second, in the later time window (500-700ms), the activation of targets in both Toneme and Contour Competitor conditions continues to increase, while the activation of target in the Segmental Competitor condition just starts to increase (WINDOW 2, Figure 5.3). This was confirmed by the significantly lower but faster increase in proportion of looks to the target in the Segmental Competitor condition than the other two competitor conditions. The data of looks to competitor also show that the participants directed more looks to the Segmental Competitor more than Toneme and Contour Competitors, suggesting a stronger competition effect between targets and Segmental Competitors.

It has been repeatedly observed in previous studies on spoken word recognition within the Visual World Paradigm that, segmental overlap between the target and the competitor gives rise to competition effect in listeners' recognition of the spoken stimuli (see e.g., Tanenhaus et al., 1995; Allopenna et al., 1998; Dahan et al., 2001a, 2001b; McMurray et al., 2002; Beddor et al., 2013). Based on recent evidence that segments and lexical tones play a similar role in constraining word recognition (Malins & Joanisse, 2011), we predicted that tonal overlap, or at least partial tonal overlap, should induce further competition effect compared to the condition with no tonal overlap. Therefore, we would have expected more competition between target and competitor in the Toneme Competitor condition and the Contour Competitor condition than in the Segmental Competitor condition. Instead, what we have observed was that within the early window (i.e., 200-400ms), there was a delayed activation of the target in the Contour Competitor and the Segmental Competitor conditions, suggesting a different processing trajectory of the target words in the face of their Toneme Competitor (i.e., earlier and greater activation) from that against a Segmental or Contour competitor. Interestingly, regardless of how far the surface $f0$ realization of the target tone deviates from the surface $f0$ realization of its Toneme Competitor (i.e., across the three target types), the pattern of early activation of target with Toneme Competitor remains. This suggests that at an earlier stage of processing, listeners are processing the tonal information, despite their contextual variation, as their underlying lexical tones. Thus even for targets with Near-Merger tone sandhi, their Contour Competitors, despite the similarity in tonal pitch contours, exert a similar competition effect just as the Segmental Competitors (i.e., segmental sharing only).

A similar delayed activation of the correct target has been observed in Malins and Joanisse (2010), where participants did not look at the target at all during the initial 200ms into the first syllable when there is a Segmental Competitor. Moreover, they also observed that compared to the baseline condition, participants initially directed more looks to the target when target and competitor have tonal overlap in addition to some segmental overlap with the targets. This is then in line with our earlier activation of target with

Toneme Competitor (although we had lexical tonal overlap as well as complete segmental overalp).

At a later time window of processing (500-700ms), however, we observed a different pattern of looks as a function of both target types and competitor types. Specifically, two observations are to be noted. First, across target types, there was a continued delay in the activation of the targets with Segmental Competitor. The activation trajectory for the target in the Contour Competitor condition shows a more comparable pattern to that in the Toneme Competitor condtion, which is a reversed pattern from the earlier window. Despite their similarity, there were subtle yet consistent differences across the target types. For the Near-Merger targets, where the *f0* contours of the target and the Contour Competitor resemble to such an extent that they are almost ambiguous, the activation of the target in the Contour Competitor condition reached a higher level than that in the Toneme Competitor condition. In the No-Merger Sandhi and No Sandhi target conditions, where the *f0* contours of the target and the competitor are clearly distinctive from each other, the Contour Competitor enhanced the activation of the target, compared to the Segmental Competitor, but to a much smaller extent than the Toneme Competitor.

