



Universiteit
Leiden
The Netherlands

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

Li, Qian

Citation

Li, Q. (2016, February 10). *The production and perception of tonal variation -: evidence from Tianjin Mandarin*. Retrieved from <https://hdl.handle.net/1887/37609>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/37609>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



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 1 INTRODUCTION

As an important means of our daily communication, spoken language is notoriously rich in its variability. Therefore a word's pronunciation is often different in connected speech from when it is produced in isolation. For example, in the phrase “green beans”, the [n] in the word “green” may assimilate its place of articulation to match that of the following consonant, making [n] more similar to the following [b], resulting the word “green” to sound more like “greem” in this context. In the last decades, such segmental variation in connected speech has received much attention in both fields of linguistics and psycholinguistics. Various theories and models have been proposed to account for the underlying mechanisms of how speakers produce and perceive the variation (see Ernestus, 2012 for a review). Variation of suprasegmental properties, however, is much less studied, probably due to the fact that suprasegmental properties such as prosody is considered to play a relatively minor role in distinguishing lexical meanings in Indo-European non-tonal languages such as English (e.g., Cutler, 1986). This has greatly limited the existing speech production and perception theories within a segment-dominant frame in general.

It is known that of all languages in the world, the majority of them are tonal languages, in which pitch patterns over syllables distinguish lexical meanings (see reviews in e.g., Fromkin, 1978; Yip, 2002; Zsiga, 2012). In some of these tonal languages (i.e., the majority of the African languages and certain American-Indian languages), tones are known to correlate with the morphology and syntax of the language; while in others, tones are exclusively used at the lexical level, which can be found in almost all languages in the Sino-Tibetan family as well as many other languages in the Southeast Asia (Wang, 1967). Emerging evidence has shown that in tonal languages (e.g., Chinese), suprasegmental features such as lexical tones play as an important role in constraining lexical meanings as segments do (e.g., Schirmer, Tang, Penney, Gunter, & Chen, 2005; Lee, 2007; Liu & Samuel, 2007). Understanding variation of lexical tones, therefore, is in great need for the construction of more powerful theories of speech production and comprehension.

One well-known example of a lexical tone system is the four full lexical tones in Standard Chinese. When produced in isolation, Tone 1 (T1) is a high level tone as in *ma*¹ (妈 ‘mother’), Tone 2 (T2) a rising tone as in *ma*² (麻 ‘hemp’), Tone 3 (T3) a dipping tone as in *ma*³ (马 ‘horse’), and Tone 4 (T4) is a falling tone as in *ma*⁴ (骂 ‘to scold’). These different pitch patterns (i.e., high level, rising, dipping, falling) are found to primarily correlate with how the rate of vocal cords vibration (i.e., *f*0) changes over time (e.g., Chao, 1920). Figure 1.1 illustrates the mean *f*0 of four lexical full tones in Standard Chinese, based on six tokens of each tone across eight speakers (Xu, 1997). As can be seen in Figure 1.1, the *f*0 contour of T1 stays at the upper *f*0 range throughout the whole syllable; T2 typically shows a rising *f*0 pattern which rises from the mid range to the upper limit; T3 is realized with a concave *f*0 shape which falls to the lower end for the first half of the *f*0 and then rises back to the mid range; for T4, the *f*0 falls from the high end of the *f*0 range to the lower end.

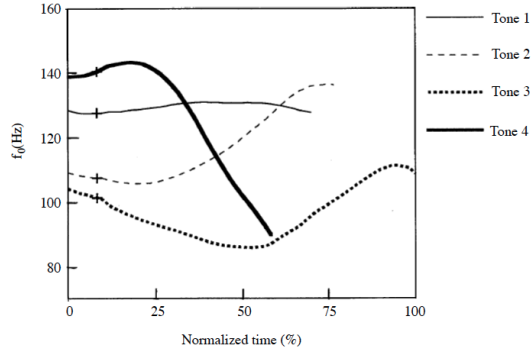


Figure 1.1 Mean f_0 contours of the four Mandarin tones in the syllable /ma/ produced in isolation (Xu, 1997). Time normalized with duration proportional to the mean duration of T3. Averaged across six tokens by eight speakers.

