Language Sciences 48 (2015) 62–69

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

Cross-linguistic perception of Mandarin tone sandhi

Ao Chen*, Liquan Liu, René Kager

Utrecht University, Institute of Linguistics, Trans 10, 3513JK, Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form 3 December 2014
Accepted 4 December 2014
Available online 12 January 2015

Keywords:
Cross-linguistic
Speech perception
Functional basis
Tone sandhi

ABSTRACT

In the current study, we examined what forms the phonological knowledge of T3 sandhi among native Mandarin listeners and we ask the question whether there are cognitively based perceptual constraints that relate to T3 sandhi. In Mandarin Chinese, T3 sandhi restricts the co-occurrence of low tones (T3), and neutralizes the rising tone (T2) and T3 when followed by another T3. We tested native Mandarin Chinese listeners and Dutch listeners on bisyllabic tone discrimination tasks. We found that the phonological knowledge of native listeners involves the neutralization between T2T3 and underlying T3T3 sequences, rather than perceiving a boundary between consecutive T3s. Dutch listener, without the phonological knowledge of Chinese tones, also perceive T2T3 as most similar to the underlying T3T3 sequence. Such cross-linguistic consistency in perception suggests a common functional basis, where memory limitation may play a role. Listeners may attend more to the offset of an incoming tone, and as T2T3 and T3T3 share the same offset, they tend to be misperceived as identical. Regardless of the cross-linguistic consistency in perception, the misperception between T2T3 and underlying T3T3 is more pronounced among Mandarin listeners, showing that phonological knowledge shapes perception in return.

1. Introduction

In tone languages, lexical tones are pitch variations that distinguish meaning at the lexical level. Tone sandhi refers to the tonal alternation conditioned by adjacent tones. Tone sandhi is pervasive in various Chinese dialects (*e.g.*, Chen, 2000). The most studied tone sandhi is that of the Mandarin T3 tone. In Mandarin Chinese, there are four lexical tones, namely, high-level (T1, 55), mid-rising (T2, 35), low-dipping (T3, 214), and high-falling (T4, 51), where the numbers in parentheses gives the pitch values of the tones, with 5 indicating the highest level and 1 indicating the lowest level. Though there is a dip along the pitch contour of T3, the distinctive feature of T3 is "low" (Wu, 1982; Shih, 1997). When T3 occurs before a pause, it is realized with a low-dipping contour whereas; when followed by another tone except T3, it changes to a low-falling tone. In Mandarin, T3 sandhi restricts the co-occurrence of two T3s, such that the first T3 changes to a T2 when followed by another T3, *i.e.*, /214 214/ → /35 214/, producing the surface sequence T2T3. For example, "good rice" in Mandarin Chinese is composed of two syllables /hau/ and /mi/, for both of which the underlying tone is T3. As T3 sandhi applies, the first T3 changes to a T2, so that /hau/ carries a T2 and /mi/ carries a T3. By changing an underlying T3 into T2 when followed by another T3, T3 sandhi neutralizes the underlying difference between T2T3 and T3T3. For example, /hau mi/ 'millimeter', is composed of two syllables /hau/ and /mi/, which are underlyingly specified as T2 and T3. For native Mandarin listeners, 'good rice' and 'millimeter' sound the same. Fig. 1 plots the pitch contours of natural productions of "milimeter" and "good rice" of a female native

* Corresponding author. Tel.: +31 30 253 5853.
E-mail address: irischen71@hotmail.com (A. Chen).

http://dx.doi.org/10.1016/j.langsci.2014.12.002
0388-0001/

speaker. Perceptually, the sandhied T3, which has a surface form of T2, becomes indistinguishable from an underlying T2 for native listeners (Ho, 1976; Wang and Li, 1967; Peng, 1996; Huang, 2004). T3 sandhi is a productive process belonging to the Mandarin phonological grammar (Cheng, 1968; Speer et al., 1989; Shih, 1986). Though T3 sandhi has been widely studied, the processing of this grammar is still not fully understood. In the current study, we examine what composes the perceptual knowledge of the T3 sandhi of native listeners and we ask whether there are possible cognitively based perceptual constraints that relate to this grammar.