Overall, our data show that the Contour Competitors (Toneme-, Contour+) play a comparable role to the Segmental Competitors (Toneme-, Contour-) during the initial activation of the target. Therefore, despite the claims in the literature that for Near-Merger tone sandhi, one lexical tone undergoes sandhi change and is realized as another lexical tone, our data suggest that upon hearing the target word, which has a similar *f0* contour as the Contour Competitor, it is the underlying lexical tone of the target word that is concurrently activated together with the segments. This explains the different activation patterns of the target word when there is a Toneme Competitor versus when there is a Contour or Segmental Competitor. This finding then lends support to the proposal in the literature that tonal variation due to tone sandhi is better viewed as allophonic variation of the underlying lexical tones and are stored together with the lexical toneme as allotones in the mental lexicon (Chen et al., 2011; Nixon et al., 2015). This is also in line with recent results of an ERP study on lexical tonal processing in Standard Chinese (X. Li & Y. Chen, in press). With an oddball paradigm, this study investigated the effect of allophonic variation on the mental representation and neural processing of lexical tones produced in isolation. All stimuli share segments /ma/ but with different lexical tones: high-level T1, rising T2, and dipping T3. The ERP results show that among the four oddball conditions (T1/T3, T3/T1, T2/T3, T3/T2; standard/deviant), the T1-T3 pair elicited symmetrical mismatch negativity effects (MMN), while the T2-T3 pair only showed asymmetrical MMN effects where there were significantly greater and earlier MMN effects in the T2/T3 condition than that in the reversed T3/T2 condition. This thus suggests the co-activation of representations of both non-sandhi T3 and T3$_{sandhi}$ even when listeners only hear a non-sandhi T3 in isolation.

Despite that tonal contour similarity (between the target and the competitor) did not seem to exert an effect on the activation of target tone within the earlier time window, it is important to note that there was a contour similarity effect at a later processing stage. Within the time window of 500-700ms, the tonal contour has been completed or is near

completion. By then, contour similarity of the Contour Competitor seemed to trigger similar patterns of target activation as in the Toneme Competitor condition, evident across the target types. This is in sharp contrast to the delayed activation of target with Segmental Competitor. While more research is needed to understand the process, one possibility to explain this differences is that the activation of the target is facilitated either with underlying lexical tonal sharing or surface contour similarity. While in the Segmental Competitor condition, the delayed activation of the targets may be due to the lack of such facilitation effect. Furthermore, listeners seemed to be sensitive to the subtle differences in the contour similarity. When the target and competitor show great contour overlap as in the Near-Merger Sandhi case, the contour overlap seems to be more likely to facilitate the target activation, compared to when the target and the Contour Competitor share less contour similarity (i.e., in the No-Merger Sandhi and No Sandhi target conditions).

Our results are thus different from conclusions made in traditional tone sandhi perception studies which usually employ meta-linguistic judgement tasks. These studies have mainly investigated the perception of sandhi tones in Standard Chinese which only has one Near-Merger sandhi (i.e., the change of a 3rd Low tone before another Low; known as Third-Tone/Low-Tone Sandhi), and have argued for complete neutralization of the Sandhi Low tone with the lexical Rising tone due to the lack of reliable perception between the two tones (e.g., Wang & Li, 1967; Speer et al., 1989; Chen, 2013; Chen et al., 2015). Given the multiple layers of tone sandhi in Tianjin Mandarin, we show, with eye tracking data which reveal the dynamic processing of spoken words, that tone sandhi does not have to be the categorical change from one lexical tone to another. Regardless of how the sandhi-derived tones resemble another lexical tone within the tonal system, native listeners consistently process the claimed output tones (in the literature) as an unrelated tone from an early processing stage and was able to fixate upon the correct target word despite great similarity between the sandhi variant and the lexical tone competitor which is claimed in the literature to be the sandhi output tone. In other words, within the Visual World Paradigm, which is much more sensitive to the dynamic activation processes given a speech input, we have shown that listeners are able to differentiate a sandhi tonal variant from its contour competitor lexical tone well above chance level. This thus echoes well with recent experimental data showing subtle but consistent acoustic differences (see e.g., Zee, 1980; Peng, 1996; Yuan & Chen, 2014; Zhang & Liu, 2011; Li & Chen, 2016) as well as different tonal encoding processes in online speech production between sandhi-derived tonal variants and the lexical tones claimed to be their sandhi output tones in the literature (Chen et al., 2011; Nixon et al., 2015; Zhang et al., 2014). This thereby argues firmly against the view that tone sandhi in these Mandarin dialects involves the categorical change of one lexical tone into another.