From the perspective of listeners, the f_0 differences provide important perceptual cues that can be utilized to identify tones (Gandour, 1978; but see e.g., Andruski, 2006; Brunelle, 2009b for tones contrasted in voice quality). These cues include the overall height of the f_0 contour (e.g., Lin & Repp, 1989; Shen, Lin, & Yan, 1993; Francis, Ciocca, & Ng, 2003; Shen et al., 2013), the direction of f_0 change (e.g., Lin & Repp, 1989; Fox & Qi, 1990; Shen & Lin, 1991; Moore & Jongman, 1997; Francis et al., 2003), and the timing of f_0 turning point (e.g., Shen & Lin, 1991; Shen et al., 1993; Moore & Jongman, 1997). For example, as introduced earlier, T2 in Standard Chinese is typically a rising tone. However, due to physiological limitations of the articulators, the f_0 onset of T2 is slightly falling prior to its significant rising f_0 contour. As a result, T2 is very often realized with a concave f_0 pattern similar to that of T3. However, a closer look suggests that the f_0 turning points of the two concave f_0 contours are at different locations, i.e., T2 has an early turning point (if any), while the turning point for T3 is relatively late, as can also be seen in Figure 1.1. Shen and Lin (1991) created a continuum ranging from T2 to T3 in Standard Chinese by manipulating the location of the f_0 turning point. Native listeners were asked to label the stimuli as either T2 or T3 in a forced choice paradigm. Results show that the location of the f_0 turning point, among others, plays an important role in distinguishing T2 vs. T3 in Beijing Mandarin. Stimuli with earlier f_0 turning points were more likely to be labelled as T2 than those with later turning points.

Having known how lexical tones are realized and perceived in isolation is important for our understanding of lexical tones; however, it is far from the whole story. As a matter of fact, in daily speech communication, it is very rare that tones only occur in isolation. Rather, they are more often produced and perceived in contexts, where their f_0 realization usually exhibits extensive variability.

At a local level, the f_0 realization of lexical tones is known to be varied greatly by its local tonal contexts via two different processes. One such local process is known as tone sandhi, which is traditionally defined as the phonological change of lexical tones (see surveys in Wang, 1967; Chen, 2000; also see Zhang, 2010 for a review). An example in Standard Chinese: when two dipping tones (T3) are combined, the first T3 is typically

realized with a rising f_0 , resembling that of the rising T2. Although previous impressionistic studies have claimed that the T3 sandhi in Standard Chinese involves the complete tonal change from T3 to T2 (e.g., Chen, 2000; Yip, 2002), this view is challenged by findings from tone production studies, which demonstrate subtle but quite consistent f_0 differences between T3_{sandhi} and the lexical T2 (e.g., Zee, 1980; Peng, 1996; Yuan & Chen, 2014). Furthermore, T3_{sandhi} and the lexical T2 have also been shown to be processed differently during speech encoding (Chen, Shen, & Schiller, 2011; Zhang, Xia, & Peng, 2014). If complete neutralization between T3_{sandhi} and the lexical T2 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.

Another local contextual process known to introduce tonal variability is tonal coarticulation, which traditionally refers to the phonetic f_0 adjustments to neighboring tonal contexts (e.g., Wu, 1985; Xu, 1997; also see reviews in Xu, 2001; Chen, 2012). Tonal coarticulation has been reported in many East Asian contour-tone languages, such as Thai (e.g., Palmer, 1969; Abramson, 1979; Gandour et al., 1993, 1996; Gandour, Potisuk, & Dechongkit, 1994; Potisuk, Gandour, & Harper, 1997; Zsiga & Nitisaroj, 2007), Vietnamese (e.g., Han & Kim, 1974; Brunelle, 2009a), Standard Chinese (e.g., Shen, 1990; Xu, 1994; Xu, 1997; Kochanski, Shih, & Jing, 2003), Taiwanese (e.g., Peng, 1997; Wang, 2002), and Malaysian Hokkien (e.g., Chang & Hsieh, 2012). Taking an example in Standard Chinese, Figure 1.2 illustrates the mean f_0 realization of the high-level T1 following (Figure 1.2a) or preceding different tones (Figure 1.2b), based on six tokens produced by eight speakers (Xu, 1997). It is clear in Figure 1.2a that, due to different preceding tones, the second high-level T1 is realized with very different f_0 contours in these contexts. The f_0 contours of T1 thus deviate in different magnitudes from the canonical level tonal shape, especially for the first half of the second T1. When T1 is followed by different tones as shown in Figure 1.2b, its f_0 also exhibits observable variation, but in a much smaller magnitude (Xu, 1997).