The various tone sandhi patterns in Chinese dialects can be divided into two types, left-dominant *versus* right-dominant (Yue-Hashimoto, 1987; Chan and Ren, 1989; Chan, 1995; Chen, 2000). Zhang (2007) argued that left-dominant sandhi involves rightward extension (*i.e.*, the tonal value of the left syllable is spread to the right syllable). In contrast, right-dominant sandhi, such as T3 sandhi, involves "default insertion and paradigmatic neutralization of non-final tones" (Zhang, 2007, p. 1). In T3 sandhi, the right T3 keeps its base tone whereas the left T3 is neutralized to T2. Zhang (2007) argues that the right dominance in tone sandhi is time-constrained, as the longer duration of the final syllable facilitates the realization of complex tonal contour (Oller, 1973; Klatt, 1975; Beckman and Edwards, 1990; Edwards et al., 1991; Wightman et al., 1992; Johnson and Martin, 2001; Lehiste, 1960; Lindblom, 1968). Contextual alternation tends to occur where such alternation is least noticeable (Hume and Johnson, 2001), and T2 and T3 are the acoustically most similar tones in Mandarin Chinese (Hume and Johnson, 2003); thus, it is not surprising that T3 sandhi neutralizes T3 to T2. The time constraint hypothesis may account for the fact that the rising tail of T3 is lost when preceding another non-T3 tone whereas T3 is fully realized before a pause, as it has more time to realize its complex contour in the latter case. However, this does not explain why the neutralization only occurs on the preceding syllable. For avoiding the co-occurrence of T3s, if the only constraint considered is time, then a T3T2 sequence is predicted to be as good a candidate as a T2T3 sequence. What we have not fully understood is why T3 sandhi operates in such an asymmetrical way. The asymmetry in Mandarin T3 sandhi is two-fold: first, the left syllable rather than the right in T3T3 undergoes tonal alternation; second, T2 can be a phonetic realization of T3 whereas T3 can never be a realization of T2.

Interestingly, although the perception of T3 sandhi has been studied extensively, the question of what forms perceptual knowledge of T3 sandhi still remains unclear. Previous literature has been limited to the perception single syllables and the comparison between sandhied T3 and underlying T2. It is widely agreed that the rising pitch contour derived by T3 sandhi is perceptually identical to an underlying T2 (Wang et al., 1963; Speer et al., 1989; Peng, 1996), although acoustically they are not completely neutralized (Chen and Yuan, 2007). However, T3 sandhi does not apply on single syllables, but only on a multi-syllable domain. Hence, the perceptual knowledge of T3 sandhi should also be examined in multi-syllable domains. At least two hypotheses about the perceptual knowledge of T3 sandhi should be examined: first, it could be that native listeners perceive T3T3 as an illegal sequence and, accordingly, when they encounter two T3s, they perceive a boundary between them. Second, it could be that native listeners neutralize the underlying T3T3, where both T3s are in their underlying form, with the surface T3T3 being perceived as if T3 sandhi had applied, namely T2T3.

As mentioned earlier, right dominant tone sandhies abound in Chinese dialects. Where right dominant sandhi occurs, insertion and paradigmatic neutralization of non-final tones is the default pattern. The consistency in neutralizing the non-final tone raises the question whether such asymmetrical neutralization has a functional basis. Ohala (1981, 1993, 2012) argued listeners form the basis for some sound structures and listeners can be a source of sound change. Similarly, Blevins (2006) proposed that sounds in human speech possibly have developed into discrete contrastive categories as a result of physical constraints of the perceptual system. It is widely accepted that the way in which sounds are structured can reflect cognitive constraints in the listeners. Steriade (1995, 1997) argued that the neutralization of laryngeal categories occurs in contexts where perceptual cues to this contrast would be diminished. Cross-linguistically, vowel systems tend to occupy the

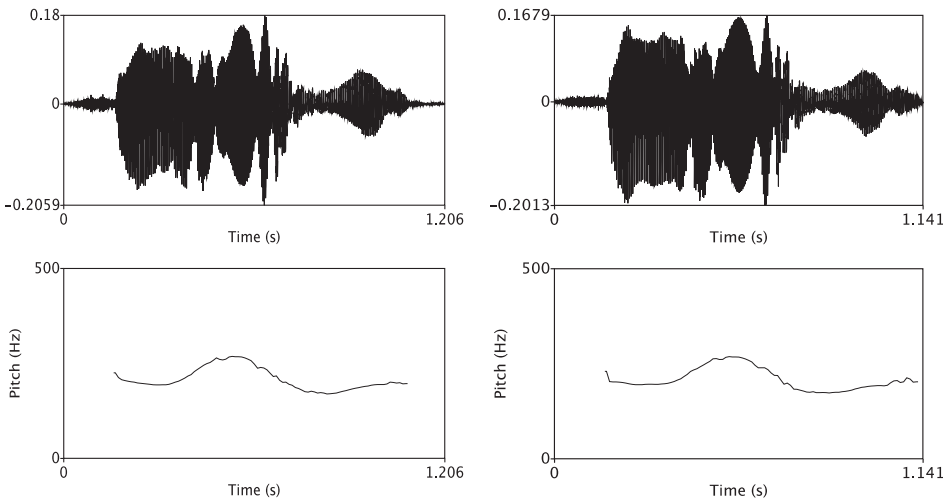

Fig. 1. Natural production of "milimeter" (left panel) and "good rice" (right panel) of a female speaker.