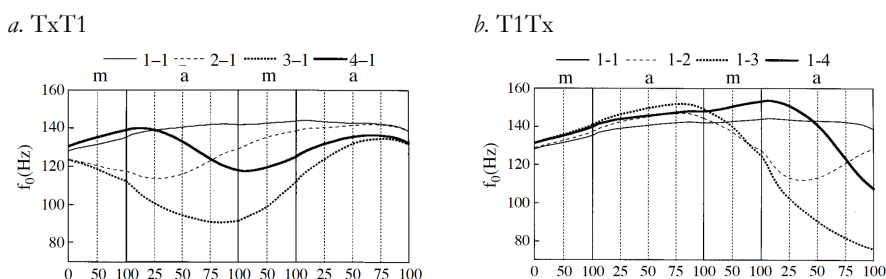


Figure 1.2 The effect of preceding (a) or following tones (b) on the f_0 realization of T1 in /mama/ sequences in Standard Mandarin (Xu, 1997). In (a), the tone in the first syllable is varied from T1 to T4, while the second tone is held constant as T1. In (b), the first tone is held constant, while the second tone is varied from T1 to T4. Averaged across six tokens by eight speakers.

From the global point of view, the f_0 realization of a lexical tone is further conditioned by the overall prosody of the whole utterance where the tone-bearing lexical

item is produced (see reviews in Xu, 2001 and Chen, 2012). Such global influences can come from the grouping of the prosodic units within the utterance (e.g., Shih, 1997; Yang & Wang, 2002; Pan & Tai, 2006; Scholz & Chen, 2014b), the prosodic prominence due to certain information structures (e.g., Xu, 1999; Chen & Braun, 2006; Chen & Gussenhoven, 2008; Chen, 2010; Wang & Xu, 2011), or even sentence types such as declaratives or interrogatives (e.g., Shen, 1991; Liu & Xu, 2005). For example, the *f0* realization of lexical tones in Standard Chinese is found to be subject to different focus statuses of the tone-bearing unit, where tones under focus are usually realized with enhanced distinctiveness of their characteristic *f0* contours compared to those not under focus (e.g., Chen & Braun, 2006; Chen & Gussenhoven, 2008; Chen, 2010).

In addition to variation of the lexical full tones, there exists a special tonal variability phenomenon called “neutral tone” (known in Chinese as 轻声 *qīngshēng*¹ ‘light tone’) (Chao, 1920). As illustrated with examples below from Standard Chinese, these syllables are typically grammatical morphemes as in *a*, the unstressed final syllable within a disyllabic lexical item as in *b*, or the final syllable of a reduplicated form as in *c*.

<i>a</i>	我的	wo ³ de	<i>possessive marker</i>
<i>b</i>	玻璃	bo ¹ li	<i>lexical item</i>
<i>c</i>	看看	kan ⁴ kan	<i>reduplicated form</i>

As can be seen from the examples, the neutral tone syllables are never produced independently, but always follow a full tone syllable within a disyllabic domain in Standard Chinese.

Compared to the four full lexical tones, the neutral tone syllables are acoustically reduced and considered to be “unstressed” (see reviews in Liang, 2008, Chen, 2015 and references therein). For example in Standard Chinese, neutral tone syllables are found to be substantially shorter than the full lexical tones (e.g., Lin & Yan, 1980; Shih, 1987). In terms of the *f0* realization, these syllables do not surface with any of the lexical *f0* patterns, but rather are realized with greatly varied *f0* depending on the tone of the preceding full lexical tone syllable (e.g., Shih, 1987; Chen & Xu, 2006 and references therein). Due to the variability, neutral tone has been regarded as “toneless” or “targetless” in most impressionistic studies, and the varied *f0* realization is attributed to the tonal spreading from the preceding lexical tone (e.g., Yip, 1980) or the phonetic interpolation of neighboring lexical tone targets (e.g., Shih, 1987; van Santen, Shih, & Möbius, 1997). However, in a recent study of neutral tone in Standard Chinese, Chen and Xu (2006) investigated the *f0* realization of long stretch of neutral tones. By increasing the number of neutral tone syllables from one to three, the authors show that the *f0* variability of the neutral tone gradually decreases over time and eventually approaches a stable mid-low *f0* range by the end of the third neutral tone syllable. This has thus suggested that neutral tone has its own mid-low tonal target that is independent from its tonal contexts.

It is to be noted that, despite the large repertoire of dialects spoken in China, only Standard Chinese has been studied in depth. Much less well-controlled data have been collected to understand how lexical tones are realized and varied in dialects. For most of the works in Chinese dialects, lexical tone descriptions are done largely within the traditional phonological framework. Variation of tones is mostly investigated on an

impressionistic and introspective basis. Detailed variation effects that are not easy to detect with unaided ears are very often ignored. Whenever certain variation is detected, it is very often described based on subjective impression and exclusively attributed to categorical changes (e.g., Chen, 2000). This not only misinforms the theoretical accounts of the variation phenomenon itself, but also greatly restricts our understanding of how listeners produce and recognize tonal variation in general.

To fill this research gap, this dissertation sets out to investigate the tonal variability in Tianjin Mandarin with a series of well-controlled experiments. Tianjin Mandarin serves as a good test case as it is well-known for its tonal sandhi patterns over disyllabic tonal sequences and the iterative and conflicting sandhi applications over trisyllabic tonal sequences, known as the “paradox” in Tianjin tone sandhi (e.g., Chen, 2000). In addition, Wang (2002) reports a special rising f_0 realization of neutral tone for Tianjin Mandarin which challenges the current understanding of neutral tone realization based on evidence from Standard Chinese. It is important to note that although tonal variation in Tianjin Mandarin seems to have been extensively studied in the literature, most previous studies are based on impressionistic data (but see Zhang & Liu, 2011 for an experimental study of tonal coarticulation and tone sandhi in Tianjin Mandarin). Further empirical evidence is thus in great need to shed more light on the underlying mechanisms of these phenomena. Last but not least, an advantage for investigating Tianjin Mandarin is that it is similar with Standard Chinese especially in the segmental aspects as will be shown in Chapter 2, making it easy to compare between the two Mandarin varieties.

Several interesting issues are to be highlighted in this dissertation concerning the production and perception of lexical tones in connected speech. The rest of this chapter briefly introduces these issues and how they have been addressed experimentally in this dissertation.

1.1 Tone sandhi vs. tonal coarticulation

As we have introduced earlier, both tone sandhi and tonal coarticulation give rise to variation in the f_0 realization of a lexical tone, so that the resulting f_0 contour may deviate with varying degrees from the canonical f_0 shape of a lexical tone when produced in isolation. One noteworthy issue is that, although tone sandhi and tonal coarticulation are recognized as two different processes by definition, little is known so far on how to categorize different f_0 variability as being due to tone sandhi or tonal coarticulation.

In the only attempt made so far concerning this issue, Shen (1992) proposes three criteria between the two processes: 1) tone sandhi is constrained by language-dependent morpho-phonemics, while tonal coarticulation is language-independent; 2) tone sandhi could have an assimilatory or dissimilatory nature while tonal coarticulation is only assimilatory; 3) tone sandhi involves the lexical tone change from one to another which is not observed for tonal coarticulation. As will be discussed in Chapter 3, none of these criteria successfully distinguish the two types of tonal variation (also see Chen, 2000 for a criticism). Chen (2000) thus argues that there is “no essential difference” between tone sandhi vs. tonal coarticulation, but defines tone sandhi as the perceptible tonal variation to

“trained but unaided ears”. Although this definition is shared among almost all studies on tone sandhi in the recent decades, it does not provide us with any real insight into different types of contextual tonal variation, as the impression of individual researchers on what constitutes tone sandhi might differ, which further leads to different theoretical accounts of the tonal alternation phenomenon.

It is notable that tonal sequences that are claimed to undergo tone sandhi by impressionistic studies are very often analyzed separately from tonal coarticulation, leaving no direct comparison of the two types of contextual variation thus far (see e.g., Xu, 1997; Zhang & Liu, 2011 for anticipatory tonal coarticulation analyses excluding tone sandhi sequences). It is therefore yet to be investigated with well-designed experimental data how tone sandhi and tonal coarticulation might individually contribute to the *f*₀ variability of lexical tones, without which one is unable to objectively attribute different types of *f*₀ variability to different sources.

Chapter 3 aims to fully understand the various patterns of *f*₀ deviation due to these two different contextual tonal variation processes. To this end, Chapter 3 investigates all possible combinations of the four lexical tones in Tianjin Mandarin. In particular, this chapter zooms into the effect of the following tones (over the second syllable) on the *f*₀ realization of the preceding tones (over the first syllable). This is a comparable context to examine both anticipatory tonal coarticulation and tone sandhi as Tianjin Mandarin shows right-dominant tone sandhi patterns. With the technique of Growth Curve Analysis (GCA) (Mirman, 2014), this chapter objectively quantifies whether and how *f*₀ contours of a lexical tone differ from each other as a function of the following lexical tones. GCA proves to be a powerful method for dealing with the time-varying nature of *f*₀ contours, which shows a great advantage over traditional statistical analyses of *f*₀ contours.

Results show that, among the four commonly accepted tone sandhi patterns over disyllabic tonal sequences proposed in the literature (e.g., Li & Liu, 1985; Chen, 2000; Ma, 2005; but see Wee, Yan, & Chen, 2005 for two more disyllabic sandhi sequences), only three have been confirmed. This chapter also yields an interesting anticipatory coarticulatory raising effect of the lexical low-falling tone which has never been reported in the literature. In addition, the confirmed disyllabic sandhi patterns can not be consistently observed within trisyllabic sequences, as claimed in the literature, suggesting a much simpler tone sandhi system of Tianjin Mandarin. Furthermore, by comparing the acoustic realization of the sandhi sequences vs. that of the claimed target patterns according to the literature, Chapter 3 shows that no tonal neutralization can be confirmed over any of the disyllabic or trisyllabic sandhi processes. Rather, based on how sandhi-derived tones resemble their claimed output tones, Chapter 3 further differentiates near-merger sandhi vs. no-merger sandhi, where a near-merger sandhi tone is realized with only subtle *f*₀ differences from that of the claimed output tone while a no-merger sandhi tone and its claimed output tone show clearer *f*₀ differences.

1.2 Effect of tonal variability on tone perception

Given the extensive contextual f_0 variation of lexical tones in connected speech, one would further ask how the lexical tones produced in connected speech are perceived. As introduced earlier, listeners are found to primarily rely on the f_0 information during tonal perception. Multiple f_0 cues have been discovered to play crucial roles in the perception of canonical lexical tones that produced in isolation. However, this is insufficient to understand how lexical tones are perceived in connected speech for the following two reasons.

First, the f_0 perceptual cues observed for tones produced in isolation have mainly indicated the most significant f_0 characteristics of the canonical tonal contours. For example in Standard Chinese, a high-level f_0 contour signals a T1 produced in isolation. However, in connected speech, T1 does not always surface with a level f_0 contour. As shown in Figure 1.2a, when T1 in Standard Chinese is preceded by four different lexical tones, only the one following T1 is realized with a high-level f_0 contour which is similar to that of the canonical T1 produced in isolation; T1 following T2, T3 or T4 is realized with a salient rising f_0 contour with different rising slopes. Then the question arises how listeners cope with the f_0 variation and identify the tone as T1 in such situation?

Second, it is known that the perception of lexical tones in connected speech is different from the perception of tones produced in isolation, because tonal contexts can greatly influence the identification of the embedded lexical tones (e.g., Lin & Wang, 1985; Fox & Qi, 1990; Xu, 1994; Moore & Jongman, 1997; Francis et al., 2003). For example, in Lin and Wang (1985), it is found that a high-level f_0 contour, which is T1 in Standard Chinese, can be perceived as a rising tone (T2) when it is followed by another tone whose f_0 onset is higher than the high-level tone. It is therefore highly doubtful that to what extent the conclusions made based on lexical tones produced in isolation can be adaptable to tones in connected speech.

Among studies on the perception of lexical tones in connected speech, the main topics discussed in the literature concern 1) how the perception of lexical tones is influenced by the tonal contexts in which they are embedded (e.g., Lin & Wang, 1985; Fox & Qi, 1990; Xu, 1994; Moore & Jongman, 1997; Francis et al., 2003), and 2) whether certain sandhi-derived tones can be distinguished from the claimed output tones according to the tone sandhi rules proposed in the literature (e.g., Wang & Li, 1967; Speer, Shih, & Slowlaczek, 1989; Chen, 2013; Chen, Liu, & Kager, 2015). These studies, though providing us with some insights into how listeners perceive lexical tones in connected speech, have left many issues open for further investigation. For example, one question is how listeners perceive distorted f_0 realization itself without information from the tonal context. Having known that there are various types of f_0 variability when lexical tones are combined so that the f_0 realization is distorted to varying degrees, one further question is whether these different types of tonal variability might affect listeners' perception of lexical tones.

One methodological limitation of previous studies on tone perception is that these studies mainly focused on listeners' end-state performances, as they were required to e.g., make choices among multiple options within a time limit (e.g., Lin & Repp, 1989; Shen et

al., 1993; Peng, 1996; Moore & Jongman, 1997; Francis et al., 2003; Chen, 2013; also see McGuire, 2000 for a methodological review on speech perception). Multiple models of speech recognition, e.g., Shortlist Model (Norris, 1994), Cohort Model (e.g., Marslen-Wilson & Tyler, 1980) and TRACE model (McClelland & Elman, 1986), have shown that speech signals are perceived in an incremental fashion, along the unfolding of which listeners actively evaluate the heard signals in real time, and gradually activate the relevant lexical candidates stored in the mental lexicon. Lexical tones are found to be perceived in the similar accumulating way (e.g., Malins & Joanisse, 2010; Shen, Deutsch, & Rayner, 2013). End-state responses, therefore, are only revealing one aspect of this process (see a detailed discussion on this issue in Spivey, 2007).

Chapter 4, via the Visual World Paradigm, investigates the effect of contextual tonal variation on speech recognition in Tianjin Mandarin by monitoring the eye movements of the listeners when presented with different auditory target stimuli. Three types of contextual tonal variability as observed in Chapter 3 have been used as the target stimuli: near-merger sandhi, no-merger sandhi, and no sandhi tonal coarticulation.

The results yield significant perceptual differences among different types of tonal variation, which affect online speech processing differently, as reflected in the different eye movement patterns. No-sandhi coarticulation was easier to recognize than tone sandhi, and between the two sandhi variation types, near-merger sandhi was more difficult to process than no-merger sandhi. This has thus not only shown that in addition to the facilitation of the contextual information in the perception of lexical tones in context, the different degrees of tonal variability exert an effect on tonal recognition, but also suggested the necessity of differentiating tonal variability of different types that have been observed in Chapter 3.

1.3 Nature of tone sandhi

Among various ongoing debates about tone sandhi, one issue further discussed in this dissertation is whether tonal neutralization is involved in the process of sandhi changes. In previous impressionistic studies, tone sandhi in many Chinese dialects has been commonly regarded as the complete tonal change from one lexical tone into another lexical tone within the tonal inventory (e.g., Chao, 1948; Shen, 1992; Yip, 2002; Chen et al., 2015; also see Chen, 2000 for a review). As introduced earlier, it is claimed that in Standard Chinese, when two T3s (dipping tone) are combined, the first T3 is changed into T2 (rising tone) due to the resembling / θ realization of T3_{sandhi} vs. the lexical T2 (e.g., Chao, 1948). Further evidence from perceptual studies also shows that native listeners cannot reliably distinguish T3_{sandhi} vs. T2 using offline meta-linguistic tonal discrimination tasks (e.g., Wang & Li, 1967; Chen et al., 2015). This has led to the traditional definition of Low tone sandhi as the complete tonal change from T3 to T2. Generally speaking, descriptions of tone sandhi in the literature have been mainly about the categorical change of one tone to another, based on which various tonal alternation theories have been proposed (e.g., Chen, 2000; Wee et al., 2005; Hyman, 2007).

However, as mentioned earlier, such a perception-based conclusion has been challenged by data from production studies. Acoustic studies on Mandarin T3 sandhi have repeatedly observed that there are subtle but consistent differences between a T3_{sandhi} and a lexical T2. For example in Yuan and Chen (2014), data from natural speech corpora show that the /*ʃ*/ realization of T3_{sandhi} and the lexical T2 are mainly different in the magnitude of /*ʃ*/ rise and the duration of the /*ʃ*/ rise. Recent online studies with psycholinguistic approaches further show that the T3_{sandhi} is in fact processed differently from the lexical T2 during production (e.g., Chen et al., 2011; Zhang et al., 2014). If the sandhi-derived tones were completely changed into another lexical tone as claimed, such systematic difference in production would have been unexpected.

Furthermore, if one regards tone sandhi to involve complete tonal neutralization, the potential issue of iterative sandhi applications is expected. The iterativity of sandhi application refers to the phenomenon that the first round of tone sandhi change creates a new context for another round of sandhi application, which is not observed in Standard Chinese but in many other dialects such as Tianjin Mandarin (e.g., Li & Liu, 1985; Hung, 1987; Tan, 1987; Zhang, 1987; Wee et al., 2005). Taking the trisyllabic tonal sequence T3T1T1 in Tianjin Mandarin for example, previous literature has claimed that within T3T1T1, the tonal sequence T1T1 to the right undergoes the first round of tone sandhi change into T3T1, resulting in the intermediate stage T3T3T1, in which the T3T3 sequence to the left further triggers another round of tone sandhi to T2T3, as shown in the claimed rule: T3+T1+T1→T3+T3+T1→T2+T3+T1 (underlines indicating possible sandhi contexts at each stage; e.g., Li & Liu, 1985; Hung, 1987; Tan, 1987; Zhang, 1987; Wee et al., 2005). There are in total five such iterative application cases claimed for Tianjin Mandarin, which has led to rather complicated tone sandhi application patterns in trisyllabic sequences. If tonal neutralization is indeed not involved in tone sandhi, the iterative tone sandhi application is then called into question.

Both Chapters 3 and 5 tap into this issue. As introduced earlier, the acoustic data in Chapter 3 show significant differences in /*ʃ*/ realization between the sandhi-derived tones and the claimed output tones in both disyllabic and trisyllabic domains. In addition, Chapter 3 shows no evidence of iterative sandhi applications in the trisyllabic tone sandhi data, further confirming the non-involvement of neutralization in tone sandhi in Tianjin Mandarin. Given the findings in Chapter 3, the question that arises is whether listeners perceive sandhi tones more similarly to the original/underlying tones or to the claimed output tones?

Chapter 5 aims to answer this question by examining the patterns of tone sandhi perception. With the same Visual World Paradigm as in Chapter 4, Chapter 5 investigates the recognition time course of sandhi target stimuli when the targets were presented with different types of competitors. Three competitor types were included for each target which all had segmental overlap with the target: 1) Toneme Competitor, with toneme overlap but /*ʃ*/ contour difference from the target (Toneme+, Contour-); 2) Contour Competitor, with toneme overlap with that of the claimed output tone, therefore with contour resemblance but toneme difference from the target (Toneme-, Contour+); 3) Segmental Competitor, with no toneme or contour overlap with the target (Toneme-, Contour-). Details aside, the

main finding in Chapter 5 is that the eye movement pattern in the Contour Competitor condition exhibits a similar pattern to that in the Segmental Competitor condition during the early stage of processing, suggesting that the claimed output tones are in fact perceived more similarly to unrelated tones. This therefore, from a perceptual perspective, further confirms that tone sandhi does not involve the categorical change from one lexical tone to another in the lexical tonal system.

1.4 Neutral tone & prosodic boundary

While the previous three issues are concerned with the two local contextual tonal variation, i.e., tone sandhi and tonal coarticulation, the current issue explores the effect of a more global factor, i.e., prosodic structure, on the $/\theta/$ realization of neutral tone. It has been extensively noted that in speech production, speech utterances are grouped into constituents of various sizes (e.g., Beckman & Edwards, 1994; Ladd, 2008). The organization of these speech constituents is called the prosodic structure; and the boundaries between different constituents are called the prosodic boundaries. Although prosodic structure can sometimes be predicted from the syntactic structure of an utterance, in many cases, it is not isomorphic with the syntactic structure (see reviews in e.g., Shattuck-Hufnagel & Turk, 1996; Frota, 2012).

The role that prosodic boundaries play in speech is not only of chunking the speech, but more of signaling the level that a particular constituent belongs to within the hierarchical system of prosodic structure. The bigger the prosodic boundary, the higher the level in the hierarchy. As a crucial indicator of the prosodic structure of an utterance, prosodic boundaries can be manifested acoustically in various ways. Cross-linguistic evidence has shown that, prosodic boundaries play an important role in conditioning the articulation of segments in ways such as initial strengthening, final lengthening, and coarticulatory resistance. Specifically, the articulation of segments in the initial position within a prosodic domain is often strengthened (e.g., Fougeron & Keating, 1997; Fougeron, 1999; Cho & Keating, 2001; Fougeron, 2001; Keating, 2003; Cho & McQueen, 2005). This strengthening effect is realized in a cumulative way, so that variation due to the boundary effect changes as a function of the hierarchical level of the boundary: the higher the boundary, the greater the strengthening effect. Besides initial strengthening effect, it has also been found in various studies that there is a tendency for segments to be longer before a prosodic boundary than in other contexts, due to the slowing down of the articulation rate before a boundary, known as final lengthening (e.g., Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992; Kuzla, Cho, & Ernestus, 2007). The degree of final lengthening also varies as a function of the prosodic level: the higher the prosodic boundaries, the greater the degree of final lengthening. Another type of segmental variation brought about by prosodic boundaries is the influence on segmental coarticulation. The extent and magnitude of coarticulation varies as a function of the prosodic boundaries intervening between segments. Evidence from numerous studies has revealed that segments tend to be cumulatively resistant to influences from the contexts. In other words, the higher the level of the prosodic boundary, the less the segmental coarticulation (e.g.,

Egido & Cooper, 1980; Byrd & Saltzman, 1998; Tabain, 2003; Cho, 2004; Pan, 2007; see also Scholz & Chen 2014a on tonal coarticulation).

Similar with segments, the f_0 realization of lexical tones is also found to be subject to the effect of prosodic boundaries. For example, in Standard Mandarin, it has been found that a rising tone could be realized with a level or even slightly falling f_0 contour between a high and low tonal context when it is located at the middle of the prosodic domain (Xu, 1994). Yang and Wang (2002) show that the f_0 minima of the pre-boundary tone changes as a function of the level of the prosodic boundary: the higher the level of prosodic boundary, the lower the f_0 minima. Furthermore, it has been reported that the Low-tone sandhi (i.e., the realization of a Low tone with a rising pitch contour before another Low tone) in Standard Mandarin could be blocked in certain prosodic groupings (e.g., Shih, 1997; Zhang, 1997; Kuang & Wang, 2006).

Note that in all studies on the effect of prosodic boundaries on lexical tone realization, the subject of study has been limited to the full lexical tones which are realized with lexically distinctive f_0 contours. None of them has examined how neutral tone realization is affected by different prosodic boundaries. This has constrained our understanding of not only the effect of prosodic boundaries on tonal realization, but also the underlying mechanism of neutral tone in various Chinese dialects. It has been reported that the phenomenon of neutral tone exists not only in Standard Chinese, but also in various Chinese dialects (see reviews in Liang, 2008 and Chen, 2015). Many of these dialects have been reported to exhibit interesting neutral tone properties that are different from those of Standard Chinese. For example in Tianjin Mandarin, previous studies have reported consistent neutral tone f_0 realization preceding a lexical low-falling tone (Wang & Jiang, 1997; Wang, 2002; Lu & Wang, 2012; Li & Chen, 2011), which is not observed in Standard Chinese. However, it is noteworthy that most previous studies on neutral tone in dialects are based on impressionistic data, in which the realization of neutral tone is often regarded as the categorical tonal change. As a result, a special rising neutral tone target has been claimed when neutral tone is followed by a lexical low-falling tone in the impressionistic studies of Tianjin neutral tone (Wang & Jiang, 1997; Wang, 2002; Lu & Wang, 2012). This has thus challenged to the mid-low neutral tone target observed in Standard Chinese. Furthermore, studies with only impressionistic observations do not take global prosodic factors (e.g., prosodic boundaries) into account.

To fill this gap, Chapter 6 shows a good example of prosodically-conditioned neutral tone realization in Tianjin Mandarin. With consideration of the effect of prosodic boundaries on neutral tone realization, Chapter 6 shows that the rising neutral tone f_0 contour before T1 does not need to be treated as a special tonal target as claimed, but rather should be due to the raising effect brought about by the following T1 as first proposed in Chapter 3. Neutral tone in Tianjin Mandarin has a mid-low neutral tone target similar to that in Standard Chinese.

In conclusion, while both local and global factors are known to affect both speech production and speech perception, much less has been investigated on how lexical tones are produced and perceived in connected speech. This dissertation aims to gain further understanding of tonal variability through tapping into the afore-mentioned four issues

with evidence from Tianjin Mandarin. In the rest of this dissertation, Chapter 2 will first give a comprehensive description of the phonological system of Tianjin Mandarin. Chapter 3 will set out to investigate the f_0 variability induced by tone sandhi vs. tonal coarticulation. The perceptual consequences of the contextual tonal variability will be addressed in Chapter 4. Chapter 5 will set out to further understand the online perception of sandhi tones. In Chapter 6, the effect of prosodic boundaries on neutral tone realization will be investigated. Chapter 7 will summarize the research questions and conclude the main findings throughout the dissertation. Note that Chapters 2-5 are prepared as individual journal articles, and therefore contain some overlapping background information.